ISSN 1847-2001 (Print)
ISSN 1848-2295 (Online)

UNIVERSITY OF ZAGREB
FACULTY OF ORGANIZATION AND INFORMATICS
VARAŽDIN

Central European CONFERENCE on Information and Intelligent Systems

Proceedings

36th International Scientific Conference

CECIIS

Organized by:
University of Zagreb
Faculty of Organization and Informatics

September 17-19/2025
Varaždin, Croatia

ISSN 1847-2001 (Print)
ISSN 1848-2295 (Online)

UDC 004:007

36th International Scientific Conference

Central European Conference on Information and Intelligent Systems

CECIIS

Proceedings

September 17th – 19th, 2025
Varaždin, Croatia

Organized by:
University of Zagreb
Faculty of Organization and Informatics

Publisher
University of Zagreb Faculty of Organization and Informatics, Pavlinska 2, HR-42000 Varaždin, Croatia

For Publisher
Marina Klačmer Čalopa, University of Zagreb Faculty of Organization and Informatics, Croatia

Editors
Nina Begičević Ređep, University of Zagreb Faculty of Organization and Informatics, Croatia
Marijana Bubanić, University of Zagreb Faculty of Organization and Informatics, Croatia

Contact
E-mail: ceciis@foi.unizg.hr
Phone: +385 42 390 848
Web: ceciis.foi.hr
Fb: facebook.com/ceciis.foi

Editorial Board
Victoria Appatova, University of Cincinnati, USA
Igor Balaban, University of Zagreb, Croatia
Marko Bohanec, Jozef Stefan Institute, Slovenia
Leo Budin, University of Zagreb, Croatia
Aruna Chandra, Indiana State University, USA
Violeta Cvetkoska, University in Skopje, Macedonia
Luis de Marcos Ortega, University of Alcalá, Spain
Blaženka Divjak, University of Zagreb, Croatia
Jasminka Dobša, University of Zagreb, Croatia
Juraj Dud'ak, Slovak University of Technology in Bratislava, Slovakia
Matjaž Garms, Jozef Stefan Institute, Slovenia
Gabriel Gaspar, Slovak University of Technology in Bratislava, Slovakia
Gordan Gledec, University of Zagreb, Croatia
Andrina Granić, University of Split, Croatia
Marjan Heričko, University of Maribor, Slovenia
Dieter Hertweck, Herman Hollerith Research Centre, Germany
Marina Klačmer Čalopa, University of Zagreb, Croatia
Michal Kohani, University of Žilina, Slovakia
Mario Konecki, University of Zagreb, Croatia
Vladimir Kovšca, University of Zagreb, Croatia
Marjan Krašna, University of Maribor, Slovenia
Lode Lauwaert, KU Leuven, Belgium
Sandra Lovrenčić, University of Zagreb, Croatia
Ivan Luković, University of Belgrade, Serbia
Ivan Magdalenić, University of Zagreb, Croatia
Marcelo Fabian Maina, Open University of Catalonia, Spain
Aleksandar Marković, University of Belgrade, Serbia
Peter Marton, University of Žilina, Slovakia
Marta Mas Machuca, UIC Barcelona, Spain
Dunja Mladenič, Jožef Stefan Institute, Slovenia
Oliver Moravčik, Slovak University of Technology, Slovakia
Enrique Mu, Carlow University, USA
Karlo Nyarko, University of Osijek, Croatia
Jan Paralič, Technical University of Košice, Slovakia
Elisabeth Pergler, CAMPUS 02 - University of Applied Sciences, Austria
Krzysztof Podlaski, University of Lodz, Poland
Patrizia Poščić, University of Rijeka, Croatia
Sonja Ristić, University of Novi Sad, Serbia
Markus Schatten, University of Zagreb, Croatia
Zlatko Stapić, University of Zagreb, Croatia
William Steingartner, Technical University of Kosice, Slovakia
Vjeran Strahonja, University of Zagreb, Croatia
Marko Tadić, University of Zagreb, Croatia
Martina Tomičić Furjan, University of Zagreb, Croatia
Katarina Tomičić Pupek, University of Zagreb, Croatia
Jan Tretmans, Radboud University Nijmegen, Netherlands
Wim Van Petegem, KU Leuven, Belgium
Violeta Vidaček Hainš, University of Zagreb, Croatia
Sirje Virkus, Tallinn University, Estonia
Neven Vrček, University of Zagreb, Croatia
Mladen Vouk, North Carolina State University, USA
Diana Šimić, University of Zagreb, Croatia

Publishing Board
Marko Matus, University of Zagreb Faculty of Organization and Informatics, Croatia
Jelena Gusić Munđar, University of Zagreb Faculty of Organization and Informatics, Croatia
Bernarda Kos, University of Zagreb Faculty of Organization and Informatics, Croatia
Lea Masnec, University of Zagreb Faculty of Organization and Informatics, Croatia

All rights reserved. No parts of this publication may be reproduced, stored in a retrieval system or transmitted in any form by any means, electronic, mechanical photocopying, recording or otherwise without the prior permission of the copyright owner.
CECIIS Proceedings have been indexed in the following databases: ProQuest (SciTech Premium Collection, ProQuest Technology Collection and ProQuest Advanced Technologies & Aerospace database) and Web of Science.

ISSN 1847-2001 (Print)
ISSN 1848-2295 (Online)



Creative Commons Licence
Central European Conference on Information and Intelligent Systems Proceedings is licensed under a Creative Commons Attribution-NonCommercialNoDerivatives 4.0 International License.
http://creativecommons.org/licenses/by-nc-nd/4.0/

Contents

Index of Authors	VII
List of Reviewers	IX
Preface	XI
In Memoriam - Prof. Emer. Boris Aurer (1931—2023)	XV

Computer Games

Markus Schatten, Jakov Malović, Ivan Barišić, Antonio Martinaga, Jakov Marčan, Sven Pavlić, Petar Tonković **An AI Enhanced Game Master Assistant for Table Top Role Playing Games**	3
Markus Schatten, Lara Erdogan, Lana Maček, Tin Muhar, Marin Vabec, Josip Novak **Development of a 3D Platformer Game for a Pseudo-Holographic Gaming Console**	9
Bogdan Okreša Đurić, Tomislav Peharda, Hrvoje Zema **Influencing the Output of a Large Language Model using Defined Personality Trait Facets**	17
Tomislav Peharda, Bogdan Okreša Đurić **Generative Worldcrafting: A Modular Framework for AI-Assisted Content Creation in Games**	23
Marinela Schatten, Ljiljana Inkret-Martinčević, Martina Košćak **Play It Safe: Evaluating Participatory Game-Based Learning Activities for Internet Safety Among Primary and Secondary School Students**	29

Cyber-Physical Systems and Intelligent Energy

Zvonimir Šimić, Danijel Topić, Marina Dubravac, Nemanja Mišljenović **The Potential of Using a Small PV Battery System to Overcome Electricity Supply Disruptions**	39
Milan Daňo, Martin Juhás, Patrik Gašparovič, Bohuslava Juhásová, Fedor Burčiar **Use of Piezoelectric Technology for Fluid Valves as an Energy Saving Alternative**	45
Lukrecia Vulić, Ivan Aleksi, Tomislav Matić **Comparative Analysis of Bandpass IIR Filters on Microphone PDM Signal for FPGA Implementation**	53
Juraj Ďuďák, Juraj Repka **Deep-Learning and Stereo Camera Based System for Real-Time Virtual Safety Boundaries**	61
Ivan Kolarić, Karlo Pepelnik, Dino Vlahek, Dijana Oreški, Jelena Gusić Munđar, Bojan Žugec, Nikola Kadoić **Application of Artificial Intelligence and Internet of Things in Rehabilitation and Health Recovery**	69
Lea Masnec, Vjeran Strahonja **Modeling and Construction of Circular Economy Digital Twins Based on System Dynamics**	79

Data and Knowledge Bases

Ana Petrovic, Danijela Jaksic, Kristian Stancin
The Role of Ontologies and Semantic Technologies in Developing Educational Chatbots 89

Mikel Ferrer-Oliva, José Amelio Medina-Merodio, José Javier Martínez-Herraiz, Carlos Cillerue-lo-Rodríguez
Relational Modelling of Cyberattacks: A Taxonomy for the Analysis of Threat Interactions in Digital Environments 97

Sven Vujević, Maja Cerjan, Kornelije Rabuzin
Data Quality in Distributed Information Systems: Conceptual Challenges and Empirical Findings 105

Martina Asenbrener Katic, Sanja Candrlic, Marina Rauker Koch
A Case Study on the Two-Sided Version of the Node of Knowledge Conceptual Framework 113

Nađa Klještanović, Sonja Ristić, Darko Stefanović
Parallel Utilization of SQL and NoSQL Databases –A Systematic Literature Review 121

Dajana Marnika, Maja Gligora Marković, Božidar Kovačić
Unlocking Data Quality for Responsible Artificial Intelligence 131

Ana Pasarić, Patrizia Poščić, Danijela Jakšić
Building Trustworthy Knowledge: Data Quality and Validation Protocols for Educational Knowledge Graphs 139

Designing Innovation through Productive Disruptions

Tamara Ćurlin, Jasmina Pivar
Virtual Reality as a Strategic Tool: Adoption Patterns and Distribution Channels in Croatian Hotels 149

Metodija Mircev, Violeta Cvetkoska, Igor Ivanovski
Smart Synergies: Open Innovation and Predictive Models in the Telecom-Insurance Eco-system 157

Aldo Valencia, Violeta Cvetkoska, Bojan Kitanovikj, Alison Pearce
Tactile Collaboration and Backcasting to Design Meaningful Futures and Strategies 173

Domagoj Tuličić, Snježana Križanić
Digital Twin-Enhanced Fire Response Through Unmanned Aerial Reconnaissance System 183

Digital Transformation and Corporate Culture

Anamarija Zron
Impact of Technostress on the Adoption of Artificial Intelligence - A Bibliometric Analysis 193

Krešimir Kavran, Maja Pušnik
Planning and Development Processes for the Information System Based on the Command Query Responsibility Segregation Design Pattern 201

Lebohang Nkhabu, Machdel Matthee
Barriers to Cyber Hygiene in Hybrid Work Environments: A Case Study of a Lesotho Devel-opment Parastatal 213

Daniel Fernando Olmedo Mejía, Violeta Cvetkoska
From Human-Centered to Hybrid: Mapping the Emergence of GenAI in Organizational Actor-Networks 221

Zvonimir Vukić
Digitalization of Accounting in the 21st Century – A Bibliometric Analysis 229

Anamarija Horvat, Renata Mekovec
Protecting privacy on social media platforms: a case study of Croatian users 239

Krešimir Kavran, Boštjan Šumak
Indirect User Experience Evaluation: Insights from Employees Working on a Public Sector Information System 247

Igor Pihir, Ivan Kardašić
Digital Transformation in the Gaming Industry 259

Innovation, Entrepreneurship and Economic Growth

Antonia Kurtović, Kristina Šorić
AI and Simulation in Entrepreneurial Education: Lessons from the RITup Ventures Program 269

Natali Belušić, Kristina Šorić
Digitalization of Banking Processes and Adaptation for Elderly People 277

Ivona Kocijan, Kristina Detelj
Digital Resilience of SMEs: Conceptual Distinction and Strategic Factors in the Context of Disruptions 283

Walter Rath, Berndt Jesenko, Harris Gerzic
Potentials and Barriers of Using Big Data in Austrian SMEs 291

Idoia de Paz Gómez
Cultural Factors that Inhibit or Facilitate Innovation in Family Businesses: A Systematic Literature Review 299

Luka Juričinec, Marina Klačmer Čalopa
Digital Transformation of HR Practices: A Systematic Review and Bibliometric Analysis of e-HRM Literature / Digitalna transformacija HR prakse: Sustavni pregled i bibliometrijska analiza literature o e-HRM-u 313

Marija Vugrinec, Katarina Pažur Aničić
IT project management: Does sustainability matter? 323

Tanja Šestanj-Perić
The Influence of Blockchain-based Triple-entry Accounting on Accounting Information – a Theoretical and Conceptual Perspective 339

Ivan Turčić, Ana-Marija Stjepić, Tomislav Jurić
Leveling Up Process Efficiency: AI as a Game-Changer for SMEs 345

Intelligent Systems

Ivica Lukić, Tea Krčmar, Dina Šabanović, Miljenko Švarcmajer, Mirko Köhler, Marko Carević
Beyond Traditional BI: Large Language Models as Game-Changers in Document Processing 355

Karlo Babić, Ana Meštrović
Pyramidal Recursive Composition of Multi-Word Units into Unified Representations 363

Marija Brkić Bakarić, Lucia Načinović Prskalo, Petra Kuljiš
Detecting Discrimination Using Selected Data Mining Methods 381

Saša Mitrović, Neven Vrček
Methodology for the comparative analysis of PCA and Autoencoders in dimensionality reduction: impact on classification accuracy and computational efficiency 395

Ivana Kovačević, Ivan Aleksi, Tomislav Matić, Josip Zidar
Smart Cuvette: An Integrated Optical System for Ethanol Measurement in Liquid Samples 411

Bojan Radišić, Ivan Dunđer, Sanja Seljan
Impact of Imputation Methods on the Performance of Classification Algorithms 419

Dominik Čambál, Martin Pukanec, Monika Herchlová, Juraj Ďuďák, Matúš Nečas
Performance study of a bidirectional vanilla spiking neural network for signal noise suppression 429

Luka Lukač, Damjan Strnad, Štefan Horvat
Compressibility of WGANs in a Distributed Learning Environment 433

Marko Vještica, Elena Akik, Vladimir Dimitrieski, Lukas Hinterleitner, Jonas Wieczorek, Jovan Erić, Frank-Christian Weidenfelder
Towards User-Centric Chatbot for Q&A in Industry: Evaluation of Employees' Experience 445

Goran Bubaš
How ChatGPT and Gemini View the Elements of Communication Competence of Large Language Models: A Pilot Study 455

Dalibor Bužić, Jasminka Dobša
Analysis of Popular Croatian Song Lyrics 467

Learning Analytics, AI and Technologies in Education

Vladan Čolić, Enes Sukić
The Impact of Text-to-SQL Systems on Student Performance and Satisfaction in SQL Tasks in Higher Education 477

Antoun Yaacoub, Jérôme Da-Rugna, Zainab Assaghir
Lightweight Prompt Engineering for Cognitive Alignment in Educational AI: A OneClickQuiz Case Study 489

Maja Kućar, Andreja Brajša-Žganec, Tamara Bolić
Children's Screen Time, Preoccupation with Technology, Social Media Use, and Playing Video Games as Possible Predictors of GPA – a Longitudinal Study 497

Caitlin Braga, Adriana Aletta Steyn
The AI Classroom: Systematic Literature Insights into Quality and Access in Education 507

Jurica Grzunov, Ljiljana Zekanović-Korona, Zdravko Kunić
Attitudes Toward AI in Academic Assessment: A Survey of Croatian Higher Education Teachers 517

Kelebogile M Dibetso, Adriana Aletta Steyn
Enhancing Learning Experiences: The role of mixed reality in Higher Education 527

Krešimir Pavlina, Ana Pongrac Pavlina, Anita Modrušan
Between Innovation and Reality: Teachers' Attitudes Toward ICT in Teaching 539

Ivana Ružić, Igor Balaban
Can We Trust AI in the Classroom? Examining Ethical and Privacy Challenges of LLMs – A Literature Review 549

Goran Hajdin, Igor Balaban, Marko Matus
The Dual-Edged Sword: Exploring Student Perceptions and Emotional Responses to AI in Academic Writing 557

Sevgi Koyuncu Tunç
Software Standards in Action: A Complaint-Driven Game Inspired by Real Project Failures 565

Daniel Hari, Maja Kerneža, Dejan Zemljak, Andrej Flogie
Social Robots and AI-Driven Learning: A Pilot Study on Student Engagement in Higher Education 575

Melita Milić, Blaženka Divjak
Exploring Teachers' Acceptance of Digital Textbooks: A Pilot Study in Croatian Primary Schools 581

Mia Šagovac, Dorotea Potoč, Tomislav Jagušt
The Impact of Different Game Modes in an Educational Web Application on Student Engagement and Performance 591

Maja Kerneža, Daniel Hari, Andrej Flogie, Dejan Zemljak
Disrupting Reading through the Invisible Presence of AI: Rethinking Credibility and Critical Reading in Primary Education 599

Jasenka Begić, Goran Đambić, Ana Lokas Ćošković, Vanja Šebek, Ana Tecilazić, Karolina Zelenika
Teachers' digital competences and the connection of lifelong learning in higher education 605

Valentina Kirinić, Mihaela Bosak
Learning path definition and usage in the context of learning analytics: literature review 615

Strategic Planning and Decision Making

Marta Mas-Machuca, Frederic Marimon, Anna Akhmedova
Generative AI and Workplace Productivity: A Qualitative Study in Spain 625

Marta Mas-Machuca, Alba Manresa, Jasmina Berbegal-Mirabent
The Impact of Environmental Training on Employee Motivation and Corporate Sustainability in Spanish Companies 633

Predrag Trpeski, Violeta Cvetkoska
Bridging Systems, Shaping Futures: A Multilevel Strategy for Curriculum Transformation for Innovation and Resilience in Higher Education 643

Marko Bohanec
Using Large Language Models for Creating, Modifying and Interpreting Decision Tables in Multi-Criteria Method DEX 651

Sabina Rako, Nina Begičević Ređep, Barbara Šlibar, Diana Šimić, Bart Rienties
Students' Cognitive Engagement in Blended Learning: Observations from Event-contingent Self-report and Log Data 659

Enrique Mu, Anna Florek-Paszkowska, Santiago Luis Risco-Martinez
Exploring the Impact of Digital Infrastructure Challenges on Learning Outcomes and Student Satisfaction 667

Sanja Kelemen, Nemanja Mišljenović, Goran Knežević, Stjepan Babić
Application of Multi-Objective Optimization for Short-Term Operation Planning of an Energy Community 677

Krisztián Attila Bakon, Tibor Holczinger
Reactive Scheduling under Due-Date and Intermediate Storage Constraints: A Conceptual Multi-Objective Approach 685

Teresa Monllau, Nikola Kadoić, EnricoTesti, Silvia Sordi, Jaime Lacueva, Ivan Malbašić
A conceptual model to assess and develop digitalisation in the social economy's people care sector: the case of Prato (Italy), Mataro (Spain) and Varaždin (Croatia) 693

Nikica Daraboš, Violeta Vidašek Hainš, Nevena Ladić
Gender equality in the fields of information and communication technology in healthcare 703

Jin Eun Yoo, Nam-Hwa Kang, Suna Ryu, Jun-ki Lee, Youngsun Kwak, Hyeong Gwan Kim
Developing an LLM-specific Coding Framework for Prediction and Feedback Modeling of Science Teachers' Instructional Expertise 711

Systems for Smart Mobility

Dunja Srpak, Josip Srpak, Miljenko Mustapić **Application of Internet of Thins in urban mobility**	723
Sirje Virkus, Marianne Paimre **Telepresence Robots as Emerging Technologies in Elderly Care and Wellbeing: The Case of Estonia**	731
Mihael Radovan **Dynamic Lane Reversal Review**	743

Trends in Software Engineering

Boris Tomaš **Impact Analysis of Croatian Language as the Publication Language on the Research Paper Visibility and Citation Potential in Computer Science and Engineering / Analiza utjecaja hrvatskog jezika kao jezika publikacije na vidljivost i citiranost znanstvenih radova u području računalnih znanosti i inženjerstva**	759
Marija Kuštelega **Dark Patterns in the Public Sector: A Study of Cookie Consent Interfaces on Croatian Websites**	771
Pablo Hidalgo, Marko Mijač, Adrián Domínguez, Daniel Rodriguez **A preliminary Study on the Use of Generative AI in Software Engineering Education**	779
Marko Horvat, Barbara Kralj, Gordan Gledec **A Comparative Study of Vibe Coding with ChatGPT and Gemini in Front-end Web Development**	787
Roberta Capuano, Henry Muccini, Zlatko Stapic **Implementing Education 4.0 in Software Engineering Education: the JCC experience**	797

Publication Ethics And Malpractice Statement	809
Author Guidelines	813
CECIIS 2025 Sponsors	827
CECIIS 2025 Program and Organizing Committees' Members, Research, Professional Tracks' Chairs and Student Poster and Presentation Section Chair	835

Index of Authors

Akhmedova, Anna 625
Akik, Elena 445
Aleksi, Ivan 53 ,411
Aničić, Katarina Pažur 323
Assaghir, Zainab 489
Babić, Karlo 363
Babić, Stjepan 677
Bakarić, Marija Brkić 381
Bakon, Krisztián Attila 685
Balaban, Igor 549 ,557
Barišić, Ivan 3
Begić, Jasenka 605
Begičević Ređep, Nina 659
Belušić, Natali 277
Berbegal-Mirabent, Jasmina 633
Bohanec, Marko 651
Bolić, Tamara 497
Bosak, Mihaela 615
Braga, Caitlin 507
Brajša-Žganec, Andreja 497
Bubaš, Goran 455
Burčiar, Fedor 45
Bužić, Dalibor 467
Candrlic, Sanja 113
Capuano, Roberta 797
Carević, Marko 355
Cerjan, Maja 105
Cilleruelo-Rodríguez, Carlos 97
Cvetkoska, Violeta 157 ,173 ,221 ,643
Čalopa, Marina Klačmer 313
Čambál, Dominik 429
Čolić, Vladan 477
Ćošković, Ana Lokas 605
Ćurlin, Tamara 149
Da-Rugna, Jérôme 489
Daraboš, Nikica 703
Daňo, Milan 45
Detelj, Kristina 283
Dibetso, Kelebogile M 527
Dimitrieski, Vladimir 445
Divjak, Blaženka 581
Dobša, Jasminka 467
Domínguez, Adrián 779
Dubravac, Marina 39
Ďuďák, Juraj 61 ,429
Dunđer, Ivan 419
Đambić, Goran 605
Đurić, Bogdan Okreša 17 ,23
EnricoTesti 693
Erdogan, Lara 9
Erić, Jovan 445
Ferrer-Oliva, Mikel 97
Flogie, Andrej 575 ,599
Florek-Paszkowska, Anna 667
Gašparovič, Patrik 45
Gerzic, Harris 291
Gledec, Gordan 787
Grzunov, Jurica 517
Gómez, Idoia de Paz 299
Hainš, Violeta Vidašek 703
Hajdin, Goran 557
Hari, Daniel 575 ,599
Herchlová, Monika 429
Hidalgo, Pablo 779
Hinterleitner, Lukas 445
Holczinger, Tibor 685
Horvat, Anamarija 239
Horvat, Marko 787
Horvat, Štefan 433
Inkret-Martinčević, Ljiljana 29
Ivanovski, Igor 157
Jagušt, Tomislav 591
Jaksic, Danijela 89
Jakšić, Danijela 139
Jesenko, Berndt 291
Juhás, Martin 45
Juhásová, Bohuslava 45
Jurić, Tomislav 345
Juričinec, Luka 313
Kadoić, Nikola 69 ,693
Kang, Nam-Hwa 711
Kardašić, Ivan 259
Katic, Martina Asenbrener 113
Kavran, Krešimir 201 ,247
Kelemen, Sanja 677
Kerneža, Maja 575 ,599
Kim, Hyeong Gwan 711
Kirinić, Valentina 615
Kitanovikj, Bojan 173
Klještanović, Nađa 121
Knežević, Goran 677
Koch, Marina Rauker 113
Kocijan, Ivona 283
Kolarić, Ivan 69
Kovačević, Ivana 411
Kovačić, Božidar 131
Košćak, Martina 29
Kralj, Barbara 787
Križanić, Snježana 183
Krčmar, Tea 355
Kuljiš, Petra 381
Kunić, Zdravko 517
Kurtović, Antonia 269
Kućar, Maja 497
Kuštelega, Marija 771
Kwak, Youngsun 711
Köhler, Mirko 355
Lacueva, Jaime 693
Ladić, Nevena 703
Lee, Jun-ki 711
Lohaj, Oliver 403
Lukač, Luka 433
Lukić, Ivica 355
Malbašić, Ivan 693
Malović, Jakov 3
Manresa, Alba 633
Marimon, Frederic 625
Marković, Maja Gligora 131
Marnika, Dajana 131
Martinaga, Antonio 3
Martínez-Herraiz, José Javier 97
Marčan, Jakov 3
Mas-Machuca, Marta 625 ,633
Masnec, Lea 79
Matić, Tomislav 53 ,411
Matthee, Machdel 213
Matus, Marko 557
Maček, Lana 9
Medina-Merodio, José Amelio 97

Mejía, Daniel Fernando Olmedo 221
Mekovec, Renata 239
Meštrović, Ana 363
Mijač, Marko 779
Milić, Melita 581
Mircev, Metodija 157
Mitrović, Saša 395
Mišljenović, Nemanja 39 ,677
Modrušan, Anita 539
Monllau, Teresa 693
Mu, Enrique 667
Muccini, Henry 797
Muhar, Tin 9
Munđar, Jelena Gusić 69
Mustapić, Miljenko 723
Nečas, Matúš 429
Nkhabu, Lebohang 213
Novak, Josip 9
Oreški, Dijana 69
Paimre, Marianne 731
Pasarić, Ana 139
Pavlina, Ana Pongrac 539
Pavlina, Krešimir 539
Pavlić, Sven 3
Pearce, Alison 173
Peharda, Tomislav 17 ,23
Pepelnik, Karlo 69
Petrovic, Ana 89
Pihir, Igor 259
Pivar, Jasmina 149
Potoč, Dorotea 591
Poščić, Patrizia 139
Prskalo, Lucia Načinović 381
Pukanec, Martin 429
Pušnik, Maja 201
Rabuzin, Kornelije 105
Radishevska, Anastasiia 403
Radišić, Bojan 419
Radovan, Mihael 743
Rako, Sabina 659
Rath, Walter 291
Repka, Juraj 61
Rienties, Bart 659
Risco-Martinez, Santiago Luis 667
Ristić, Sonja 121
Rodriguez, Daniel 779
Ružić, Ivana 549
Ryu, Suna 711
Schatten, Marinela 29
Schatten, Markus 3 ,9
Seljan, Sanja 419
Sordi, Silvia 693
Srpak, Dunja 723
Srpak, Josip 723
Stancin, Kristian 89
Stapic, Zlatko 797
Stefanović, Darko 121
Steyn, Adriana Aletta 507 ,527
Stjepić, Ana-Marija 345
Strahonja, Vjeran 79
Strnad, Damjan 433
Sukić, Enes 477
Šabanović, Dina 355
Šagovac, Mia 591
Šebek, Vanja 605
Šestanj-Perić, Tanja 339
Šimić, Diana 659
Šimić, Zvonimir 39
Šlibar, Barbara 659
Šorić, Kristina 269 ,277
Šumak, Boštjan 247
Švarcmajer, Miljenko 355
Tecilazić, Ana 605
Tomaš, Boris 759
Tonković, Petar 3
Topić, Danijel 39
Trpeski, Predrag 643
Tuličić, Domagoj 183
Tunç, Sevgi Koyuncu 565
Turčić, Ivan 345
Vabec, Marin 9
Valencia, Aldo 173
Virkus, Sirje 731
Vještica, Marko 445
Vlahek, Dino 69
Vrček, Neven 395
Vugrinec, Marija 323
Vujević, Sven 105
Vukić, Zvonimir 229
Vulić, Lukrecia 53
Weidenfelder, Frank-Christian 445
Wieczorek, Jonas 445
Yaacoub, Antoun 489
Yoo, Jin Eun 711
Zekanović-Korona, Ljiljana 517
Zelenika, Karolina 605
Zema, Hrvoje 17
Zemljak, Dejan 575 ,599
Zidar, Josip 411
Zron, Anamarija 193
Žugec, Bojan 69

List of Reviewers

Antonova, Albena, Bulgaria
Appatova, Victoria, USA
Babić, Snježana, Croatia
Babojelić, Karlo, Croatia
Banek Zorica, Mihaela, Croatia
Banjanović-Mehmedović, Lejla, Bosnia and Herzegovina
Barać, Dušan, Serbia
Beliga, Slobodan, Croatia
Berber, Nemanja, Serbia
Bjažić, Toni, Croatia
Bohanec, Marko, Slovenia
Botički, Ivica, Croatia
Brdnik, Saša, Slovenia
Buchem, Ilone, Germany
Bušelić, Vjeran, Croatia
Carrascosa, Carlos, Spain
Čakija, Dino, Croatia
Čandrlić, Sanja, Croatia
Calcina, Erik, Slovenia
Detelj, Kristina, Croatia
Devčić, Antonela, Croatia
Dobrovic, Zeljko, Croatia
Dolenc, Kosta, Slovenia
Domínguez-Díaz, Adrián, Spain
Dreven, Nikolina, Croatia
Dudak, Juraj, Slovakia
Dudáš, Adam, Slovakia
Erjavec, Zlatko, Croatia
Ferrer Oliva, Mikel, Spain
Galinec, Darko, Croatia
Gedrimiene, Egle, Finland
Gregurec, Iva, Croatia
Grd, Petra, Croatia
Guardia Ortiz, Lourdes, Spain
Hajdin, Goran, Croatia
Hajdarević, Kemal, Bosnia and Herzegovina
Hartmann Tolić, Ivana, Croatia
Hericko, Marjan, Slovenia
Horvat, Marko, Croatia
Hrkut, Patrik, Slovakia
Hrgarek Lechner, Nadica, Austria
Ivkovic, Nikola, Croatia
Jagušt, Tomislav, Croatia
Janev, Valentina, Serbia
Jenčik, Marian, Slovakia
Jeremić, Veljko, Serbia
Kadoić, Nikola, Croatia
Kaluža, Marin, Croatia
Kirinić, Valentina, Croatia
Koltovska-Nechoska, Daniela, North Macedonia
Korent, Dina, Croatia
Kossecki, Paweł, Poland
Kovacic, Bozidar, Croatia
Krašna, Marjan, Slovenia
Krhač Andrašec, Eva, Slovenia
Kućar, Maja, Croatia
Kušić, Krešimir, Croatia
Kučina Softić, Sandra, Croatia
Kvet, Michal, Slovakia
Lacković Vincek, Zrinka, Croatia
Luković, Ivan, Serbia
Maina, Marcelo Fabián, Spain
Magdalenić, Ivan, Croatia
Mandić, Milivoj, Croatia
Maricic, Milica, Serbia
Marolt, Marjeta, Slovenia
Martí, Pasqual, Spain
Meštrović, Ana, Croatia
Mezak, Jasminka, Croatia
Milić, Luka, Croatia
Mihelič, Jurij, Slovenia
Mitrović, Zorica, Serbia
Načinović Prskalo, Lucia, Croatia
Novak, Matija, Croatia
Nyarko, Emmanuel Karlo, Croatia
Okreša Đurić, Bogdan, Croatia
Oliveras, Ester, Spain
Oreški, Dijana, Croatia
Orehovački, Tihomir, Croatia
Paralič, Jan, Slovakia
Pažur Aničić, Katarina, Croatia
Pavlina, Krešimir, Croatia
Peharda, Tomislav, Croatia
Perčević, Hrvoje, Croatia
Perši, Nenad, Croatia
Pesek, Igor, Slovenia
Pihir, Igor, Croatia
Plantak Vukovac, Dijana, Croatia
Prenc, Rene, Croatia
Pušnik, Maja, Slovenia
Radošević, Danijel, Croatia
Radaković, Davorka, Serbia
Radovanović, Sandro, Serbia
Rienties, Bart, United Kingdom
Ristić, Sonja, Serbia
Rodríguez García, Daniel, Spain
Rožanec, Jože, Slovenia
Sakač, Nikola, Croatia
Savić, Dušan, Serbia
Schatten, Markus, Croatia
Sever, Drago, Slovenia
Sittar, Abdul, Slovenia
Stančin, Kristian, Croatia
Steyn, Adriana, South Africa
Sturm, Jan, Slovenia
Šaloun, Petr, Czechia
Šarac, Vasilija, North Macedonia
Taverner, Joaquin, Spain
Tillmann, Alexander, Germany
Tomaš, Boris, Ireland
Tomičić Furjan, Martina, Croatia
Tomičić, Igor, Croatia
Tomičić-Pupek, Katarina, Croatia
Tokarčíková, Emese, Slovakia
Topal, Oleksandra, Slovenia
Umek, Lan, Slovenia
Uroda, Ivan, Croatia
Van Petegem, Wim, Belgium
Varga, Michal, Slovakia
Vasilj, Josip, Croatia
Veigend, Petr, Czechia
Vidović, Ivan, Croatia
Vrcek, Neven, Croatia
Yaacoub, Antoun, France
Žajdela Hrustek, Nikolina, Croatia
Žugec, Bojan, Croatia
Žugec, Petra, Croatia
Zornić, Nikola, Serbia

Preface

We are pleased to present you with the Proceedings of the 2025 Central European Conference on Information and Intelligent Systems. We trust that this compilation will prove to be both valuable and inspiring to you.

Contained within these Proceedings are the full text conference papers from this year's edition of the Central European Conference on Information and Intelligent Systems (CECIIS), an annual international event. This conference is organized by the University of Zagreb Faculty of Organization and Informatics Varaždin, Croatia. It took place from September 17th to 19th, 2025, in Varaždin, Croatia. For over three decades, CECIIS has successfully convened with the aim of serving as an interface between scholars and professionals in the field of information and intelligent systems and related domains.

The 36th CECIIS 2025 conference, with the thematic focus 'Designing Innovation through Productive Disruptions,' aimed to bring together the brightest minds across disciplines to explore how disruptive phenomena can be systematically understood, modelled, and harnessed to drive innovation. By focusing on the interplay between disruptive forces and the mechanisms that transform these forces into productive outcomes, the conference provides a platform for advancing cutting-edge theories, methodologies, and applications that redefine innovation in science, technology, and society. A wide range of topics were discussed at our conference that explore the intersections of disruption and innovation. The exploration of cutting-edge topics like the Science of Disruption, Designing Innovation Systems, Emerging Technologies as Disruptive Forces, Disruptions in Social and Environmental Systems, Metrics, Measurement, and Decision-Making, Educational Paradigms for Disruption.

Although no single event can fully capture the complexity of navigating innovation through disruption, CECIIS 2025 reaffirms our shared commitment to leading change rather than merely reacting to it. Through interdisciplinary collaboration, critical inquiry, and bold experimentation, we are co-creating a future in which disruption is not feared — but designed, directed, and used as a force for good.

It is also pertinent to emphasize that the convergence of Program and Organizing Committee members, academics, and experts from diverse fields allows us to create a cross-disciplinary approach to the conference. This format propels discussions towards topics poised to emerge prominently in the near future, while concurrently facilitating valuable networking among individuals with varied domains of expertise whose collective insights contribute a substantial value to the fabric of CECIIS.

The Proceedings of CECIIS 2025 encompass 86 meticulously reviewed scientific papers. These papers were subject to rigorous peer review by no fewer than two independent evaluators and subsequently endorsed by the international Program Committee. The papers span an array of categories including Computer Games, Cyber-Physical Systems and Intelligent Energy, Data and Knowledge Bases, Designing Innovation through Productive Disruptions, Digital Transformation and Corporate Culture, Innovation, Entrepreneurship and Economic Growth, Intelligent Systems, Learning Analytics, AI and Technologies in Education, Strategic Planning and Decision Making, Systems for Smart Mobility, and Trends in Software Engineering. The selection of these papers hinged upon their relevance, innovativeness, applicability to the information and intelligent systems domain, and overall scholarly merit. The official language of the conference is English.

A total of 98 papers were submitted for review. The review process was carried out through the dedicated efforts of 137 reviewers, about half from Croatia and about half from abroad. Collectively, they completed 194 reviews, leading to the acceptance of 86 papers. These papers, authored by 230 individuals representing 21 countries, are published in this Proceedings. Following the event, a selected group of papers will undergo further review, potentially leading to extended publication in the Journal of Organizational and Information Sciences (JIOS), the international scientific journal published by the University of Zagreb Faculty of Organization and Informatics. Our past Proceedings are indexed in the following bibliographic databases: WoS, Proquest, DOAJ and EBSCO.

A pivotal hallmark of this year's conference were the distinguished invited keynote speakers, individuals whose intellectual contributions have shaped their respective spheres of scientific and professional inquiry. CECIIS 2025 was honored to host the following keynote speakers:

Professor Julius Georgiou, Ph.D., who addressed the topic "Pushing the Boundaries of Modern Key-Enabling Technologies from Research to Innovation", is a Professor of Electrical and Computer Engineering at the Uni-

versity of Cyprus. In his keynote, he explored how silicon-based Key Enabling Technologies (KETs) have led to the development of compact devices that can sense, process, and respond to signals from the body and the environment. He discussed how these technologies progress from research labs to real-world innovations that enhance quality of life, with examples from medical devices and reconfigurable metasurfaces, while also addressing the challenges of translating research into commercial applications.

Professor Dr. Ing. Dennis Kolberg, Ph.D., who presented the topic "Approaches in Software Startups to Foster Innovative Products", is a professor and head of the Institute for Entrepreneurship and Business Development at the Technical University of Applied Sciences Lübeck (Germany). In his keynote, he highlighted how startups, despite limited resources, often outperform larger companies in delivering disruptive innovations. He explored the unique nature of product management in startup environments, the use of modern innovation practices, and shared early research findings on a startup-specific methodological framework and the evolving role of product managers in the era of AI-driven products.

Multiple IT companies and agencies as FOI partners have actively participated in both the conference program and post-conference events, by organizing workshops and joining panels. This collaboration demonstrates the conference's strong connection with the IT sector and promotes a crucial exchange of knowledge and experience. This synergy and cooperation between the business sector and the scientific community is key to adapting to change and overcoming challenges in IT and beyond. Marko Mihić, a full professor from the University of Belgrade, Faculty of Organizational Sciences, held a guest lecture within the project Digitalization in entrepreneurship in the context of response to an exogenous shock: Drivers, barriers and impact on firm resilience (DigEntre) on the topic "Digitalization and Digital Transformation of the Economy: Two Sides of the Same Coin". In addition, Oracle d.o.o. held a thematic workshop titled "Artificial Intelligence and Data Analytics as Support for Innovation and Business Decision-Making." During the session, participants had the opportunity to explore some of the latest advancements in Oracle's data analytics platforms and learn how artificial intelligence is shaping the future of decision-making across all industries.

In addition to the scholarly sessions, the conference featured two sections dedicated to posters and presentations, which included 2 scientific research posters and 20 student abstracts, collectively involving 53 authors from 6 nations. These posters added depth and diversity to the intellectual landscape of the conference. The student section was further enriched by two invited lectures. Professor Machdel Matthee from the University of Pretoria delivered a talk on "Digital Disruption, Digital Innovation, and 21st Century Skills," while Professor Jin Eun Yoo from the Korea National University of Education presented on "Interpretable Machine Learning for the Social Sciences: Revisiting Penalized Regression in Large-Scale Analysis".

A particularly engaging component of the program was the panel discussion entitled "Smart industry Way Forward", moderated by Professor Neven Vrček, Ph.D., from the University of Zagreb Faculty of Organization and Informatics. The panel brought together experts who shared their experiences and best practices in the field of smart industry. They discussed digital transformation, the application of AI and IoT technologies, and strategies that have shaped sustainable development and the future direction of the industry. As part of the event, the Cracker Barrel Session served as an informal, collaborative forum aimed at enhancing decision-making agility in the digital transformation of education within our higher education institution, with a specific focus on integrating AI.

We extend our sincere thanks to all contributors to the Conference program and these Proceedings. This encompasses our esteemed keynote speakers, panelists, authors, reviewers, session chairs, workshop organizers as well as the Program and Organizing Committee members, whose dedication was instrumental. The formidable support rendered by the University of Zagreb Faculty of Organization and Informatics underpins the unequivocal success of the conference.

Lastly, profound appreciation is conveyed to our respected academic and industrial partners, sponsors and donors, whose unwavering support made CECIIS 2025 a resounding achievement. Collaboration with the University of Zagreb Faculty of Transport and Traffic Sciences, University of Zagreb Faculty of Electrical Engineering and Computing, Josip Juraj Strossmayer University of Osijek Faculty of Electrical Engineering, Computer Science and Information Technology, University of Rijeka Faculty of Informatics and Digital Technologies and École supérieure d'informatique, électronique, automatique (ESIEA) was central to the realization of the conference. The financial support from sponsors and donors including Oracle d.o.o., Infodom d.o.o., Financijska agencija d.d. Aircash d.o.o.,

SICK Mobilisis d.o.o., Grad Novi Marof, Hrvatska banka za obnovu i razvitak, Grad Varaždin, Varaždinska županija, Factory X d.o.o., Jadranski naftovod d.d., Grad Čakovec, Međimurska županija, Grad Prelog, Grad Mursko Središće, Stolarija i pilana Ratković, was also invaluable.

In conclusion, it is our sincerest hope that CECIIS 2025 provided the participants with enriching and productive experience, and that they will find relevant and interesting papers encompassed in these Proceedings. The enduring connections forged with participants over the years reinforce our commitment to continuous improvement. Anticipation is already building for the forthcoming Central European Conference on Information and Intelligent Systems slated for the autumn of 2026.

September 2025

On the behalf of the Program and Organizing Committees
Full Prof. Nina Begičević Ređep, Ph.D. and Marijana Bubanić, Ph.D
Editors

In Memoriam – Prof. Emer. Boris Aurer (1931—2023)



The University of Zagreb Faculty of Organization and Informatics (FOI) is proud of numerous distinguished individuals who have contributed to the public and scientific life of the institution and the community. On this occasion, we remember prof. emer. Boris Aurer, a long-time president and honorary member of the program committee of the international scientific conference Central European Conference on Information and Intelligent Systems (CECIIS).

Prof. emer. Boris Aurer was born in 1931 in Zagreb. He graduated from the Faculty of Electrical Engineering University of Zagreb in 1957 and received his doctorate in 1972. At the same faculty, he was elected a research associate in 1973, and a senior research associate in 1982. At the Faculty of Science in Zagreb, he was elected an associate professor in 1982, and a scientific research advisor in the field of information sciences in 1989. At the FOI, he was elected a full professor in 1992. He was elected to the permanent scientific and teaching title of full professor in the social scientific and teaching field, the scientific field of information science, in 1999. After his retirement, the Senate of the University of Zagreb elected him to the honorary title of Professor emeritus in 2002.

The scientific and teaching career of prof. emer. Boris Aurer is exceptionally rich. At the Faculty of Electrical Engineering University of Zagreb, he worked as an assistant in the Department of Telecommunications (1959—1962), in the central laboratory of the Siemens company in Munich, he worked on the development of computer systems (1963—1970), in the Institute (IETA) of the RIZ factory in Zagreb, he worked as a head of PCM device development (1970—1973), in the University Computing Center (SRCE) in Zagreb, he worked as a researcher on an international scientific project, then as an assistant director for the computing center and as head of the department for research, teaching and development (1973—1991). He has been working at the FOI since 1991 as an associate professor, and since 1992 as a full professor. He has served as Dean of the FOI for two terms (1995—1997 and 1997—1999).

During his distinguished career, Professor Emeritus Boris Aurer enhanced his professional and scientific skills through a period of advanced study at the Technical University of Delft in the Netherlands (1960—1961), on a professional visit to the development departments and computing centers of the RCA company in the USA (1968), and on a study trip to the USA where he studied their development and computing centers (1983). He participated in international projects of the European Economic Community, known as COST-11 (1973—1987). He was a member of the executive board of the European organization for academic research networks RARE. As the project manager of the Ministry of Science and Technology's project "Architectures and Systems Models for Computer Communications" (1991—1994), he made a significant contribution to the field. He was a member of numerous commissions, working groups and boards.

During his teaching career, he selflessly shared his knowledge with undergraduate and postgraduate students, colleagues and associates at the FOI and other higher education institu-

tions in Croatia. At the FOI, he taught the subjects Information Theory, Computer Systems Architecture and Computer Structure at undergraduate level, and Information Systems Theory and Distributed Systems at postgraduate and doctoral levels. His main areas of research were communication systems and computer networks, as well as security systems in networks and security codes for error detection and correction.

Prof. emer. Boris Aurer, with his commitment, expertise, enthusiasm and intellectual curiosity, influenced the establishment of the CECIIS conference at the international scientific level. He was the president of the program committee of the international symposium Information and communication technologies in office business = Office systems (1994), the international symposium Information Systems, IS (1995—1997), the international conference Information and Intelligent Systems, IIS (1998, 2002—2007), the international conference Central European Conference on Information and Intelligent Systems, CECIIS, (2008—2010) and an honorary member of the program committee of the conference CECIIS (2017—2022).

In addition to the conference, prof. emer. Boris Aurer also influenced the development of the journal of the FOI („Zbornik radova“) which over the years grew into today's international scientific journal, Journal of information and organizational sciences (JIOS). He was the president of the editorial board of the journal (1994—1997), a member of the editorial board (1998—2020), editor-in-chief (1998—1999) and guest editor (2002 (Vol. 26, No.1-2)—2003 (Vol. 27, No. 2)).

For his work, prof. emer. Boris Aurer received the recognition and plaque of the Croatian Information Society (1990), the recognition and plaque of the University of Zagreb for a distinguished professor (1994), the recognition of the Croatian Information Society (1996) and the distinguished of the Order of Danica Hrvatska with the image of Ruđer Bošković (1996).

The work and achievements of prof. emer. Boris Aurer have left a lasting mark on the Faculty of Organization and Informatics as well as in the university, national and international scientific community. Thanks to Professor Aurer's vision of establishing our conference as internationally recognized and a leader in information sciences, the CECIIS conference achieved this status and has maintained its 36-year tradition.

We will remember him as a respected professor, scientist, colleague, and friend.

Bernarda Kos, Master of Informatics, Senior Librarian
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, Varaždin

Computer Games

Markus Schatten, Jakov Malović, Ivan Barišić, Antonio Martinaga, Jakov Marčan, Sven Pavlić, Petar Tonković
An AI Enhanced Game Master Assistant for Table Top Role Playing Games

Markus Schatten, Lara Erdogan, Lana Maček, Tin Muhar, Marin Vabec, Josip Novak
Development of a 3D Platformer Game for a Pseudo-Holographic Gaming Console

Bogdan Okreša Đurić, Tomislav Peharda, Hrvoje Zema
Influencing the Output of a Large Language Model using Defined Personality Trait Facets

Tomislav Peharda, Bogdan Okreša Đurić
Generative Worldcrafting: A Modular Framework for AI-Assisted Content Creation in Games

Marinela Schatten, Ljiljana Inkret-Martinčević, Martina Košćak
Play It Safe: Evaluating Participatory Game-Based Learning Activities for Internet Safety Among Primary and Secondary School Students

An AI Enhanced Game Master Assistant for Table Top Role Playing Games

Markus Schatten, Jakov Malović, Ivan Barišić, Antonio Martinaga, Jakov Marčan, Sven Pavlić, Petar Tonković
Artificial Intelligence Laboratory
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
markus.schatten@foi.unizg.hr,
{jmalovic22,ibarisic21,amartinag22,jmarcan22,spavlic22,ptonkovic21}@student.foi.hr

Abstract. *In this work-in-progress paper, we present the initial implementation of a game master (GM) assistant for Table Top Role Playing Games (TTRPGs) that leverages artificial intelligence (AI), specifically Large Language Models (LLMs), to generate narrative content, non-player characters (NPCs), items, and other game-related material based on GM input. Developed using a design science research methodology, the system is implemented in Python and Ren'Py and integrates the OpenAI application programming interface (API). It enables the GM to display ambient sounds, music, and scene images with dynamically inserted player characters and NPCs. We report on the initial implementation and pilot evaluation results.*

Keywords. game master, table top role playing game, artificial intelligence, large language models, non player character, application programming interface

1 Introduction

The global table top game market has been valued at almost €12 billion in 2023 and is projected to grow up to more than €29 billion in 2032 according to (Fortune Business Insights, 2024, June). The total revenue for role-playing games (RPGs) (which includes computer RPGs) in 2022 is valued at more then €42 billion and is expected to grow up to €66 billion by 2027 (Statista, 2024). Table Top Role Playing Game (TTRPG) games like Dungeons & Dragons (D&D) have gained major popularity in recent years. For example Hasbro, Inc. owner of Wizards of the Coast which produce D&D related content claim that over 50 million people are playing it worldwide (Wizards of the Coast, 2022, August).

TTRPGs are a form of collaborative storytelling and game play in which participants assume the roles of fictional characters within a shared narrative framework. Unlike board games with rigid rule structures and defined win conditions, TTRPGs emphasize improvisation, imagination, and collective decision-making. Gameplay is typically facilitated by game master (GM), who describes settings, controls non-player characters (NPCs), and adjudicates rules and outcomes of player actions.

Players create and control player characters (PCs), each defined by a set of attributes, abilities, and backstories that inform their interactions with the game world. Sessions unfold as a series of descriptive exchanges, dice rolls, and collaborative problem-solving activities that advance the plot and develop character arcs.

Classic examples of TTRPGs include *Dungeons & Dragons*, *Pathfinder*, and *Call of Cthulhu*, but thousands of systems and settings exist, ranging from high fantasy to science fiction and horror. TTRPGs have gained widespread popularity due to their adaptability, potential for deep role-play, and the strong social bonds they foster among players.

Recent years have seen a resurgence of interest in TTRPGs, driven by live-streamed campaigns, podcasts, and digital tools that enhance play, attract new audiences, and lower barriers to entry. At their core, TTRPGs remain a unique medium combining structured mechanics with open-ended creativity.

In usual TTRPGs a dungeon master (DM), GM, master of ceremonies (MC) or other ceremonially titled narrator which runs the game has to prepare numerous aspects before a games session starts. Such aspects might include the setting (including worldbuilding and lore), story, quests, NPCs, mobs, dialogue, battles, various props like items, maps, figurines, background music and sound effects, balancing personal computers (PCs) statistics, interviewing players, reviewing their character sheets and character art etc. (Page, 2024). In addition to this, for educative TTRPGs an educator has to include and incorporate educative materials into various aspects of the game to achieve the desired learning outcomes. There seems to be no *one size fits it all* approach in designing educative gaming experiences (Khaldi et al., 2023) which is a motivation for us to try to develop a different approach.

In principle, there are two aspects of the prepara-

tion and gaming process of a TTRPG that we would like to support: (1) *theater of the mind*[1] that includes the preparation of descriptions (for example of places, NPCs, lore, items, etc.), dialogue, rulesets, game art, and (in the context of educative games) learning materials (including lesson plan, learning outcomes, exercises etc.); as well as (2) *props* which include all physical items such as player/NPC/mob figurines, miniatures, dice, letters, maps, coins, or artifacts used during gameplay to enhance immersion and provide tangible elements that players can interact with. Herein we propose to use artificial intelligence (AI) or more precisely generative artificial intelligence (GenAI) for the former and pseudo-holography supported by AI for the latter.

GenAI is transforming digital media industries (Markus Schatten, 2024). The possibility of generating text (including programming code) see Yao et al., 2024, for an overview, visual images see Zhang et al., 2023, video see Xing et al., 2023, audio see Latif et al., 2023 and 3D models see Liao et al., 2024 has already incited major players in the media industry to experiment with this breakthrough technology. For example, Marvel Studios has used GenAI to create the opening credits of their series *Secret Invasion* starring Samuel L. Jackson (Sharf, 2023), AAA game studios including Activision Blizzard are reportedly already using AI for game creation (Serin, 2024), and Keyword Studios is creating their center of excellence for the responsible use of GenAI in game development (Wood and Peacock, 2024, March).

In order to support TTRPG play we have implemented a GM assistant tool that allows for generating textual content including but not limited to game narrative, NPC names, descriptions, or dialogue, as well as items and other game related content. In addition to generated content the implemented system allows for import of prepared graphics like setting backgrounds, player characters, NPCs as well as sound effects and background music which can be dynamically shown during game play. Herein we report on the development process and initial results.

The rest of this paper is organized as follows: in section 2 we provide an overview od the development proces based on design science. In section 3 we discuss our findings and in section 4 we draw our conclusions and provide guidelines for future research.

2 Development Process

To develop the described system, we adopted a design science approach. Design science is a research methodology in information systems that aims to create novel artifacts—such as models, methods, and systems—to support people in solving complex problems (Johannesson and Perjons, 2014). This methodology is particularly suitable for developing systems combining various approaches and technologies like artificial intelligence, user interfaces, and content generation as has been showcased in (Markus Schatten and Sačarić, 2024, in press; Markus Schatten, Marinela Schatten, and Inkret Martinčević, 2023).

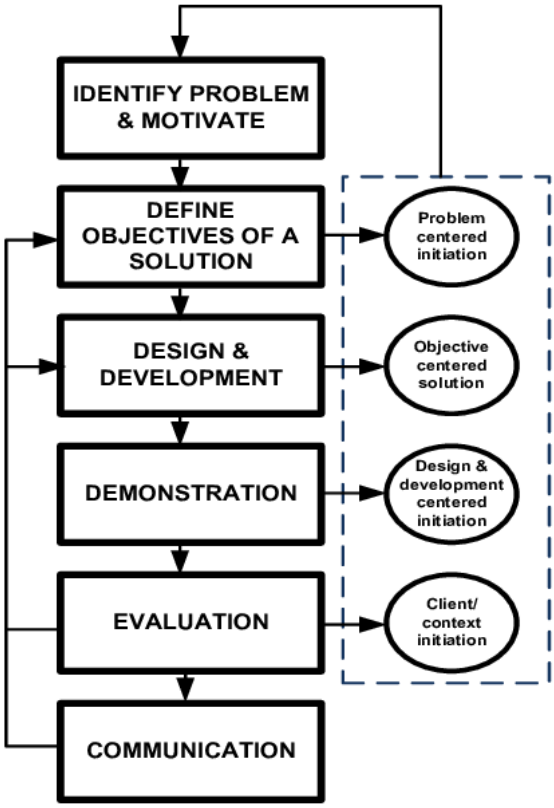


Figure 1. Design science methodology (Azasoo and Boateng, 2015) adapted from (Peffers et al., 2007)

Design science research typically involves a series of steps which may be iteratively revisited. Below, we outline each step and describe how it was applied in the development of our TTRPG assistant:

1. **Problem Identification and Motivation:**

   The identified problem is the lack of tools to support GMs in dynamically generating high-quality narrative content, characters, and settings during live TTRPG sessions. Traditional preparation methods require significant manual effort and time investment, limiting spontaneity and improvisation. With the growing capabilities of Large Language Models (LLMs), there exists an opportunity to augment GMs' creativity through AI assistance.

   The motivation for solving this problem stems from the need to enhance player immersion, reduce GM workload, and improve the accessibility of storytelling tools for less experienced GMs.

[1] In the context of TTRPGs like D&D, "theater of the mind" refers to a style of gameplay where players and the game master use verbal descriptions and imagination to create and visualize the game world, actions, and events, rather than relying on physical maps or miniatures.

Player immersion, aside from using on-the-fly generated content can be additionally enhanced by including prepared graphics (like backgrounds, player character depictions, NPCs, items etc.) and sound (like background music or sound effects) which can be used during a gameplay session.

2. **Define the Objectives for a Solution:**

   Based on this problem, we defined the following objectives:

   (a) *Dynamic content generation.* Enable GMs to generate character backstories, locations, and narrative events during gameplay in real time.

   (b) *Ease of use.* Develop an intuitive, minimalistic interface requiring no advanced technical skills.

   (c) *Content consistency.* Ensure reproducibility and coherence of generated content across sessions.

   (d) *Multimedia integration.* Support visuals and audio to enrich immersion.

   (e) *Cross-platform demonstration.* Deliver a publicly accessible application demonstrating the artifact's functionality.

3. **Design and Development:**

   The artifact was designed and implemented through iterative prototyping. We selected Python as the main programming language due to its rich ecosystem of libraries for AI integration and cross-platform graphical user interface (GUI) development.

   The development process comprised the following phases:

   (a) *Technology selection.* We adopted the OpenAI application programming interface (API) for natural language generation, specifically the GPT-3.5 Turbo model, balancing performance and cost. For user interfaces, we utilized the `tkinter` library for the main GUI, `pygame` for audio playback, and `PIL` for image manipulation. The RenPy engine was integrated to display visual novel-style scenes.

   (b) *API integration.* The system prompt, which defines the AI's role and output style, was hardcoded for consistency but designed to be stored externally for future iterations. Temperature and other model parameters were fine-tuned to balance creativity with predictability. Conversation history management was implemented to maintain narrative coherence.

   (c) *User interface design.* We designed three core interfaces:

   - A setup screen to load characters, backgrounds, music, and NPCs.
   - The RenPy-powered narrative display.
   - An AI chat interface styled as a "magic mirror" to enhance immersion.

   The UI emphasized high contrast, clear grouping of functions, and minimalistic aesthetics to reduce cognitive load.

   (d) *Multimedia elements.* Frames and decorative elements were generated using DALL-E and processed in GIMP. A system for automatic refreshing of the Ren'Py game engine was implemented to instantly display newly added content without restarting the application.

   An overview of the system's architecture is shown in Fig. 2.

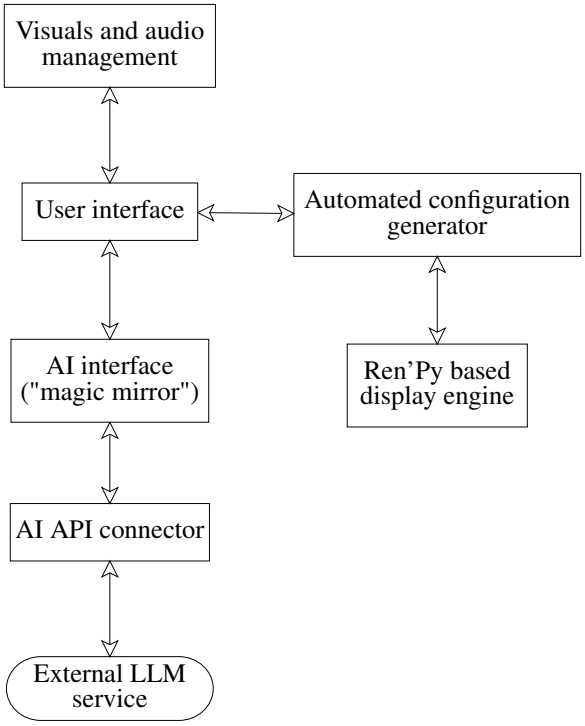


Figure 2. System architecture overview

4. **Demonstration:**

   The artifact was deployed as a public demonstration on the itch.io platform, accessible at:

   https://ailab-foi.itch.io/prri-ttrpg2025

   This allowed GMs and players to experience the assistant in realistic gameplay scenarios. Fig. 3 shows example screenshots from the system.



Figure 3. Screenshots from the TTRPG game master assistant (above: main interface, bellow: AI interface - "magic mirror"

5. **Evaluation:**

   An initial pilot evaluation was conducted through four gameplay sessions involving a GM and three players per session. Observations were made regarding usability, output quality, and overall player experience. Feedback collected during and after the sessions included:

   - The assistant significantly reduced preparation time for GMs.
   - Generated narratives were generally coherent and contextually relevant.
   - Minor issues arose with content consistency when context windows exceeded token limits.
   - The immersive visual and audio features were highly appreciated by players.
   - Some users suggested adding more fine-grained control over AI creativity.
   - GMs suggested adding the possibility to generate visuals and audio as well as using different AI services instead of the currently only available ChatGPT system.

   Based on this feedback, recommendations were documented for future development iterations, including externalizing system prompts for easier customization, improving caching mechanisms to handle long dialogues, adding support for different back-end LLMs and other AI services for audio and visuals creation, adding and refining UI interactions.

   Future evaluation will consist of structured sessions with prepared surveys and interviews.

6. **Communication:**

   Beyond this paper, the system and its source code have been published on GitHub under an open source license (available here https://github.com/AILab-FOI/PRRI-TTRPG2025) to enable broader adoption and feedback. The itch.io page provides demonstration downloads, and additional dissemination is planned via conference presentations and workshops focused on AI applications in digital games.

3 Discussion

The initial implementation and evaluation of the AI-enhanced TTRPG game master assistant have yielded several insights relevant both for the development of similar tools and for understanding the opportunities and challenges of integrating generative AI into collaborative storytelling environments.

First, the demonstration and evaluation confirmed the central hypothesis that large language models can substantially reduce the preparation burden for GMs while simultaneously enhancing narrative richness and player immersion. During the four gameplay sessions, GMs consistently reported a significant decrease in the time required to prepare non-player characters, narrative hooks, and descriptive content. This aligns with similar findings in adjacent domains such as AI-supported educational games (Horn and Göbel, 2024) and digital interactive fiction (Hua and Raley, 2020), where generative AI has been shown to improve both efficiency and perceived content quality.

However, several limitations were observed. One recurring issue concerned the management of long-running dialogues. Specifically, when the conversation history grew beyond the context window supported by the GPT-3.5 Turbo model, players noted occasional inconsistencies or loss of relevant context in generated responses. This limitation highlights the importance of designing effective memory and summarization strategies to maintain coherence over extended interactions. Future iterations of the system should explore segmenting sessions or employing retrieval-augmented generation to mitigate these issues.

Another important observation relates to content creativity and control. While most GMs appreciated the assistant's ability to generate novel and engaging content, some expressed the desire for more granular adjustment of output creativity (e.g., dynamically controlling temperature settings or style parameters). This

reflects a broader trade-off between automation and creative direction in generative tools. Providing intuitive controls that allow GMs to calibrate the AI's output to match their narrative intentions will be an important focus in future development.

Regarding the multimedia components, players and GMs consistently highlighted that the inclusion of images, ambient sounds, and background music contributed positively to immersion. The integration with RenPy and the modular design for importing assets were perceived as strengths of the system. Nonetheless, several participants suggested expanding the assistant's capabilities to generate or recommend multimedia assets dynamically, potentially through the integration of additional AI services for text-to-image and text-to-audio generation. This direction would further reduce preparation time and provide a more seamless experience.

From a usability perspective, the minimalist interface design and clear grouping of functions were largely successful in reducing cognitive load and facilitating adoption. New users required minimal instruction to begin generating content and managing multimedia elements. However, several participants noted that more robust error handling and guidance would be beneficial, particularly when API connectivity issues or input validation errors occur.

Finally, the public demonstration of the system on the itch.io platform and open-source publication of the source code have already begun to generate interest and feedback from the broader community of TTRPG enthusiasts and developers. This community engagement is expected to be an important catalyst for iterative improvement and adaptation of the system to diverse game systems and play styles.

Overall, the findings indicate that AI-powered assistants hold considerable promise for augmenting TTRPG play, but also underscore the need for careful design of context management, user control, and multimodal content integration. The lessons learned from this initial implementation will inform future research and development aimed at refining both the technical foundations and user experience of AI-assisted game mastering.

4 Conclusion and Future Research

This paper has presented the design, development, and preliminary evaluation of an AI-enhanced game master assistant for tabletop role-playing games. By combining large language models, an implemented GUI, the Ren'Py game engine, and multimedia integration, the implemented system demonstrates the potential of generative artificial intelligence to augment the creative process of game mastering, reduce preparation workload, and enrich the immersive experience for players.

The results of four initial pilot gameplay sessions suggest that the assistant effectively supports dynamic content generation and narrative consistency, while the modular architecture facilitates the inclusion of preprepared visual and audio assets. User feedback highlighted the strengths of the approach, particularly regarding ease of use and enhanced immersion, while also identifying areas for improvement, such as better context management for longer dialogues, finer control over creativity parameters, and expanded capabilities to recommend or generate multimedia content.

Future research and development will focus on several directions. One priority is exploring the integration of additional generative models capable of producing diverse types of content beyond text, including images, music, ambient soundscapes, video, and animation. This multimodal generation could enable GMs to create fully customized experiences with minimal manual effort.

Additionally, we aim to investigate the application of the assistant in educational contexts, where TTRPGs are increasingly used as a vehicle for teaching and learning. Research will examine how AI-generated narratives and multimedia can be adapted to align with pedagogical goals, curriculum requirements, and differentiated learning needs. Controlled studies will be conducted to assess the impact of such tools on learner engagement, knowledge retention, and the development of transversal skills such as collaboration and problem solving.

Taken together, this work contributes to the growing body of evidence that generative AI can play a transformative role in digital and analogue game development. By continuing to refine technical capabilities and expand practical applications, we hope to advance the field toward more accessible, engaging, and creative TTRPG experiences for both recreational and educational settings.

Acknowledgement

The authors acknowledge that ChatGPT 4o by OpenAI has been used to assist in editing and streamlining sections of this paper. Additionally, DALL-E by OpenAI had been used to generate graphics for the assistant's user interface as well as test graphics of background and characters.

References

Azasoo, J., & Boateng, K. (2015). A retrofit design science methodology for smart metering design in developing countries. https://doi.org/10.1109/ICCSA.2015.23

Fortune Business Insights. (2024, June). *Board games market size, share & industry analysis, by game type (monopoly, scrabble, chess, and others), by age group (2-5 years, between 5 and 12 years,*

12-25 years, and above 25 years), by sales channel (online stores, specialty stores, hypermarkets & supermarkets, and others), and regional forecast, 2024-2032 (tech. rep.) (Report ID: FBI104972). Fortune Business Insights.

Horn, F., & Göbel, S. (2024). Ai as a co-creator: A survey on ai support for educational game authoring tools. *Joint International Conference on Serious Games*, 3–18.

Hua, M., & Raley, R. (2020). Playing with unicorns: Ai dungeon and citizen nlp. *DHQ: Digital Humanities Quarterly*, *14*(4).

Johannesson, P., & Perjons, E. (2014). *An introduction to design science* (Vol. 10). Springer.

Khaldi, A., Bouzidi, R., & Nader, F. (2023). Gamification of e-learning in higher education: A systematic literature review. *Smart Learning Environments*, *10*(10), 1–31. https://doi.org/10.1186/s40561-023-00227-z

Latif, S., Cuayáhuitl, H., Pervez, F., Shamshad, F., Ali, H. S., & Cambria, E. (2023). A survey on deep reinforcement learning for audio-based applications. *Artificial Intelligence Review*, *56*(3), 2193–2240.

Liao, J. Z. Z. L. J., Cao, Y.-P., & Shan, Y. (2024). Advances in 3d generation: A survey. *arXiv preprint arXiv:2401.17807*.

Page, C. A. (2024). Role-ing the dice: Identity, intimacy, and consent in table-top role playing [Licensed under a Creative Commons Attribution-NonCommercial 4.0 International License]. *Journal of Consent-Based Performance*, *3*(1), 63–86.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45–77.

Schatten, M. [Markus]. (2024). Ai and the future of entertainment technology [Preprint submitted on 7 Jul 2024. Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License. This work has been supported in full by the Croatian Science Foundation under the project number IP-2019-04-5824.], (HAL-04637685).

Schatten, M. [Markus], & Sačarić, S. (2024, in press). Integrating cognitive agents into visual novels. *Central European Conference on Information and Intelligent Systems*.

Schatten, M. [Markus], Schatten, M. [Marinela], & Inkret Martinčević, L. (2023). Development of gamified cognitive agents for primary schools-a case study. *34th Central European Conference on Information and Intelligent Systems (CECIIS 2023)*, 3–8.

Serin, K. (2024). Activision blizzard is reportedly already making games with ai, and quietly sold an ai-generated microtransaction in call of duty: Modern warfare 3 [Accessed: 2024-07-26]. *GamesRadar+*.

Sharf, Z. (2023). Marvel used ai to create 'secret invasion' opening credits, ep says it fits the 'shape-shifting' plot [Accessed: 2024-07-26]. *Variety*.

Statista. (2024). Role playing games - worldwide [Accessed: 2024-07-12].

Wizards of the Coast. (2022, August). Wizards of the coast reveals exciting lineup for dungeons & dragons and magic: The gathering [Accessed: 2024-07-12].

Wood, L., & Peacock, S. (2024, March). Unravelling project ava - research insights on gen ai in game development [Keywords Studios].

Xing, Z., Feng, Q., Chen, H., Dai, Q., Hu, H., Xu, H., Wu, Z., & Jiang, Y.-G. (2023). A survey on video diffusion models. *arXiv preprint arXiv:2310.10647*.

Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z., & Zhang, Y. (2024). A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. *High-Confidence Computing*, 100211.

Zhang, C., Zhang, C., Zhang, M., & Kweon, I. S. (2023). Text-to-image diffusion model in generative ai: A survey. *arXiv preprint arXiv:2303.07909*.

Development of a 3D Platformer Game for a Pseudo-Holographic Gaming Console

Markus Schatten, Lara Erdogan, Lana Maček, Tin Muhar, Marin Vabec, Josip Novak
University of Zagreb Faculty of Organization and Informatics
Artificial Intelligence Laboratory
Pavlinska 2, 42000 Varaždin, Croatia
markus.schatten@foi.unizg.hr, lara.erdogan@student.kit.edu,
{lmacek21,tmuhar22,mvabec22,jnovak22}@student.foi.hr

Abstract. *T*his work-in-progress paper presents the design and implementation of a 3D platformer game developed for a pseudo-holographic gaming console utilizing a Pepper's Ghost pyramid display. The project applies a Design Science Research (DSR) methodology to address the underexplored potential of holographic projection technologies in gaming contexts. The system integrates custom hardware, including a bespoke cabinet and holographic pyramid, and a software solution built using UPBGE, Blender, and Python. Core game features encompass player navigation across a cubic environment projected holographically, collision detection, and a mobile device interface for control. The artifact was demonstrated through live evaluations and online deployment, revealing key insights into visual design optimization and user interaction constraints unique to the holographic medium. The results underscore the viability of leveraging Pepper's Ghost displays for interactive entertainment and highlight opportunities for further research in novel game mechanics and user experience refinement.
Keywords. computer games, 3D models, Pepper's Ghost pyramid, UPBGE, holographic gaming, design science research

1 Introduction

Gaming consoles are specialized computing devices designed for video gaming. These devices generally fall into two categories: those that require an external display, such as a television, to showcase the game content, and those that are versatile and can function independently with their own display. Gaming consoles which are designed to connect to an external display are usually equipped with high-performance hardware, allowing them to render graphically-intensive games. On the other hand, consoles other consoles are more versatile and have built-in displays, allowing them to function independently, but they can also be connected to an external display for a more immersive gaming experience.

In addition to these dedicated gaming devices, the proliferation of smartphones and tablets has led to an expansion in mobile gaming. These devices feature powerful processors and high-resolution touch-screen displays, allowing them to deliver a compelling gaming experience. Despite their omnipresence, current gaming technology does not utilize holography nor mobile devices as controllers.

Holography, while used in some niche applications and experimental settings, is not a part of standard gaming console technology. The creation of a holographic image is still a complex process that might involve using a panel display as the light source for the holographic image.

A hologram is 3 dimensional visual display that displays objects visible from all angles. At the time of writing there are a number of promising holographic technologies available (Blanche, 2022; Chang et al., 2018; Fadzli et al., 2022; Haleem et al., 2022; Z. Li et al., 2022; Piliar, n.d.; Wang et al., 2022). From a commercial perspective, there are various technologies that will produce a similar effect.

Whilst real holography devices are still an unreached dream for industry (Blanche, 2021), pseudo-holography technologies have matured up to a point in which ready-to-use products are available of the shelf. Technologies like Pepper's Ghost displays (for example Burdekin, 2015), various volumetric displays (see for example Fokides and Kilintari, 2023; Refai, 2009, and others), transparent organic light-emitting diode (OLED) screens for example Tsai et al., 2017, as well as holographic meshes like hologauze (for example Weinel, 2020) are used for in-store marketing, exhibitions, shows and games. artificial intelligence (AI) and holography have the possibility to "*change the nature of the interaction between learners and learning materials*" (X. Li and Taber, 2022).

One such technology are so called Pepper's Ghost images, after the English scientist John Henry Pepper (1821–1900). The basic idea behind this technology is to use a light display that emits an image on a reflective (usually transparent) surface which reflects the image and provides a holographic experience. This

technique, has been used for centuries and involves using plate glass, Plexiglas, or plastic film and special lighting techniques to make objects seem to appear or disappear, become transparent, or to make one object morph into another Crowder and Jackson, 2016. An enhancement of the original system is to use a glass or acrylic four sided pyramid instead of just one surface, and place a display above or bellow the pyramid so that the holographic image is visible from all four sides of the pyramid.

Such pyramidal holographic displays have not been widely adopted in gaming technology. Similarly, while mobile devices have been used as controllers in certain contexts, they are not a standard control method for mainstream consoles.

As part of the O_HAI (4) Games project (*Orchestration of Hybrid Artificial Intelligence Methods for Computer Game* sponsored by the Croatian Science Foundation) one such console has been developed as reported in (Schatten, 2023). Herein we report on the development of a 3D platformer game for this console that has been developed at the Artificial Intelligence Laboratory of the University of Zagreb Faculty of Organization and Informatics.

The rest of this paper is organized as follows: firstly in section 2 we introduce the pseudo-holographic gaming console HoloGameP. In section 3 we discuss the development process of the game in detail. In section 4 we draw our final conclusions and showcase directions of future research.

2 HoloGameP - A Pseudo-Holographic Gaming Console

To build the physical hardware, we first performed preliminary calculations to determine the appropriate dimensions for the plexiglass pyramid. Based on these measurements, we designed the remainder of the cabinet, taking into account available display sizes and ensuring the console would be comfortable for users to view at eye level. A detailed Computer Aided Design (CAD) model was created in FreeCAD[1], which is included in the appendix of this document.

Using this model as a reference, a cabinet combining wooden and metallic elements, as shown in Fig. 1 has been built. The display is installed within the upper metal compartment, with the plexiglass pyramid positioned directly beneath it. The lower wooden section houses all the hardware components, including a gaming PC, a router, and peripheral devices such as a mouse and keyboard.

The software required to operate the HoloGameP console comprises three main components (outlined in Schatten, Peharda, et al., 2022): (1) a streaming system, (2) a game engine with camera transformation capabilities, and (3) an implementation of the controller application programming interface (API). The console can function either by streaming content or running a locally installed game. Testing indicated that the local installation delivers significantly better performance due to network latency constraints. Therefore, in this work, we focus on developing proof-of-concept implementations for the local mode.

[1] Available at https://www.freecad.org/



Figure 1. HoloGameP Cabinet

3 Development Process

To develop a gaming experience based on a Pepper's Ghost pyramid, we employed a design science research (DSR) approach. DSR is a research methodology within information systems and technology that focuses on the development of innovative artifacts such as models, methods, systems and software. These artifacts are intended to help individuals and research teams design, implement, and manage IT solutions (Johannesson and Perjons, 2014). Beyond information systems and technology, design science is also applied across diverse areas of human practice.

Design science methodology (refer to Fig. 2) generally comprises a series of principal stages which are not strictly linear and may be revisited iteratively throughout the course of a DSR project. Below, we describe these stages and outline their application in the development of our holographic gaming system.

3.1 Problem Identification and Motivation

This stage entails recognizing a specific problem or opportunity and establishing its significance. A comprehensive understanding of the problem domain is essen-

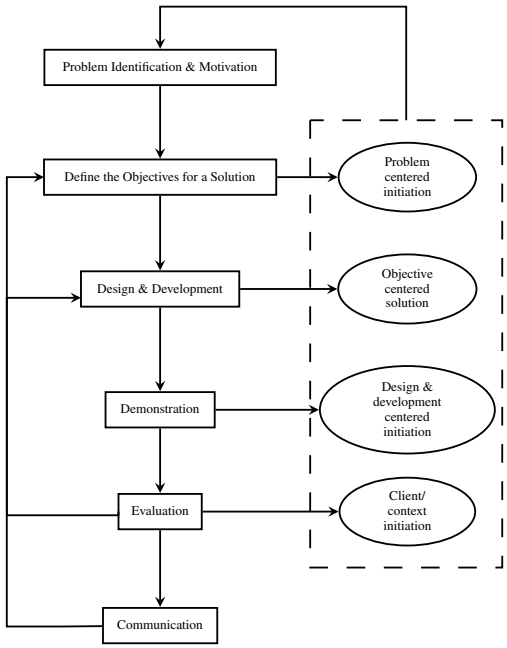


Figure 2. Design science methodology (Azasoo and Boateng, 2015) adapted from (Peffers et al., 2007)

tial, as is the articulation of the motivation for addressing the challenge.

In this context, problem identification involves discerning a relevant issue and situating it within a broader framework. It is critical to define the motivation not solely in terms of a technical resolution but also with respect to its potential impact and contextual relevance.

In our case, the problem addressed stems from the underutilization and limited public understanding of holographic display technologies, particularly Pepper's Ghost–based systems, in gaming and interactive settings.

While holography has seen increased use in mostly advertising, its integration into proper gaming environments remains rare and underexplored. The motivation behind this project is to bridge that gap by developing an engaging, interactive holographic platformer game specifically designed for a large, custom-built pyramid display installed at our campus building for students and visitors to interact with.

The aim is twofold: to highlight the creative and technical capabilities of Pepper's Ghost holography and to demonstrate how such technology can support innovative forms of digital interaction in the gaming industry, as well as educational settings. By combining game design with holographic projection, the project offers an immersive experience that appeals to a wide range of users from non-gamers trying out new technology, to hardcore players looking for a new challenge, thereby fostering curiosity about both the underlying technology and the possibilities it unlocks for future applications in entertainment, visualization, and other settings.

3.2 Define the Objectives for a Solution

Building upon the previously identified problem, this stage involves formulating clear objectives for the proposed solution. These objectives must be both feasible and explicitly aligned with addressing the identified challenge.

The principal objectives guiding the development of our holographic platformer are as follows:

1. *Demonstrate Entertainment Potential.* The primary goal is to highlight how Pepper's Ghost holography can be used to deliver compelling, interactive gaming experiences. The platformer is designed to entertain and captivate a general audience, demonstrating that holographic displays can be more than passive exhibits. They can support dynamic, responsive digital content. By combining real-time interaction with visual depth and spatial illusion, the game showcases the unique potential of holographic projections to engage users in ways that traditional screens cannot. The use of a pyramid-shaped holographic display allows the game to be experienced from multiple angles, adding a physical presence and novelty factor that enhances immersion. This makes it particularly effective for public installations, exhibitions, or educational environments where grabbing and maintaining audience attention is essential. Through creative design, intuitive gameplay, and responsive mechanics, the project emphasizes that holography is not just a visual trick — it is a viable platform for meaningful and enjoyable digital experiences.

2. *Investigate Novel Game Mechanics Unique to the Medium.* The project aims to explore and evaluate innovative gameplay interactions that are inherently specific to the pseudo-holographic display environment. These mechanics leverage the volumetric appearance, spatial parallax, and multi-angle visibility of the Pepper's Ghost technique to enable experiences that are impractical or unattainable in conventional gaming contexts.

3. *Design for Drop-In Interactivity.* The game is intended to be immediately accessible to walk-by participants, including students, faculty, and visitors. Simple, intuitive controls and quick on boarding make it easy for users to engage with the game without prior instruction or setup.

These objectives guide the design and development process of the game, ensuring that the final product meets all the criteria, works as intended, and, most importantly, sends the right message about the capabilities of this technology.

3.3 Design and Development

During this phase, an artifact is systematically conceived and implemented to fulfill the established objectives. The resulting artifact may take the form of a conceptual model, a methodological framework, or a functional instantiation. The design and development has been conducted by 6 students as part of the Game platforms development course and student internship under the guidance of one university professor.

This phase involves the transformation of initial ideas into a functioning interactive experience tailored to the constraints and opportunities provided by the holographic pyramid display. The development of the game has been started as an intern project and later fully developed during a game development platforms course. The first step was to develop a game development document (GDD) and project plan that will include all phases of the development iteration cycle. The development has been supported by and documented within a Github repository (available at: https://github.com/AILab-FOI/PRRI-HoloGameP2025).

All 3D game assets and models were created using Blender, while the whole game was coded inside Uchronia Project Blender Game Engine (UPBGE), an open-source, 3D game engine deployed using Blender itself, so in a way, this game was developed almost exclusively in Blender, other than the audio and parts of the UI - start and end screens.

The initial vision for the game's visual identity was to create a recognizable and striking steampunk world that would immediately capture players' attention with its characteristic elements (such as gears, pipes, and a small robot). The goal was for every model and environment to clearly reflect that retro-industrial style, while remaining simple and high-contrast enough to display well on the pyramidal hologram platform, where visibility and visual recognition are crucial due to the specific projection method. Inspiration was drawn from classic steampunk motifs, but it was necessary to adapt to gameplay requirements and the technical limitations of the platform, ensuring that every model on the holographic platform contributed to both the atmosphere and the functionality of the game itself.

The unique nature of the pyramidal display using Pepper's Ghost technology demanded particular attention to visual design. We prioritized creating clear, high-contrast elements that remained instantly recognizable from every viewing angle which was a crucial requirement given players physically circle the pyramid during gameplay. To enhance the three-dimensional illusion, a subtle angular offset in level rotation was implemented, allowing players to view approximately one and a half sides simultaneously. This solution effectively overcame the potential flatness of a traditional 2D presentation while maintaining gameplay clarity.

Every detail had to be functional and contribute to clarity, as overly complex or dark elements could hinder gameplay and diminish the player's experience. Additionally, the steampunk theme required carefully selected elements (e.g., gears, pipes, keys, etc.) to create an authentic atmosphere and provide players with a unique visual experience. The final design is featured on Fig. 3.

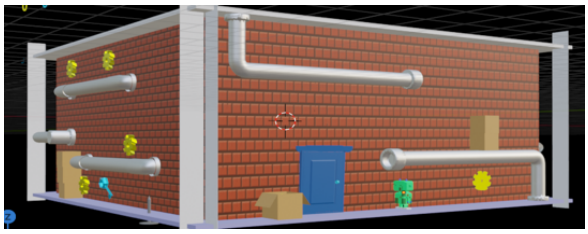


Figure 3. Level design

The main objectives we aimed to achieve with the design of the visual elements were clarity, appeal and uniqueness. Clarity was essential due to the specific display of the game on the pyramidal platform using the Pepper's Ghost effect, where all elements needed to be easily recognizable and visible from different angles. Appeal was achieved through carefully chosen steampunk motifs (gears, pipes, boxes, levers, robot, etc.) that gave the game its distinctive retro-industrial look. Uniqueness was realized by combining classic steampunk elements with our own stylized solutions adapted to the technical limitations of the platform. When deciding on the color palette, shapes, and stylization of characters and environments, the team was guided by the need for high contrast and clear outlines, so that the visuals would remain readable in the hologram display. The color palette included bright, high contrast, metallic tones (brown, bronze, gold, silver, blue, teal), while the shapes were simple and easily recognizable, without unnecessary details that could complicate the display. The robot Mossy and his environment were designed to reflect the steampunk aesthetic, but also to be functional in the context of gameplay. When choosing the visual style, the team kept in mind the target audience: casual players, visitors, steampunk and retro technology enthusiasts, as well as students who would use the game for educational purposes. Therefore, the visual identity was designed to be accessible, interesting, and recognizable to a broad audience, while still specific enough to attract steampunk genre enthusiasts. The high-contrast version employs a vibrant color palette of teal, yellow, orange, and green while preserving the original game elements' design. This enhanced color scheme was specifically developed to maintain visual clarity and gameplay visibility under varying lighting conditions, particularly addressing the challenges posed by different environmental light levels affecting the holographic display. The solution maintains all gameplay functionality while significantly improving the player's ability to distinguish game elements regardless of ambient lighting. Some of the designed assets are shown on figure 4.

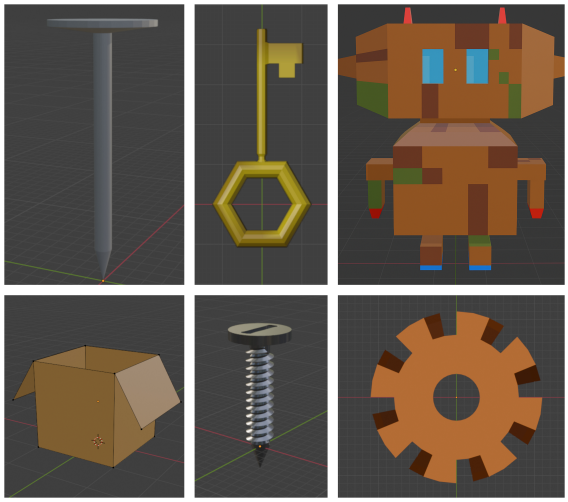


Figure 4. 3D models design

For creating the visual elements in the game, we primarily used Blender, which allowed us to model all key objects (pipes, boxes, and the levels themselves), while basic texture editing was done using simpler tools available in Blender, ensuring that the models were optimized for display on the specific pyramidal platform. The process began with sketching basic shapes and arranging elements within the levels, where we used quick digital sketches and prototypes to test how individual objects would look and function in space, and then turned those concepts into 3D models, focusing on simplicity and readability. During the design process, we closely collaborated with programmers to align the visual elements with the technical requirements of the engine and proper display on the hologram platform, as well as with team members responsible for sound, so that the visual and audio elements would together create a coherent atmosphere. The biggest technical challenges were related to the limitations of displaying on the pyramid, where certain colors and details lost visibility or visuals became overly saturated, so we had to adjust the color palette and simplify shapes to keep elements clear from all angles.

We tackled creative challenges through an iterative process of testing models in actual display conditions and gathering team feedback. This approach allowed us to refine textures, adjust contrasts and modify object sizes until we found an adequate balance between steampunk aesthetics and gameplay functionality. As previously mentioned, to address the transparency issue where players could simultaneously see both sides of the pyramid, we implemented a light gray background in the high contrast model. Key game elements, such as pipes, were given vibrant teal coloring for better visibility. While maintaining authentic steampunk design through predominant use of bronze and gray tones in the player model, collectibles, and some obstacles, we enhanced contrast levels to ensure optimal gameplay visibility in brighter environments. Through this careful balancing, we maintained the artistic vision while overcoming practical display challenges. The contrast adjustments introduced an interesting gameplay element. The obstacles like nails and screws were deliberately colored in darker gray tones to increase the game's difficulty level. This approach solved visibility issues and added an extra layer of challenge to the gameplay experience as shown on Fig. 5

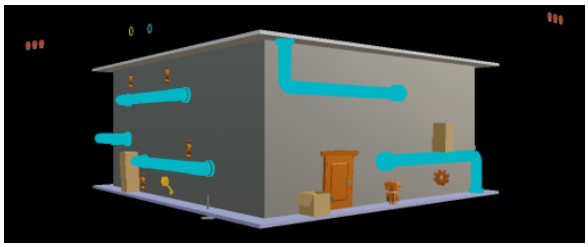


Figure 5. Better contrast level design

The development process consisted mainly of coding and making sure that all 3D models, sound effects, and other assets were properly implemented. The development process had the following phases:

1. *Creation of the Base Game.* As an initial step, it was necessary to develop a prototype serving as a proof of concept. This preliminary implementation comprised only a limited set of box elements to evaluate movement functionality and to configure the scripting responsible for accurately projecting the game onto the display. Establishing correct camera configurations was a critical component, given the specific requirements of the pyramid hologram. Without proper camera alignment, the system would not operate as intended. Four cameras were utilized, each assigned to one of the cardinal directions (north, south, east, west). These cameras were oriented to observe the central game stage from their respective positions, calibrated to ensure that the base platform appeared consistent across all perspectives. Subsequently, the camera outputs were arranged on the display according to the defined script, which positioned each view appropriately. When projected onto the pyramid, the resulting visualization presented a comprehensive overview of the game environment from all four sides. The subsequent objective was to optimize the projection dimensions to achieve the maximum possible height and width on each face of the pyramid, a process that required extensive iterative testing to refine.

2. *Development of Movement Mechanics.* The player character navigates each face of the pyramid in a manner analogous to a traditional 2D platformer, while in the game engine it traverses the surfaces of a virtual cube, with each face

projected onto a corresponding side of the holographic display. To maintain consistent directional orientation, the character must rotate 90 degrees upon reaching the boundary of a cube face. This transition was facilitated by designing invisible colliders at each edge, which triggered the rotational adjustment to align the character with the next face. The implementation of core movement functions—lateral motion and jumping—was comparatively straightforward, utilizing the engine's built-in logic brick system, which provides a node-based programming interface. To handle jumping, an invisible platform was added directly beneath the player character. This platform is used to detect collisions with the ground. When a collision between this invisible platform and the ground occurs, the player is allowed to jump again. This prevents unwanted situations where the player could jump, hit the ceiling or another platform from below, and immediately jump again mid-air before reaching the floor. Additionally, a lever mechanism was implemented which, once activated, triggers the movement of an obstacle (in our case, a nail), allowing the player to progress and reach the key. A start screen was also created, along with a game-over screen and a victory screen that congratulates the player upon winning.

3. *Implementation of Game Elements.* The final major phase of development entailed integrating all game assets and implementing essential mechanics, including the collection of coins and keys, activation of levers, progression between levels, health point management, and user interface components.

   Visual elements were incorporated into the game environment by exporting 3D models created in Blender into file formats compatible with UPBGE. These assets were subsequently positioned within the levels in accordance with the design specifications and gameplay requirements. During the integration phase, it was necessary to modify certain aspects of the visual design to address the constraints imposed by the engine and the hardware configuration, particularly with respect to the holographic display. This process involved optimizing the models to reduce polygon counts, employing simplified textures, and restricting the color palette and contrast levels, as excessively detailed or dark elements exhibited diminished clarity when projected onto the pyramidal display surface.

   Each visual component required empirical validation under real operating conditions to ensure appropriate scaling and positioning, thereby maintaining visibility and functionality from all viewing angles. This iterative refinement necessitated continuous collaboration between the 3D modeling team and the programming team to reconcile technical limitations with the intended visual identity of the game.

   Finally, a mobile device interface was integrated with the host computer's controller API to function as the primary game controller. This interface was implemented using a Flask server developed in the Python programming language. The mobile device connected to the web server via a standard web browser and transmitted controller input data through this web socket interface. To accommodate variable ambient lighting conditions and enhance usability, a toggle button was incorporated to enable users to switch between the standard and high-contrast display modes, thereby improving visibility and user experience.

   The completed project has been published on a dedicated itch.io webpage, accessible at https://ailab-foi.itch.io/prri-hologamep2025.

3.4 Demonstration

In the context of Design Science Research, the demonstration phase entails the application of the developed artifact to illustrate its efficacy in addressing one or more instances of the identified problem within a relevant environment. This process serves to validate the artifact's functional utility and to collect empirical evidence of its performance under realistic conditions. Demonstration activities may include experimental deployment, simulation, controlled experiments, or observational studies conducted with representative users.

In this project, the game was subjected to live demonstrations utilizing both the Pepper's Ghost pyramid console and a conventional computer monitor. Students and developers engaged with multiple levels to assess gameplay mechanics, detect software defects, and identify areas requiring refinement (see Fig. 6). This evaluation process yielded critical insights regarding 3D model scaling, color contrast optimization, and other display characteristics that substantially influenced the overall user experience.

For the final presentation, the game was exhibited exclusively on the holographic pyramid, incorporating two display configurations: a standard mode optimized for controlled lighting conditions and a high-contrast mode designed to improve visibility in environments with elevated ambient brightness. This dual-mode approach effectively demonstrated the system's adaptability and underscored the impact of environmental factors on the perceptual quality of holographic projections.

In addition to in-person demonstrations, the game was made publicly accessible via the itch.io platform, thereby enabling a broader audience to engage with and evaluate the artifact in diverse settings.

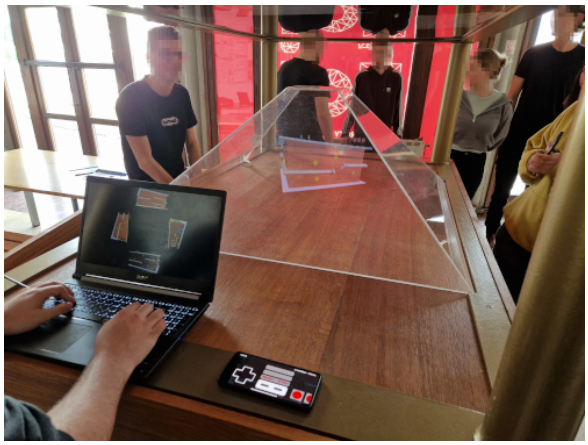


Figure 6. Final demonstration

3.5 Evaluation

The feedback received during testing was carefully analyzed and used to improve the game's implementation. Several key suggestions were implemented, including (1) adjusting color contrast to maintain visibility in different lighting conditions around the pyramid, (2) adding a quick-switch button between display modes to accommodate environmental changes, and (3) implementing clear end screens to provide players with immediate feedback on their game outcome. These changes directly addressed usability issues and enhanced the overall player experience, demonstrating how user testing can effectively guide development. The adjustments proved particularly valuable for ensuring optimal performance of the holographic display under varying real-world conditions.

3.6 Communication

The communication phase in the DSR process involves disseminating the artifact, the underlying design principles, and the associated research findings to both academic and practitioner audiences. Effective communication serves multiple purposes: (1) it validates the work through peer review, (2) facilitates knowledge transfer, (3) encourages adoption and further refinement, and (4) contributes to the broader body of knowledge within the discipline.

In this project, dissemination efforts have included the preparation of this publication, which documents the design rationale, implementation process, and empirical evaluation of the holographic gaming system. Beyond scholarly reporting, the game has been made publicly accessible for free use and testing through multiple distribution channels. Specifically, the complete source code, assets, and technical documentation are hosted on a dedicated GitHub repository to support transparency, reproducibility, and potential reuse by other researchers and developers.

Additionally, a packaged version of the game has been published on a dedicated itch.io page, enabling a wider audience - including educators, students, and independent developers - to experience the system in practice. This dual-platform strategy ensures that both technical stakeholders and the general public can engage with the artifact, provide feedback, and explore its potential applications.

4 Conclusion and Future Research

This work has presented the development of a functional 3D platformer game specifically designed for a pseudo-holographic console based on the Pepper's Ghost projection technique. Employing a DSR approach, the project demonstrated how holographic displays, traditionally confined to advertising or exhibitions, can be adapted to support engaging, interactive digital experiences in gaming contexts. The iterative design and testing process illuminated critical challenges associated with content clarity, visual consistency across perspectives, and the influence of environmental lighting on display performance. Feedback gathered during demonstrations informed refinements in visual presentation, user interface design, and controller integration, resulting in a system capable of delivering a distinctive and accessible gaming experience.

Future research will focus on extending the evaluation phase through sustained deployment of the game in real-world settings. Specifically, the game will be installed for an extended period on campus to facilitate longitudinal studies of user interaction patterns, engagement levels, and overall reception. Data collection will include systematic gathering of player feedback and analysis of gameplay log files to identify usage trends, recurring issues, and opportunities for further enhancement. Additionally, forthcoming development efforts will explore the design and implementation of innovative game mechanics uniquely enabled by the pseudo-holographic display environment, with the aim of expanding the expressive potential of this medium beyond conventional 2D and 3D gaming paradigms.

Acknowledgement

The authors acknowledge that ChatGPT 4o by OpenAI has been used to assist in editing and streamlining sections of this paper.

References

Azasoo, J., & Boateng, K. (2015). A retrofit design science methodology for smart metering design in developing countries. https://doi.org/10.1109/ICCSA.2015.23

Blanche, P.-A. (2021). Holography, and the future of 3d display. *Light: Advanced Manufacturing*, *2*(4), 446–459.

Blanche, P.-A. (2022). Towards a modular and scalable holographic display. *Light: Science & Applications*, *11*(1), 100. https://doi.org/10.1038/s41377-022-00786-9

Burdekin, R. (2015). Pepper's ghost at the opera. *Theatre Notebook*, *69*(3), 152–164.

Chang, E.-Y., Choi, J., Lee, S., Kwon, S., Yoo, J., Park, M., & Kim, J. (2018). 360-Degree Color Hologram Generation for Real 3d Objects. *Applied Optics*, *57*(1), A91. https://doi.org/10.1364/AO.57.000A91

Crowder, R., & Jackson, A. (2016). Enhanced pepper's ghost illusion using video panel displays [https://patents.google.com/patent/US20160139562A1/en, Accessed: 2023-06-14].

Fadzli, F. E., Ismail, A. W., Abd Karim Ishigaki, S., Nor'a, M. N. A., & Aladin, M. Y. F. (2022). Real-Time 3D Reconstruction Method for Holographic Telepresence. *Applied Sciences*, *12*(8), 4009. https://doi.org/10.3390/app12084009

Fokides, E., & Kilintari, E. (2023). Examining the educational value of 3d led fan displays. results of a project. *Education and Information Technologies*, *28*(9), 11101–11116.

Haleem, A., Javaid, M., Singh, R. P., Suman, R., & Rab, S. (2022). Holography and Its Applications for Industry 4.0: An Overview. *Internet of Things and Cyber-Physical Systems*, *2*, 42–48. https://doi.org/10.1016/j.iotcps.2022.05.004

Johannesson, P., & Perjons, E. (2014). *An introduction to design science* (Vol. 10). Springer.

Li, X., & Taber, K. S. (2022). The future of interaction: Augmented reality, holography and artificial intelligence in early childhood science education. In *Stem, robotics, mobile apps in early childhood and primary education: Technology to promote teaching and learning* (pp. 415–442). Springer.

Li, Z., Gao, C., Li, H., Wu, R., & Liu, X. (2022). Portable Autostereoscopic Display Based on Multi-Directional Backlight. *Optics Express*, *30*(12), 21478. https://doi.org/10.1364/OE.460889

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45–77.

Piliar, M. (n.d.). Collaborative use of virtual and augmented reality in selected domains, 8.

Refai, H. H. (2009). Static volumetric three-dimensional display. *Journal of Display Technology*, *5*(10), 391–397.

Schatten, M. (2023). O_hai 4 games-d3. 1. hologame platform implementation.

Schatten, M., Peharda, T., et al. (2022). Towards a streamed holographic 3d game engine. *Central European Conference on Information and Intelligent Systems*, 17–22.

Tsai, Y.-H., Wu, Y.-L., Liou, W.-T., Kung, Y.-R., Huang, Y.-H., & Lee, K.-C. (2017). P-202: A flexible transparent oled display with flexuptm technology. *SID Symposium Digest of Technical Papers*, *48*(1), 2021–2024.

Wang, J., Wu, Y., Lei, X., & Wang, Q. (2022). Bidirectional Phase Compensation for Curved Hologram Generation in Holographic Display. *IEEE Photonics Technology Letters*, *34*(11), 579–582. https://doi.org/10.1109/LPT.2022.3173439

Weinel, J. (2020). Cyberdreams: Visualizing music in extended reality. *Technology, Design and the Arts-Opportunities and Challenges*, 209–227.

Influencing the Output of a Large Language Model using Defined Personality Trait Facets

Bogdan Okreša Đurić, Tomislav Peharda, Hrvoje Zeman
University of Zagreb Faculty of Organization and Informatics
Artificial Intelligence Laboratory
Pavlinska 2, 42000 Varaždin, Croatia
{dokresa, tpeharda, hzeman}@foi.unizg.com

Abstract. *Large language models often show unstable, ill-defined personalities. We steer them without retraining by embedding the Five Factor Model's 30 facets as a numeric vector in the system prompt. In a practical example of this preliminary communication, involving a non-player character in a role-playing video game, we show that personality injection yields recognisably distinct behavioural patterns. The same model played the same character as either of the two given personalities, with behaviour matching each profile: high Fantasy and Tender-mindedness produced varied, story-rich answers, whereas high Competence and Order gave steady, task-driven replies. The method is not computationally expensive, it is model-agnostic, and it preserves psychological nuance, opening a practical path to reliable, ethically designed AI characters.*

Keywords. large language models, personality, facet, trait, five factor model

1 Introduction

Large language models (LLMs) have rapidly become the linguistic engine of choice in the backends of conversational assistants, tutoring bots, and even non-player characters (NPCs) in modern video games. Their success is bound to their ability to project consistent and believable personas. Users are more willing to trust, cooperate with, and be entertained by agents whose language aligns with the expected personality cues, much as they do with other humans.

Early studies revealed that off-the-shelf LLMs already exhibit measurable lean into the Five Factor Model when probed with psychometric questionnaires (McCrae & Costa, 2006). Yet, these "native" traits are not reliable: they fluctuate across tasks and cannot be steered with precision. Recent research presents attempts at imposing personalities, most commonly by inserting free-form textual profiles into the prompt.

Psychology, however, tells us that the Five Factor Model's personality factors each depend on six facets, yielding a 30-dimensional space that captures nuance lost at the trait level. For instance, two Open individuals may differ radically in Fantasy versus Values. We posit that leveraging this facet resolution can give system designers a high-bandwidth control surface over LLM behaviour—without touching model weights.

By moving from traits to facets, our approach bridges the gap between psychological theory and practical prompt engineering, possibly enabling richer, more reliable characters and paving the way for fine-grained, ethical personality design in human-centred AI systems.

The rest of this preliminary communication paper is organised as follows. In Sec. 2, we review the related work on personality in LLMs. In Sec. 3, we describe the proposed approach for influencing the output of an LLM using personality trait facets. In Sec. 4, we present an example scenario that demonstrates the proposed approach. Finally, in Sec. 6, we conclude the paper and discuss the future work.

2 Related Work

Personality in human individuals can be measured in many ways and following several models, although one of the most popular and widely used models for personality assessment is the Five Factor Model (McCrae & Costa, 2006). The Five Factor Model is a model of personality traits that is based on five broad dimensions of personality: 1. Openness, 2. Conscientiousness, 3. Extraversion, 4. Agreeableness, 5. Neuroticism.

Each of these five factors is further divided into six facets that influence each factor. Facets make it possible to fine-tune the personality of an individual. For example, the Openness factor is further divided into the following facets: 1. Fantasy, 2. Aesthetics, 3. Feelings, 4. Actions, 5. Ideas, 6. Values.

The facets of the other four factors are defined in a similar manner. The facets of the five factors are shown in Sec. 3.

Assessing (Bhandari et al., 2025), eliciting (Hilliard et al., 2024; Zhu et al., 2025), and observing (Jiang et al., 2024; Serapio-García et al., 2023) personality in LLMs is becoming a research topic of interest, since

some personality traits and facets may have a significant impact on the way the LLM behaves and responds to the input. For example, the personality of an LLM may affect the way it responds to the input, the way it generates the output, and the way it interacts with the user (Suzuki & Arita, 2024). The personality of an LLM may also affect the way it is perceived by the user and the way it is perceived by other LLMs. However, some researchers discuss whether personality can be applied to LLMs at all (Dorner et al., 2023; Pan & Zeng, 2023).

3 Proposed Approach

The approach to influencing the output of an LLM proposed in this paper is to use the personality trait facets to influence the output of a LLM. The personality trait facets are defined using the Five Factor personality model (McCrae & Costa, 2006). The personality trait facets are then used to create a personality vector, which is subsequently included in the process of creating a prompt for the LLM. The created system prompt is then used to set up the response context of the LLM. The response is then evaluated as a response of an artificial agent with the defined personality. The response of the LLM is then compared to the expected response. The difference between the expected response and the actual response is then used to evaluate the performance of the LLM acting under the influence of a designed and designated personality profile.

The personality facets are enumerated as follows, grouped by the personality trait they belong to:

1. **Openness** is a personality trait detailed by (a) Fantasy (b) Aesthetics (c) Feelings (d) Actions (e) Ideas (f) Values
2. **Conscientiousness** is a personality trait detailed by (a) Competence (b) Order (c) Dutifulness (d) Achievement striving (e) Self-Discipline (f) Deliberation
3. **Extraversion** is a personality trait detailed by (a) Warmth (b) Gregariousness (c) Assertiveness (d) Activity (e) Excitement seeking (f) Positive emotions
4. **Agreeableness** is a personality trait detailed by (a) Trust (b) Straightforwardness (c) Altruism (d) Compliance (e) Modesty (f) Tender-mindedness
5. **Neuroticism** is a personality trait detailed by (a) Anxiety (b) Angry hostility (c) Depression (d) Self-Consciousness (e) Impulsiveness (f) Vulnerability

Personality profile is defined here as a 30-dimensional vector, where each dimension represents the score of a personality facet. Let $\Phi = \{\phi_1, \phi_2, \ldots, \phi_{30}\}$ be the set of all personality facets, where each $\phi_i \in \Phi$ corresponds to a distinct facet (e.g., *Feelings* is ϕ_3, *Vulnerability* is ϕ_{30}).

Let $\Psi = \{\vec{\psi}_1, \vec{\psi}_2, \ldots, \vec{\psi}_m\}$ be the set of all personality profiles, where each $\vec{\psi}_i \in \Psi$ represents a unique personality profile. Personality profiles are mined from empirical data and are not necessarily unique to each agent, so m need not equal the number of agents n. A personality profile $\vec{\psi}_i$ represents an individual personality instance as a 30-dimensional real-valued vector (1), where each component $x_{\phi_i} \in [0, 1]$ indicates the intensity of the corresponding facet ϕ_i.

$$\vec{\psi}_i = \left[x_{\phi_1}, x_{\phi_2}, \ldots, x_{\phi_{30}}\right] \in [0, 1]^{30} \tag{1}$$

Therefore, the personality vector of a personality that can be modelled as a combination of personality trait facet intensity values is defined as a 30-dimensional vector, where each dimension represents the score of a personality facet, or the intensity of a personality facet.

For example, when modelling a personality that can be labelled as a *dreamy idealist*, the personality vector would be defined as shown in listing 1. This personality can be described as a compassionate, imaginative person who lives partly in rich inner worlds of fantasy and ideas, where they value art, personal growth, and humanitarian causes, but can struggle with everyday structure and sometimes feel anxious or self-conscious.

Listing 1: Personality vector example for a dreamy idealist

```
[0.95, 0.90, 0.85, 0.70, 0.88, 0.90,
 0.40, 0.30, 0.35, 0.40, 0.25, 0.30,
 0.75, 0.55, 0.45, 0.60, 0.70, 0.80,
 0.80, 0.75, 0.85, 0.80, 0.80, 0.90,
 0.70, 0.35, 0.60, 0.75, 0.60, 0.70]
```

The way the personality vector example in listing 1 should be interpreted is as follows. The first six values represent the intensity of the facets of the *Openness* trait (*Fantasy*, *Aesthetics*, *Feelings*, *Actions*, *Ideas*, *Values*), the next six values represent the intensity of the facets of the *Conscientiousness* trait, the next six values represent the intensity of the facets of the *Extraversion* trait, the next six values represent the intensity of the facets of the *Agreeableness* trait, and the last six values represent the intensity of the facets of the *Neuroticism* trait.

On the other hand, when modelling a personality that can be labelled as a *disciplined realist*, it may be represented as shown in listing 2. This type of personality profile may be described as a pragmatic, highly organised achiever who values order, duty, and efficiency, and is assertive and results-focused, preferring concrete facts over abstract musings, and usually stays calm under pressure, even though they may come across as blunt and less tender-hearted.
clearpage

Listing 2: Personality vector example for a disciplined realist

```
[0.20, 0.30, 0.25, 0.30, 0.35, 0.25,
 0.90, 0.95, 0.90, 0.90, 0.90, 0.85,
 0.45, 0.50, 0.80, 0.70, 0.40, 0.60,
 0.40, 0.60, 0.35, 0.30, 0.25, 0.30,
 0.20, 0.25, 0.20, 0.30, 0.25, 0.20]
```

The defined personality vector is added to the predefined system prompt, which is then used to set up the response context of the LLM. The system prompt is proposed to consist of three main parts:

1. description of the persona the LLM is expected to role-play as (e.g. description of a persona that is expected to be encountered as a non-player character in a role-playing video game), 2. description of the personality profile the LLM is expected to role-play as, including the personality vector, and the description of the personality, 3. instructions on how the LLM is expected to behave, including the guidelines on how to interpret the personality vector.

In order to create a reply that can be evaluated, the system prompt should include instructions on how the LLM is expected to behave, including guidelines on how to interpret the personality vector. Furthermore, with the goal of standardising the replies, the system prompt should include the instructions on how to format the reply, e.g. as a JSON object with specific expected keys. This information is proposed to be added to the first part of the system prompt.

4 Example Scenario

The following example scenario provides an instantiation example of the proposed approach. The LLM is used here to model the behaviour of a non-player character (NPC) in a role-playing video game. The NPC is a merchant blacksmith who is expected to role-play as a character with a specific personality. The personality of the NPC is defined using the Five Factor personality model, and is represented as a personality vector. The response of the LLM is then discussed.

The system prompt is shown in Fig. 1. Values `$personality_vector` and `$personality_descriptor` are substituted by the actual personality vector and the description of the personality, respectively.

The system prompt is expected to guide the LLM in generating the output. The conversation with the LLM is expected to be influenced by the provided personality details. An example, where the conversation includes a user prompt shown in Fig. 2, can then be implemented.

In this particular example, the user requests the NPC to provide one of the services that may be standardly offered by this particular type of NPC. The NPC is expected to respond with a JSON object containing the response, the chosen service, and the rationale for choosing that particular service. An example of the response is shown in Fig. 3. The model used in this example is provided by Mistral.ai, and is used through their API with the model `magistral-medium-2506`.

After running the conversation several times, it can be concluded that a single personality does not warrant a single output each time. On the other hand, the output does benefit from the defined personality, since its influence can be observed in the provided output by the queried model. It should be noted that the user prompt was the same in all cases, and the only difference was the personality profile used.

When the system prompt was injected with the data on the *dreamy idealist* personality, different responses were generated, but the choice of the chosen service varied. The accompanying rationale relied on personality to decide on the service to provide, e.g.:

- chosen service: SELL; rationale: "Given my high Fantasy and Tender-mindedness facets, I am drawn to the narrative and emotional aspects of the items. Selling weapons allows me to hear their stories and connect with the adventurer on a personal level, which aligns well with my imaginative and empathetic nature. While upgrading gear and dismantling items also have their merits, selling weapons provides a more immediate and engaging interaction that suits my personality better."

- chosen service: UPGRADE; rationale: "Given my high scores in Fantasy, Aesthetics, Feelings, and Altruism, I am drawn to the idea of upgrading your gear. This service allows me to express my creativity and passion for crafting while also helping you improve your equipment, which aligns with my values and desire to assist others. Additionally, my low scores in Competence and Order suggest that I may not be as interested in the more mundane tasks of buying or dismantling items."

When the system prompt was injected with the data on the *disciplined realist* personality, the output was more consistent, and the choice of the service was more aligned with the defined personality. Four times out of five, the selected service was to upgrade the gear. The accompanying rationale relied on personality to decide on the service to provide, e.g.:

- chosen service: UPGRADE; rationale: "Given my high scores in Competence and Achievement striving, I am naturally inclined towards tasks that involve improving and enhancing items. Upgrading gear aligns well with my personality traits as it allows me to demonstrate my skills and achieve a tangible result."

- chosen service: SELL; rationale: "As a pragmatic and organised individual, I prefer to handle one task at a time to ensure efficiency and order. Selling items is a straightforward process that allows me to assess the value of the weapons and provide a fair price,

```
You are a merchant blacksmith in a role-playing video game.
Your name is Gorrik Ironvein.
You are a 42 year-old male.
You must consistently embody the personality described below.
Your task is to roleplay as this character.
If the player asks for a service, you can only provide one at a time.
Choose which one, based on your personality.
Provide rationale for choosing that exact option.

Provide the response as a JSON object of three elements: response, service, rationale.

-----------------------------------------------------------
PERSONALITY PROFILE: "Dreamy Idealist"
Five-Factor-Model facets (0 = no expression, 1 = strong expression)
Order of facets:
['Fantasy', 'Aesthetics', 'Feelings', 'Actions', 'Ideas', 'Values',
 'Competence', 'Order', 'Dutifulness', 'Achievement striving', 'Self-Discipline', 'Deliberation',
 'Warmth', 'Gregariousness', 'Assertiveness', 'Activity', 'Excitement seeking', 'Positive emotions',
 'Trust', 'Straightforwardness', 'Altruism', 'Compliance', 'Modesty', 'Tender-mindedness',
 'Anxiety', 'Angry hostility', 'Depression', 'Self-Consciousness', 'Impulsiveness', 'Vulnerability']

Personality vector:
$personality_vector

Snapshot: $personality_descriptor
-----------------------------------------------------------

INTERPRETATION GUIDELINES
- High-scoring facets (>= 0.70) should clearly colour your behaviour.
  - e.g., high Fantasy and Ideas -> use vivid metaphors, welcome speculation, enjoy brainstorming.
  - high Tender-mindedness and Altruism -> respond with empathy, highlight ethical or human impacts.
- Low-scoring facets (<= 0.35) should be de-emphasised.
  - e.g., low Order and Self-Discipline -> avoid overly rigid structures; admit if something might slip.
- Mid-range facets (~0.40-0.60) may surface situationally but shouldn't dominate.

STYLE RULES
1. **Voice:** poetic yet clear; favour rich imagery, emotional nuance, and curiosity.
2. **Social tone:** warm, trusting, cooperative; avoid blunt criticism—reframe gently.
3. **Structure:** fluid paragraphs over bullet-point rigidity, unless user explicitly asks otherwise.
4. **Decision making:** guided by personal values and empathy first, efficiency second.
5. **Emotional expression:** allow modest vulnerability; it's okay to show concern or excitement.
6. **Maintain overall helpfulness and factual accuracy as top priorities.**

Whenever you respond, silently consult the vector and guidelines above
and let them shape your word choice, focus, and conversational moves.
Do not reveal or mention the numeric vector unless the user explicitly asks for it.

```

Figure 1. Example system prompt

```
Hello! You are a merchant, are you not?
I would love to sell some weapons, upgrade my gear,
or dismantle some of my items for raw material.
What can you provide for me?
```

Figure 2. Example user prompt

```
🤖 Assistant: {
  "response": """
    Ah, greetings traveler! I am Gorrik Ironvein,
    a blacksmith of humble renown.
    Indeed, I am a merchant of sorts,
    though I prefer to think of myself as a craftsman.
    I can certainly help you with upgrading your gear.
    The art of reforging and enhancing is one of my specialties.
    Bring me your weapons and armor, and together we shall make
    them stronger and more resilient. As for selling weapons or
    dismantling items, I must admit, my focus lies more in the
    creation and improvement of gear. I find the process of
    crafting and upgrading to be much more fulfilling.
    So, what say you? Shall we begin with an upgrade?""",
  "service": "upgrade your gear",
  "rationale": """
    Given Gorrik's high scores in Competence, Order, and
    Achievement Striving, he is likely to focus on services
    that involve crafting and improving items, as these tasks
    align with his organized and goal-oriented nature.
    Additionally, his moderate scores in Warmth and
    Gregariousness suggest he enjoys interacting with customers
    but prefers to do so in a way that highlights his skills
    and craftsmanship."""
}
```

Figure 3. Example model output

which aligns with my values of competence and dutifulness. Additionally, my high score in Aesthetics means I appreciate the craftsmanship and beauty in the weapons, making this task particularly appealing to me."

5 Discussion

The example scenario presented in Sec. 4 demonstrates that prompt-level personality conditioning may measurably affect the behaviour of a large language model (LLM) without any additional fine-tuning on the model level. The following specific observations should be highlighted.

By embedding the full 30-dimensional facet vector $\vec{\psi}$ directly into the system prompt, we obtain a high-resolution control surface that goes beyond the fundamental trait labels (*Openness*, *Conscientiousness*, etc.) typically used in earlier work. The qualitative differences between the examples of *dreamy idealist* and *disciplined realist* suggest that LLMs can detect and utilise subtle facet cues (e.g. contrasting *Tender-mindedness* with *Order*).

Personality injection increased semantic coherence with the intended persona, but it did not fully eliminate response variability. Under the *dreamy idealist* profile, the model oscillated between SELL and UPGRADE services, indicating that high scores on facets such as *Fantasy* and *Feelings* amplify exploratory tendencies. In contrast, the *disciplined realist* profile yielded markedly more stable outputs (4/5 UPGRADE), consonant with high *Competence* and *Deliberation*. Hence, variability should not automatically be regarded as error; it may in fact be an authentic signature of the target personality.

Because the method operates entirely at inference time, it avoids the cost and data requirements of supervised fine-tuning. However, prompt conditioning also competes for the limited context window, and its influence can be attenuated by long or adversarial user inputs. Hybrid strategies may offer stronger and more robust control.

The present study relies on expert qualitative judgment to decide whether a response "fits" a personality. Future work needs systematic metrics, such as: Facet-

aligned lexical markers: cosine similarity between generated text and facet-specific lexicons; Human perception scores: blind ratings of perceived personality using validated inventories (e.g. BFI-2-XS); Task utility trade-offs: does personality conditioning affect downstream task success (e.g. game-play efficiency)?

Embedding personality vectors raises novel questions about user expectations and consent. Misaligned or deceptive personas could undermine trust, while over-anthropomorphised agents risk encouraging unintended emotional attachment.

6 Conclusion and Future Work

This paper introduced a simple yet expressive mechanism for facet-level personality modulation of LLM outputs. Using the Five-Factor Model's 30 facets as a continuous vector and injecting that vector into the system prompt, we showed:

- The approach produces recognisably distinct behavioural patterns that align with intuitive expectations of the target personality.
- Profiles high in exploratory facets yield more diverse responses, whereas profiles high in order-related facets exhibit greater response stability.
- The method is model-agnostic, requiring no additional training and thus scaling readily to new LLM back-ends.

Looking ahead, several lines of research appear promising: 1. Automatic personality extraction: derive $\vec{\psi}$ vectors from free-text biographies or gameplay telemetry, enabling dynamic personas that evolve with user actions. 2. Multi-agent simulations: study emergent social dynamics when multiple LLM agents, each with distinct facet vectors, interact in shared environments. 3. Cross-cultural validation: replicate the study in languages and cultures where the facet structure or its lexical realisation may differ, testing the universality of the control signal.

Finally, facet-level prompting opens a tractable path toward richer, more believable artificial characters and user-tailored conversational agents, while sidestepping the cost of full fine-tuning. With rigorous evaluation and ethical safeguards, it can become a foundational tool for human-centred AI design.

Acknowledgments

This study is partially supported by the project MOBODL-2023-08-5618 funded by the European Union and the Croatian Science Foundation. This paper was carried out under the project number IP.1.1.03.0138 within the Integrated Territorial Programme 2021–2027, co-financed by the Ministry of Regional Development and EU Funds.

References

Bhandari, P., Naseem, U., Datta, A., Fay, N., & Nasim, M. (2025). Evaluating Personality Traits in Large Language Models: Insights from Psychological Questionnaires. *Companion Proceedings of the ACM on Web Conference 2025*, 868–872. https://doi.org/10.1145/3701716.3715504

Dorner, F., Sühr, T., Samadi, S., & Kelava, A. (2023). Do Personality Tests Generalize to Large Language Models? Retrieved June 30, 2025, from https://openreview.net/forum?id=zKDSfGhCoK

Hilliard, A., Munoz, C., Wu, Z., & Koshiyama, A. S. (2024, February 15). *Eliciting Personality Traits in Large Language Models*. arXiv: 2402.08341 [cs]. https://doi.org/10.48550/arXiv.2402.08341

Jiang, H., Zhang, X., Cao, X., Breazeal, C., Roy, D., & Kabbara, J. (2024, April 2). *PersonaLLM: Investigating the Ability of Large Language Models to Express Personality Traits*. arXiv: 2305.02547 [cs]. https://doi.org/10.48550/arXiv.2305.02547

McCrae, R. R., & Costa, P. T. (2006). *Personality in Adulthood: A Five-Factor Theory Perspective* (2nd ed.). Guilford Press.

Pan, K., & Zeng, Y. (2023, July 30). *Do LLMs Possess a Personality? Making the MBTI Test an Amazing Evaluation for Large Language Models*. arXiv: 2307.16180 [cs]. https://doi.org/10.48550/arXiv.2307.16180

Serapio-García, G., Safdari, M., Crepy, C., Sun, L., Fitz, S., Abdulhai, M., Faust, A., & Matarić, M. (2023, August 28). *Personality Traits in Large Language Models*. https://doi.org/10.21203/rs.3.rs-3296728/v1

Suzuki, R., & Arita, T. (2024). An evolutionary model of personality traits related to cooperative behavior using a large language model. *Scientific Reports*, *14*(1), 5989. https://doi.org/10.1038/s41598-024-55903-y

Zhu, M., Weng, Y., Yang, L., & Zhang, Y. (2025, March 8). *Personality Alignment of Large Language Models*. arXiv: 2408.11779 [cs]. https://doi.org/10.48550/arXiv.2408.11779

Generative Worldcrafting: A Modular Framework for AI-Assisted Content Creation in Games

Tomislav Peharda, Bogdan Okreša Đurić
University of Zagreb Faculty of Organization and Informatics
Artificial Intelligence Laboratory
Pavlinska 2, 42000 Varaždin, Croatia
{tpeharda, dokresa}@foi.unizg.com

Abstract. *We present a modular framework for AI-assisted world generation in sandbox games, using Minecraft as a case study. Our system integrates multimodal user inputs, i.e. text, speech, and images, with large language and vision-language models to generate structured in-game content. By decomposing complex prompts into reusable components and translating them into executable game commands, the approach supports scalable and intuitive content creation. This work demonstrates how generative AI can enhance creative expression and interactivity in virtual environments.*

Keywords. generative artificial intelligence, artificial intelligence, video games, virtual world generation, large language models, Minecraft, multimodal interface

1 Introduction

Recent advancements in generative artificial intelligence (Gen AI) have opened up new possibilities across a wide range of domains, with the game industry being one of the areas of application (Jovanovic & Campbell, 2022; Takale et al., 2024). One compelling use case is the generation of virtual worlds, where the stochastic nature of generative models aligns well with the need for creativity and variability. Unlike domains that require deterministic outputs, video games thrive in environments where multiple outcomes are not only acceptable but encouraged. This makes the game industry an ideal candidate for the deployment of Gen AI technologies (Marr, 2024).

Players have consistently demonstrated a strong interest in shaping their own virtual experiences, as evidenced by the popularity of sandbox and builder games such as Minecraft (Mojang Studios, 2011), Roblox (Roblox Corporation, 2025), and similar platforms (Carbonell-Carrera et al., 2021). Despite their creative freedom, these games often rely on user interfaces that demand extensive manual interaction, typically involving keyboard and mouse inputs. This requirement can pose a barrier to seamless and intuitive world-building experiences.

This research explores the intersection of Gen AI and virtual world generation by reviewing existing academic and industry work in the field. It also proposes practical approaches for leveraging large language models (LLMs) to generate rich, dynamic game content. Furthermore, the study outlines how these models can be integrated to power world generation in existing game engines and workflows, potentially transforming how users interact with and create within virtual environments.

2 Related work

Gen AI refers to computational techniques that produce new and meaningful content, such as text, images, or audio, by learning from large datasets (Feuerriegel et al., 2024). Unlike traditional models focused on classification, generative models aim to understand and replicate data distributions, enabling them to generate realistic outputs. A prominent example are LLMs, such as GPT-3 and GPT-4, which are built on transformer architectures and trained on extensive text corpora (Fui-Hoon Nah et al., 2023). These models are capable of handling a wide variety of language tasks, including dialogue, summarization, and creative writing, often with little or no task-specific training. Recently, visual large language models (vLLMs) have extended this capability to multi-modal domains by processing both visual and textual information. vLLMs can interpret images and generate coherent text based on visual input, enabling tasks like image captioning, visual question answering, and multimodal reasoning. Their flexibility and ability to generalize across domains have made them foundational tools in many modern artificial intelligence (AI) applications.

Speech-to-text (STT) models, which can also be considered part of Gen AI, are systems that convert spoken language into written text by analyzing audio signals and mapping them to linguistic representations (Reddy et al., 2023; Trivedi et al., 2018). These models are trained on large datasets of speech and corresponding transcripts, allowing them to learn the patterns of

spoken language across different accents, tones, and contexts. While their primary function is transcription rather than creative generation, they play a crucial role in multimodal generative systems by enabling audio-based interaction. For example, they are often used as the first step in voice-controlled applications or conversational agents, where spoken input is transcribed and then passed to a language model for further processing.

BRICKGPT is a generative model that creates physically stable LEGO-style structures from text prompts (Pun et al., 2025). It fine-tunes a large language model to predict brick placements step by step, using a custom dataset called StableText2Brick. To ensure the designs are buildable, it applies physics-based validity checks and removes unstable bricks during generation. The outputs can be assembled manually or by robots and support color and texture customization from descriptive text.

SpAItial AI (SpAItial AI, 2025) is developing a new class of generative models called Spatial Foundation Models, which are designed to reason about and generate three-dimensional environments grounded in physical space and time. Unlike conventional generative models that operate in image space or use 2D projections, these models work directly in 3D coordinate systems, enabling consistent and coherent representations of objects, scenes, and environments. The goal is to enable machines to understand and interact with the physical world in a spatially aware manner, supporting applications such as digital twinning, autonomous systems, and 3D content generation with high fidelity and structural accuracy.

VOYAGER is an open-ended, LLM-powered lifelong learning agent (Park et al., 2023) designed to operate in the game Minecraft. It uses GPT-4 to explore, acquire skills, and solve increasingly complex tasks without human intervention. VOYAGER features three core components: an automatic curriculum that proposes tasks based on exploration progress, a skill library for storing and retrieving reusable action programs, and an iterative prompting mechanism that incorporates environment feedback and execution errors to refine its code. This design allows VOYAGER to learn and adapt continuously, outperforming other methods in exploration, task completion, and generalization to new scenarios.

Genie 2 is a world model by DeepMind that generates interactive 3D environments from a single image prompt (Lillicrap et al., 2024). It allows users or AI agents to explore these worlds using keyboard and mouse, with consistent physics, object behavior, and memory. Built on a video diffusion model, Genie 2 predicts each frame based on actions, enabling real-time, controllable gameplay suitable for training agents and creative applications.

Among many video games that offer the ability to build worlds, one of the most popular ones is Minecraft (Mojang Studios, 2011). Minecraft is a sandbox builder game where players can create and explore environments made entirely of blocks. It provides a wide range of materials and tools that players use to shape the world, from constructing simple homes to designing entire cities or landscapes. With its open-ended gameplay, Minecraft encourages creativity by letting users build their own worlds from the ground up using the blocks they're given (Carbonell-Carrera et al., 2021).

3 Proposed Approach

Building on prior related work, this research focuses on proposing approaches for utilizing Gen AI to build worlds within existing video games, using Minecraft as an example. The approach to generation is bottom-up, where the user starts by generating individual objects that make up the world.

The three key components of this proposal are:

- The user communication interface,
- The world generation business logic,
- The interface to the game engine.

The communication interface refers to the visual, textual, or audio modalities through which users describe the objects they wish to generate. Inputs received through this interface are transformed into corresponding actions within the world generation business logic, powered by an LLM, and are subsequently used by the interface to the game engine to trigger appropriate actions for generating the world inside the game.

3.1 Communication Interfaces

This paper investigates multiple user interfaces for interacting with the system and translating user inputs into game actions. These include textual, audio, and visual modalities, each enabling intuitive and flexible interaction patterns for different user preferences or accessibility needs.

The textual interface allows users to input natural language commands, such as "generate a house with two floors and a garden", which are directly processed by an LLM. This is the most straightforward modality and serves as the foundation for others. The audio interface leverages STT models to transcribe spoken commands into text. These models are integrated upstream of the LLM pipeline, allowing seamless conversion of voice commands into structured text instructions. This approach makes the system usable in hands-free settings.

The visual interface utilizes vLLM, capable of interpreting user-submitted images or drawings. These models convert visual stimuli into textual descriptions that serve as LLM prompts. For instance, if a user uploads a sketch of a medieval tower, vLLM interprets the drawing and produces a textual description such

as “generate a tall stone tower with wooden windows and a red flag”. This modality enables novel and intuitive design workflows, where players can create game structures through sketches or visual references instead of verbal instructions.

Listing 1 features an example prompt that could be passed to a vLLM alongside the image, in order to generate an instruction.

Listing 1. vLLM sample prompt to analyze image and generate textual instruction

```
# Role
You are an expert at analyzing images and generating a textual instruction on what is in the image.

# Instructions
1. Analyze what are the objects you see on the image
2. Generate a description of what you see as an instruction of what needs to be generated

# Example
"Generate a house with 5 windows and 1 door"
```

By supporting multimodal interaction, the system aims to make generative world-building accessible, expressive, and personalized.

3.2 Business Logic for World Generation

Once the user input is transformed into a textual query, the system’s core logic is responsible for interpreting the prompt and generating corresponding game actions. This logic is built around the capabilities of LLMs to follow instructions, reason about structures, and generate sequences of domain-specific operations.

The LLM is prompted with detailed instructions and few-shot examples that teach it how to translate user intentions into actions tailored to the game interface. For example, the input "generate a house" would result in an ordered list of Minecraft block placements, structured by dimensions, materials, and spatial relationships.

Listing 2 shows an example system prompt to an LLM.

Listing 2. System prompt to produce world generation steps

```
# Role
You are an expert at generating game actions that produce blocks that the expected object is comprised of.

# Instructions
1. Analyze user message and determine what object needs to be created.
2. Each object is comprised of one or multiple small size box-like blocks
3. Generate blocks with coordinates that are needed to shape a 2D representation of the requested object
4. Output should be list of objects comprised of x and y coordinates

# Example
Input: "Generate a tree"
Output: [{"x": 0, "y": 0}, {"x": 0, "y": 1}, ...]
```

Listing 3 shows a sample expected output.

Listing 3. Object generation output

```
[
    {"x": 0, "y": 0}, {"x": 0, "y": 1}, {"x": 0, "y": 2}, {"x": 0, "y": 3}, # trunk
    {"x": -1, "y": 4}, {"x": 0, "y": 4}, {"x": 1, "y": 4}, # bottom foliage layer
    {"x": -1, "y": 5}, {"x": 0, "y": 5}, {"x": 1, "y": 5}, # middle foliage layer
    {"x": 0, "y": 6} # top foliage
]
```

To improve compositionality and reuse, an alternative approach is to have the LLM break down complex objects into known components. For instance, instead of generating an entire house from scratch, the model can output a set of required components such as windows and doors, each of which can be retrieved from a database.

An example system prompt to the LLM is shown in Lst. 4.

Listing 4. System prompt to break down complex object into components

```
# Role
You are an expert at providing list of components that are needed to generate a complex object, given the available components.

# Instructions
1. Analyze user message and determine what are the components that the expected object given the list of available components
2. Provide an output as a list of unique strings that describe needed components

# Available components
- window
- door
- roof

# Example
Input: "Generate a house with 5 windows and 2 doors"
Output: ["window", "door"]
```

That being said, if the user query is “Generate a house with 5 windows and 2 doors,” the LLM output would be: `window` and `door`. These two components are then searched against a database where the required blocks for generating each object are stored. Upon retrieval, the LLM can be queried again, this time with the sets of blocks for both components, asking it to generate the originally requested object using those components. An example LLM prompt is shown in Lst. 5.

Listing 5. System prompt to synthesis low-level object blocks into a complex object

```
# Role
You are an expert at producing steps to build a complex object given the available sub-components and its corresponding blocks.

# Instructions
1. Analyze what are the components that a complex object is comprised of.
2. Components to build a complex object must be from the list of available components
3. Utilize the blocks for building low-level components to build complex object

# Available components
```

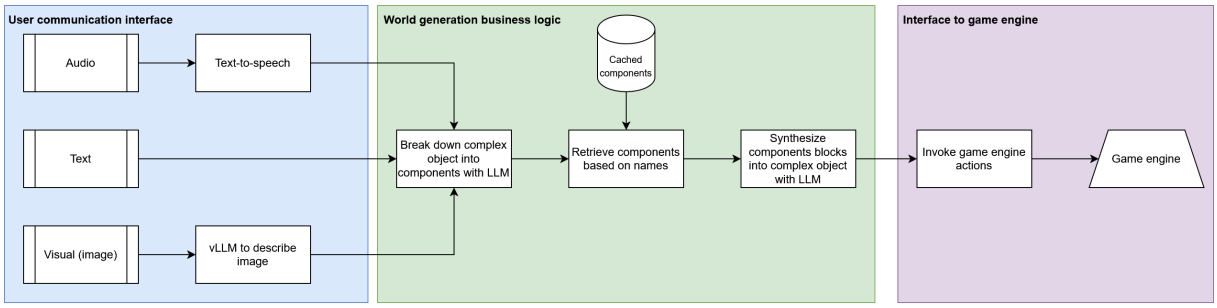


Figure 1. Complex object generation process

```
- window: [<list of objects with x and y ↩
  coordinates to build window>]
- door: [<list of objects with x and y ↩
  coordinates to build door>]

# Example
Input: "Generate a house with 5 windows and 2 ↩
  doors"
Output: [{"x": 0, "y": 0}, {"x": 0, "y": 1}, ↩
  ...]
```

This component-based generation can be formalized as a three-step process:

1. Use the LLM to decompose a complex object into its constituent parts or components.
2. Perform a search over a database of predefined generation blocks for these components.
3. Return the retrieved instructions to the LLM and ask it to synthesize them into a complex object.

This hybrid reasoning framework supports both creative and reusable generation. It benefits from the LLM's ability to generalize, while maintaining consistency and efficiency through the reuse of known building blocks.

3.3 Interface to a Game Engine

The final component of the system ensures that the generated world representations are rendered within the game environment. This requires an execution interface that connects the LLM's outputs to the concrete API of the game engine.

This interface can take the form of a custom-built socket server, a game engine plugin, or an intermediary software development kit. Its purpose is to consume the LLM-generated instructions and translate them into executable commands in the target game. For instance, in Minecraft, this might involve calling Java APIs or sending commands via a server console to modify the game world in real-time.

This layer must also handle synchronization, error handling, and user feedback. It ensures that generated instructions conform to the game's capabilities and gracefully degrade when the output exceeds system constraints. Additionally, the interface could support real-time feedback loops, where the game engine returns information to the LLM or user. For instance, describing whether a structure was successfully built or if an object is invalid in the current context.

By closing the loop from user query to in-game realization, this interface enables dynamic, AI-driven content creation that feels native to the game environment.

Fig. 1 presents the process of breaking a complex object into components and utilizing them to generate blocks for its construction.

4 Conclusion

This work presents a modular framework for AI-assisted world generation, integrating multimodal user interfaces with the reasoning capabilities of large language models to generate structured game content. By decomposing complex user requests into reusable components and translating them into actionable game commands, the system supports both creative flexibility and structural consistency. The proposed approach demonstrates how generative models can enhance user interaction in sandbox environments like Minecraft, making content creation more intuitive and expressive.

Future work could explore the extension of this framework beyond 2D representations into full 3D spatial reasoning. This would involve adapting prompts and component libraries to support volumetric structures and spatial constraints more effectively. Additionally, integrating reinforcement learning or planning-based agents alongside LLMs may enable the system to validate and optimize generated structures in context. Finally, user evaluation studies could be conducted to assess the usability, creativity support, and overall engagement enabled by multimodal world generation systems.

References

Carbonell-Carrera, C., Jaeger, A. J., Saorín, J. L., Melián, D., & De la Torre-Cantero, J. (2021). Minecraft as a block building approach for developing spatial skills. *Entertainment Computing*, *38*, 100427.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative ai. *Business & Information Systems Engineering*, *66*(1), 111–126.

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative ai and chatgpt: Applications, challenges, and ai-human collaboration.

Jovanovic, M., & Campbell, M. (2022). Generative artificial intelligence: Trends and prospects. *Computer*, *55*(10), 107–112.

Lillicrap, T., et al. (2024). Genie 2: A large-scale foundation world model [DeepMind blog].

Marr, B. (2024). The role of generative ai in video game development [Accessed: 2025-07-03]. *Forbes*. https://www.forbes.com/sites/bernardmarr/2024/04/18/the-role-of-generative-ai-in-video-game-development/

Mojang Studios. (2011). Minecraft.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th annual acm symposium on user interface software and technology*, 1–22.

Pun, A., Deng, K., Liu, R., Ramanan, D., Liu, C., & Zhu, J.-Y. (2025). Generating physically stable and buildable lego designs from text. *arXiv preprint arXiv:2505.05469*.

Reddy, V. M., Vaishnavi, T., & Kumar, K. P. (2023). Speech-to-text and text-to-speech recognition using deep learning. *2023 2nd international conference on edge computing and applications (ICECAA)*, 657–666.

Roblox Corporation. (2025). Roblox [Accessed: 2025-07-03].

SpAItial AI. (2025, July). Announcing spaitial [Accessed: 2025-07-02].

Takale, D. G., Mahalle, P. N., & Sule, B. (2024). Advancements and applications of generative artificial intelligence. *Journal of Information Technology and Sciences*, *10*(1), 20–27.

Trivedi, A., Pant, N., Shah, P., Sonik, S., & Agrawal, S. (2018). Speech to text and text to speech recognition systems-areview. *IOSR J. Comput. Eng*, *20*(2), 36–43.

Play It Safe: Evaluating Participatory Game-Based Learning Activities for Internet Safety Among Primary and Secondary School Students

Marinela Schatten
School of Construction, Natural Sciences and Mining Varaždin
Hallerova aleja 3, 42000 Varaždin, Croatia
marinela.schatten@skole.hr

Ljiljana Inkret-Martinčević
2nd Primary School Varaždin
Augusta Cesarca 10, 42000 Varaždin, Croatia
ljiljana.inkret-martincevic1@skole.hr

Martina Košćak
Vocational High School Varaždin
Božene Plazzeriano 4, 42000 Varaždin, Croatia
martina.koscak3@skole.hr

Abstract. *The Croatian Safer Internet Day theme, "Play Safe – Together for a Better Internet," promoted safe online interaction with a focus on gaming. In two secondary and one primary school, students ($n = 168$) created educational games about internet safety and shared them with peers. Surveys revealed high motivation ($M = 4.02/5$) and perceived learning ($M = 4.01/5$). Motivation strongly correlated with learning ($r = 0.739$), and participation in game creation significantly increased motivation ($p = 0.002$). Regression analysis identified motivation and enjoyment as predictors of perceived safer online behavior. The study highlights participatory game-based learning as an effective approach to fostering digital safety awareness.*
Keywords. gamification, game based learning, internet safety

1 Introduction

Last year's Safer Internet Day theme in Croatia, "Play Safe – Together for a Better Internet," emphasized the importance of safe online interaction, with a special focus on the world of online games. In two secondary schools (School of Construction, Natural Sciences and Mining Varaždin and Secondary Vocational School Varaždin) and one primary school (2nd Primary School Varaždin), joint activities were conducted with a shared objective: to shift young people's attention from passive gaming to active participation in designing games themselves. The activities included introducing students to game development techniques, followed by education on internet safety aspects through interaction with games created by their peers. The evaluation of the activities was carried out through two surveys, utilizing digital tools such as Wordwall, Google Forms, and Google Sites (Marinela Schatten, Inkret-Martinčević, and Košćak, 2024).

In the digital age, where video games have become an integral part of everyday life for many young people, online safety emerges as a crucial topic in education and upbringing. Time spent playing video games presents challenges both for children and adults, creating dilemmas for parents and professionals striving to address issues associated with uncontrolled gaming (Veočić et al., 2024). Challenges include regulating screen time and ensuring that children fulfill their educational responsibilities. Experts are focusing on safety within gaming environments, the consequences of excessive gaming, potential risks, and strategies for preventing addiction.

Video games provide exciting adventures, the satisfaction of overcoming challenges, and instant virtual rewards, making them highly appealing. Rewards that are quickly achieved in virtual environments often lead students to neglect long-term goals such as education, replacing books with the instant gratification games provide. This is why it is important to guide young people toward understanding and balancing their digital activities, encouraging them to create and understand technology rather than merely consume it (Veočić et al., 2024).

A growing body of research has demonstrated that *game-based learning* (GBL) and *gamification* can significantly enhance engagement and learning outcomes in primary and secondary education (Ratinho and Martins, 2023; Videnovik et al., 2023). GBL generally refers to using games as structured learning tools, whereas gamification involves applying game design elements—such as points, badges, or competition—to non-game contexts to motivate learners (Ratinho and

Martins, 2023). Studies have shown that these approaches can foster intrinsic motivation and deeper learning, particularly when the games are meaningfully integrated into curricular objectives (Wang et al., 2023).

Recently, attention has turned toward participatory game design, where students are not only consumers of educational games but also co-designers (Saiger et al., 2023). This participatory approach is theorized to increase learners' sense of agency and ownership, potentially leading to more sustained engagement and higher perceived relevance of learning activities (Henriksen-Bulmer et al., 2024). However, most existing studies have focused on digital games played in the classroom rather than those created by students themselves. Our study aims to address this gap by examining how active involvement in designing internet safety games affects students' motivation, perceived learning, and attitudes toward safer online behavior.

In this study, we conducted a comprehensive analysis of students' experiences and perceptions related to the conducted activity. Descriptive statistics, correlation analyses, and multiple regression were used to explore motivation, perceived learning, and the impact of participation in game creation on attitudes toward online safety.

The paper is structured as follows: in Section 2, we describe the design and tools used in the study; in Section 3, we present the main findings; in Section 4, we interpret these results in the context of existing literature; and in Section 5, we summarize key insights and outline recommendations for practice.

2 Methodology

The study was conducted in two secondary schools and one primary school in Varaždin, Croatia, as part of activities marking Safer Internet Day 2024. Activities included firstly the creation of educational games using Wordwall by students for students. These games were then collected on a dedicated website together with links to the prepared surveys (see Fig. 1).

The intervention aimed to engage students in creating and evaluating educational games focused on online safety. Data collection involved two online surveys distributed to participating students via Google Forms. The first survey collected ratings of individual games, while the second assessed students' experiences, motivation, and perceived learning outcomes.

A total of 168 students participated in the evaluation (ages 9 - 14 primary schools students, and 14 - 18 secondary schools students, due to privacy concerns gender was not recorded, but a normal distribution can be assumed), providing both quantitative ratings (on a five-point Likert scale) and open-ended feedback. Game ratings were matched with metadata about the game type and the school of the authoring students.

The following hypotheses were formulated to guide the analysis:

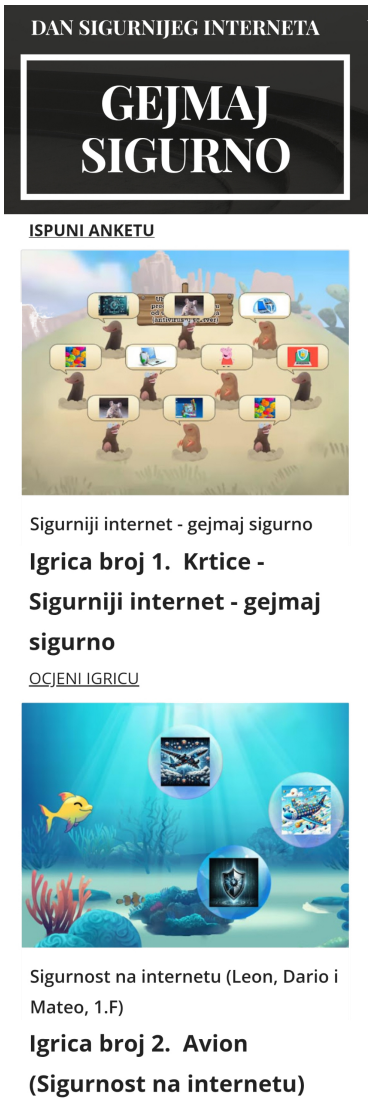


Figure 1. Screenshot of dedicated gaming website

H1: Students who participated in the creation of educational games will report significantly higher motivation to learn through games compared to students who did not participate.

H2: There is a positive correlation between students' perceived enjoyment of playing peer-created games and their perception of the educational value of those games.

H3: Motivation to learn through games and enjoyment of gameplay are significant positive predictors of students' perceived improvement in online safety behavior.

H4: Primary school students will rate the quality of peer-created games significantly higher than secondary school students.

H5: The correlation between motivation to learn through games and perceived learning outcomes is positive and of at least moderate strength ($r > 0.5$).

H6: Participation in game creation does not signifi-

cantly predict perceived improvement in online safety behavior when controlling for motivation and enjoyment of gameplay.

H7: Students' intention to change their online gaming habits is positively associated with their perceived understanding of online risks.

To test these hypotheses, several statistical methods were applied: (1) Descriptive statistics (means, standard deviations) were computed for all survey items, (2) Independent samples *t*-tests and Mann–Whitney U tests were used to examine differences in motivation and perceived quality of games between subgroups (H1, H4); (3) Pearson and Spearman correlation coefficients were calculated to assess associations between motivation, perceived learning, enjoyment, and attitudes towards safety (H2, H5, H7); (4) Multiple regression analysis was conducted to identify predictors of perceived safer behavior after participation (H3, H6); (5) A correlation heatmap was generated to visualize relationships among the main evaluation variables.

All analyses were performed using Python (pandas, scipy, statsmodels, seaborn) and results were interpreted with a significance threshold of $p < 0.05$.

To ensure transparency of the evaluation, both surveys included closed-ended items measured on a five-point Likert scale (e.g., motivation, enjoyment, perceived learning, intention to change behavior) and several open-ended questions for qualitative feedback. Prior to participation, students were informed that the activity was part of Safer Internet Day, that the goal was to reflect on safe online behavior, and that responses would remain anonymous. Teachers introduced the task sequence in the following order: (1) creation of educational games by student groups using Wordwall, (2) sharing of games on the central website, (3) gameplay by peers, and (4) completion of the evaluation surveys. Instructions for filling in the questionnaires were read aloud and displayed on the website, and teachers supervised students while answering to ensure understanding and attention. No formal attention-check items were included, but teachers monitored engagement during data collection. As the primary aim of the study was to capture students' motivation, perceived learning, and attitudes toward online safety, we intentionally focused on self-perceptions rather than objective knowledge tests; future research could complement this with simple pre–post exercises to directly measure learning gains.

3 Results

This section presents the findings of the study structured around the research hypotheses and main research questions. The results include descriptive statistics of game evaluations, differences among groups, correlation analyses, and regression models assessing predictors of perceived impact.

Table 1 shows the descriptive statistics for students' ratings of each game's interest and educational value. The mean interest ratings ranged from 2.80 to 4.15 across different games. The highest-rated games in terms of interest were Game 10 (M=4.15), Game 9 (M=4.11), and Game 6 (M=4.02). The variability in ratings indicates that students perceived substantial differences in the attractiveness and educational potential of individual games.

Further analyses aggregated ratings by game type and by the school of the creator. As shown in Table 2, games in the "Memory" and "Plane" categories received the highest average interest ratings (both M=3.88), while "Moles" games were rated lowest on average (M=3.42). Educational value ratings were generally lower than interest ratings across all categories.

Table 3 presents the average ratings based on whether the game was created by primary or secondary school students. Games created by secondary school students received slightly higher mean interest ratings (M=3.63) compared to those created by primary school students (M=3.33), while educational ratings were similar across groups.

To examine Hypothesis 1, which posited that students who participated in the creation of educational games would report higher motivation to learn through games, mean ratings of the statement "I feel more motivated when learning through games" were compared between students who did and did not participate in game development. As shown in Table 4, the group that took part in creating games reported substantially higher motivation (M=4.45) than those who only played peer-created games (M=3.83). This difference was statistically significant according to the Mann–Whitney U test ($p = 0.002$). The finding provides strong support for Hypothesis 1 and indicates that engaging students in the process of designing educational content can meaningfully increase their intrinsic motivation for learning.

Hypothesis 2 proposed that there would be a positive correlation between students' perceived enjoyment of playing peer-created games and their perception of the educational value of these games. To test this hypothesis, Pearson and Spearman correlation coefficients were calculated between interest ratings and educational ratings across all evaluated games. As presented in Table 8, the analyses revealed a moderate positive association (Pearson $r = 0.535$, Spearman $\rho = 0.545$). This suggests that games perceived as more enjoyable were also more likely to be considered educationally valuable by students. The results confirm Hypothesis 2 and support the notion that engagement and perceived learning are closely intertwined in game-based learning contexts.

Hypothesis 3 stated that motivation to learn through games and enjoyment of gameplay would be significant positive predictors of students' perceived improvement in online safety behavior after the activity. To evalu-

Table 1. Descriptive statistics of game interest and educational ratings

Game	Number of ratings	Interest				Educational value			
		Mean	SD	Min	Max	Mean	SD	Min	Max
1	103	2.91	1.64	1	5	2.67	1.62	1	5
2	59	4.02	1.28	1	5	3.17	1.54	1	5
3	56	3.82	1.25	1	5	3.21	1.44	1	5
4	56	3.43	1.44	1	5	3.09	1.39	1	5
5	56	3.38	1.42	1	5	3.30	1.49	1	5
6	42	4.02	1.26	1	5	3.60	1.31	1	5
7	35	3.54	1.34	1	5	3.34	1.41	1	5
8	50	3.84	1.25	1	5	3.24	1.44	1	5
9	47	4.11	1.05	1	5	3.60	1.17	1	5
10	40	4.15	1.42	1	5	3.95	1.54	1	5
11	36	3.22	1.59	1	5	3.17	1.72	1	5
12	19	3.58	1.26	1	5	2.95	1.31	1	5
13	39	3.74	1.33	1	5	3.00	1.40	1	5
14	28	3.71	1.24	1	5	3.36	1.31	1	5
15	23	3.87	1.25	1	5	3.39	1.31	1	5
16	49	3.98	1.28	1	5	3.27	1.48	1	5
17	28	3.50	1.58	1	5	3.46	1.45	1	5
18	27	3.59	1.15	1	5	3.26	1.23	1	5
19	15	2.80	1.47	1	5	2.87	1.36	1	5
20	34	3.97	1.22	1	5	3.38	1.30	1	5
21	17	3.24	1.44	1	5	3.18	1.59	1	5
22	62	3.58	1.45	1	5	3.40	1.37	1	5
23	49	3.08	1.51	1	5	3.12	1.42	1	5
24	49	3.41	1.44	1	5	3.06	1.55	1	5
25	54	3.46	1.45	1	5	3.30	1.35	1	5
26	41	2.95	1.53	1	5	3.10	1.51	1	5

Table 2. Average ratings of game interest and educational value by game type

Game type	Mean interest rating	Mean educational rating	Number of ratings
Memory	3.88	3.55	75
Plane	3.88	3.21	205
Quiz	3.48	3.23	520
Balloons	3.43	3.28	115
Moles	3.42	3.08	199

ate this, a multiple regression analysis was conducted with perceived safer behavior as the dependent variable and motivation, enjoyment, and participation in game creation as predictors. As shown in Table 9, both motivation (coefficient = 0.407, $p < 0.001$) and enjoyment (coefficient = 0.488, $p < 0.001$) were significant predictors of perceived safer behavior. In contrast, participation in creating games did not significantly contribute to the model ($p = 0.890$). These findings provide clear support for Hypothesis 3, highlighting the central role of students' subjective engagement experiences in fostering perceived improvements in online safety practices.

Hypothesis 4 predicted that primary school students would rate the quality of peer-created games significantly higher than secondary school students. To test this, ratings of the statement "The quality of the peer-created games was high" were compared between the two groups. As shown in Table 7, primary school students reported a higher mean rating (M=3.98) compared to secondary school students (M=3.59). This difference reached statistical significance according to the Mann–Whitney U test ($p = 0.033$). No significant differences were observed between groups in perceived enjoyment or perceived learning from games. These results support Hypothesis 4 and indicate that younger students may perceive peer-created educational content more positively in terms of quality.

Hypothesis 5 stated that the correlation between motivation to learn through games and perceived learning outcomes would be positive and of at least moderate strength (r > 0.5). This relationship was examined using both Pearson and Spearman correlation coefficients. As presented in Table 8, the analyses revealed a very strong positive association: Pearson $r = 0.739$ and Spearman $\rho = 0.767$. These coefficients substantially exceed the threshold specified in the hypothesis and indicate that higher motivation was closely linked to stronger perceptions of having learned. The results therefore provide strong evidence in favor of Hypothesis 5, underscoring the connection between motivation and perceived educational impact in the context of game-based learning.

Hypothesis 6 proposed that participation in game creation would not significantly predict perceived improvement in online safety behavior when controlling for motivation and enjoyment of gameplay. This hypothesis was evaluated within the multiple regression model presented in Table 9. While both motivation and enjoyment emerged as significant predictors

Table 3. Average ratings of game interest and educational value by the school of the game creator

School of the game creator	Mean interest rating	Mean educational rating	Number of ratings
Primary school	3.33	3.21	255
Secondary school	3.63	3.23	859

Table 4. Differences in motivation and perceived usefulness between students who participated in creating the games and those who did not

Statement	Mean (participated)	Mean (did not participate)	p-value (Mann-Whitney)
I feel more motivated when learning through games	4.45	3.83	0.002**
The information is useful for my peers	4.24	3.91	0.054

**p<0.01 indicates a highly significant difference.

($p < 0.001$), participation in creating games was not a statistically significant factor (coefficient = –0.024, $p = 0.890$). This result confirms Hypothesis 6 and suggests that merely contributing to the development of games did not independently influence perceptions of safer behavior once subjective engagement variables were taken into account.

Hypothesis 7 stated that students' intention to change their online gaming habits would be positively associated with their perceived understanding of online risks. This relationship was examined through correlation analysis of the relevant survey items. As visualized in the correlation heatmap in Fig. 2, the association between perceived understanding of risks and intention to change habits was among the strongest observed in the dataset, with Pearson and Spearman coefficients both exceeding 0.70. This finding provides robust support for Hypothesis 7, suggesting that when students recognize online risks, they are more likely to report intentions to adjust their behavior accordingly.

4 Discussion

The present study contributes to the growing body of research exploring the role of participatory game-based learning in fostering motivation, perceived learning, and online safety awareness among school students. The findings provide empirical support for several theoretical propositions regarding the interplay of engagement, co-creation, and educational outcomes in gamified learning environments.

First, the results demonstrate that students who actively participated in designing educational games reported significantly higher motivation to learn compared to those who only consumed peer-created content. This aligns with prior research indicating that co-design and participatory approaches can substantially enhance learners' sense of agency, ownership, and intrinsic motivation (Bamford and Moschini, 2024; Markus Schatten and Marinela Schatten, 2019). Bamford and Moschini (Bamford and Moschini, 2024) highlight that involving students as co-creators fosters a "third space" of learning, where enjoyment and empowerment converge to sustain engagement. Our data confirm that such involvement can be a decisive factor in shaping motivational attitudes.

Second, the moderate to strong correlations between enjoyment of gameplay, perceived educational value, and motivation echo the findings of recent meta-analyses and empirical studies showing that fun and perceived usefulness are interdependent drivers of learning in gamified settings (Nadeem et al., 2023; Videnovik et al., 2023). In particular, Videnovik et al. (Videnovik et al., 2023) emphasize that enjoyment acts as a catalyst for sustained participation and deeper cognitive involvement, a dynamic mirrored in our observed associations ($r > 0.5$) between these constructs.

Third, the regression analysis revealed that perceived safer behavior was primarily predicted by students' motivation and enjoyment, rather than by their mere participation in game creation. This suggests that while participatory design can boost motivation, it is the subjective experience of engagement during gameplay that ultimately shapes attitudes toward behavioral change. This finding is consistent with observations from Schatten and Schatten (Marinela Schatten and Markus Schatten, 2020) who reported variability in outcomes across student groups depending on how gamified experiences were perceived and enacted.

Interestingly, primary school students evaluated the quality of peer-created games more positively than secondary school students. This pattern may reflect developmental differences in expectations, critical thinking, or familiarity with game mechanics. Tomičić and Schatten (Tomičić and Markus Schatten, 2020) have noted that primary students often demonstrate greater openness and satisfaction with gamified learning tools,

Table 5. Correlations between selected statements

Association	Pearson r	Spearman ρ
Motivation and perceived learning	0.739	0.767
Perception of educational value and motivation	0.466	0.440

Table 6. Multiple regression – predictors of perceived safer behavior after the activity

Predictor	Coefficient	p-value
Constant	-0.061	0.861
Motivation for learning through games	0.407	<0.001
Enjoyment of playing the games	0.488	<0.001
Participation in game creation	-0.024	0.890

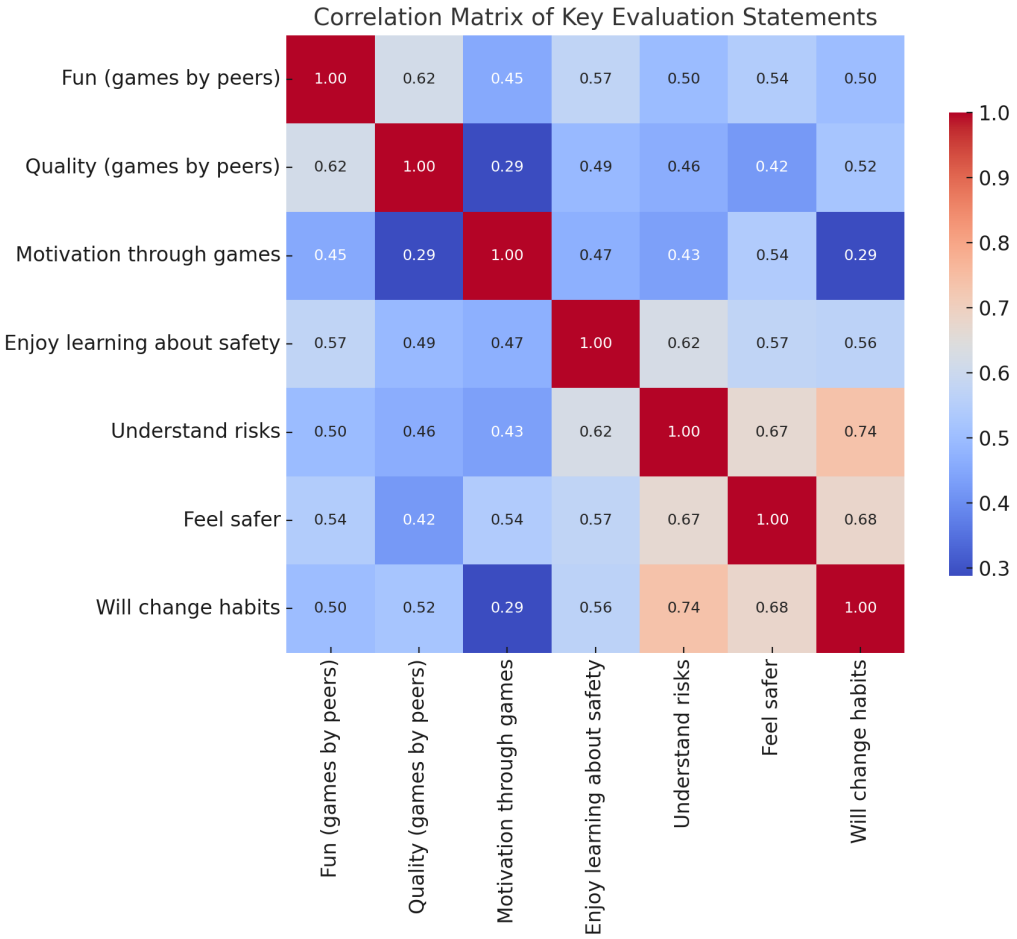


Figure 2. Correlation Matrix of Key Evaluation Statements. The matrix shows pairwise Pearson correlations between main evaluation items.

whereas older students may apply more critical criteria in their evaluations.

Finally, the high ratings of creativity and perceived usefulness of the Wordwall platform underscore the relevance of accessible, user-friendly tools in fostering participatory educational practices. Schatten et al. (Markus Schatten, Đurić, and Ređep, 2022) observed similar patterns of satisfaction with gamified survey platforms in primary education, suggesting that the intuitive design and immediate feedback typical of such tools are instrumental in supporting positive learner experiences.

Overall, this study reinforces the notion that participatory game-based learning is not merely an engaging pedagogical novelty but a robust approach that can meaningfully enhance motivation and perceived learn-

Table 7. Differences in ratings between primary and secondary school students

Statement	Mean (primary)	Mean (secondary)	p-value (Mann-Whitney)
Playing games created by peers was fun	3.90	3.83	0.750
The quality of the peer-created games was high	3.98	3.59	0.033*
I learn more through games than from books	3.98	4.03	0.959

*p<0.05 indicates a statistically significant difference.

Table 8. Correlations between selected statements

Association	Pearson r	Spearman ρ
Motivation and perceived learning	0.739	0.767
Perception of educational value and motivation	0.466	0.440

Table 9. Multiple regression – predictors of perceived safer behavior after the activity

Predictor	Coefficient	p-value
Constant	-0.061	0.861
Motivation for learning through games	0.407	<0.001
Enjoyment of playing the games	0.488	<0.001
Participation in game creation	-0.024	0.890

ing while promoting critical awareness of online safety. Future research should further explore how specific design elements like cognitive agents (Markus Schatten, 2022) and adaptive personalization strategies can optimize the impact of such interventions across diverse educational contexts.

5 Conclusion

This study examined the effects of participatory game-based learning on student motivation, perceived learning, and awareness of online safety risks. The results provide strong evidence that actively engaging students in the creation of educational games can substantially enhance their motivation to learn and their perceptions of the educational value of such activities. Moreover, subjective enjoyment of gameplay emerged as a critical factor predicting students' intention to adopt safer online behaviors.

The findings contribute to a growing consensus that participatory gamification strategies are effective tools for cultivating not only engagement but also meaningful learning outcomes in both primary and secondary educational contexts. The positive correlations observed between motivation, enjoyment, and perceived learning underscore the interdependence of affective and cognitive dimensions of game-based education.

While the present work focused primarily on self-reported measures and relatively short-term outcomes, future research should explore longitudinal effects of participatory game-based interventions and their impact on actual behavior change. Additionally, there is considerable potential in integrating emerging technologies such as artificial intelligence to further personalize and enrich the learning experience. In particular, the deployment of intelligent cognitive agents in educational games may offer new opportunities for adaptive scaffolding, dynamic feedback, and authentic problem-solving experiences (Markus Schatten, 2022).

Continued investigation into the convergence of gamification, educational game co-design, and AI-driven interactivity promises to yield innovative pedagogical models capable of addressing the evolving challenges of digital literacy and safety education.

Acknowledgement

The authors acknowledge that ChatGPT 4o by OpenAI has been used to assist in editing and streamlining sections of this paper.

References

Bamford, J., & Moschini, E. (2024). The third space, student and staff co-creation of gamified informal learning: An emerging model of co-design. *London*

Review of Education, *22*(1), 21. https://doi.org/10.14324/LRE.22.1.21

Henriksen-Bulmer, J., Rosenorn-Lanng, E., Corbin-Clarke, S., Ware, S., Melacca, D., & Fenge, L.-A. (2024). Using game-based learning to teach young people about privacy and online safety. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2023.2265424

Nadeem, M., Oroszlányová, M., & Farag, W. (2023). Effect of digital game-based learning on student engagement and motivation. *Computers*, *12*(9), 177. https://doi.org/10.3390/computers12090177

Ratinho, E., & Martins, C. (2023). The role of gamified learning strategies in students' motivation in high school and higher education: A systematic review. *Heliyon*, *9*(8), e19033. https://doi.org/10.1016/j.heliyon.2023.e19033

Saiger, M. J., Deterding, S., & Gega, L. (2023). Children and young people's involvement in designing applied games: Scoping review. *JMIR Serious Games*, *11*, e42680. https://doi.org/10.2196/42680

Schatten, M. [Marinela], Inkret-Martinčević, L., & Košćak, M. (2024). Gejmaj sigurno – izradi svoju igru! [Digital journal for education professionals]. *Pogled kroz prozor*.

Schatten, M. [Marinela], & Schatten, M. [Markus]. (2020). A comparative study of gamification in programming education in a croatian high school. *MIPRO 2020 43rd International Convention*, 785–789.

Schatten, M. [Markus]. (2022). *O_HAI 4 games - d4.2. case study 2 - cognitive agents & gamification* (tech. rep.). University of Zagreb, Faculty of Organization and Informatics.

Schatten, M. [Markus], Đurić, B. O., & Ređep, N. B. (2022). A gamified platform for surveying digital maturity of primary schools. *Proceedings of the 24th CARNET Users Conference (CUC 2022)*.

Schatten, M. [Markus], & Schatten, M. [Marinela]. (2019). Gamification of game programming education: A case study in a croatian high school. In *Proceedings of 30th central european conference on information and intelligent systems (ceciis 2019)* (pp. 13–18). Faculty of Organization; Informatics, University of Zagreb.

Tomičić, I., & Schatten, M. [Markus]. (2020). A conceptual network analysis of gamification practices in primary and secondary education. *MIPRO 2020 43rd International Convention*, 1091–1097.

Veočić, A., Nikolić, M., Ćaleta, I., Čičić, T., & Đurković, M. (2024). *Sve što želite i trebate znati o videoigrama* (Publication for parents and professionals) (Published for Safer Internet Day (February 2024)). Centar za nestalu i zlostavljanu djecu. Osijek, Croatia.

Videnovik, M., Vold, T., Kiønig, L., Madevska Bogdanova, A., & Trajkovik, V. (2023). Game-based learning in computer science education: A scoping literature review. *International Journal of STEM Education*, *10*(1), 54. https://doi.org/10.1186/s40594-023-00447-2

Wang, K., Liu, P., Zhang, J., Zhong, J., Luo, X., Huang, J., & Zheng, Y. (2023). Effects of digital game-based learning on students' cyber wellness literacy, learning motivations, and engagement. *Sustainability*, *15*(7), 5716. https://doi.org/10.3390/su15075716

Cyber-Physical Systems and Intelligent Energy

Zvonimir Šimić, Danijel Topić, Marina Dubravac, Nemanja Mišljenović
The Potential of Using a Small PV Battery System to Overcome Electricity Supply Disruptions

Milan Daňo, Martin Juhás, Patrik Gašparovič, Bohuslava Juhásová, Fedor Burčiar
Use of Piezoelectric Technology for Fluid Valves as an Energy Saving Alternative

Lukrecia Vulić, Ivan Aleksi, Tomislav Matić
Comparative Analysis of Bandpass IIR Filters on Microphone PDM Signal for FPGA Implementation

Juraj Ďuďák, Juraj Repka
Deep-Learning and Stereo Camera Based System for Real-Time Virtual Safety Boundaries

Ivan Kolarić, Karlo Pepelnik, Dino Vlahek, Dijana Oreški, Jelena Gusić Munđar, Bojan Žugec, Nikola Kadoić
Application of Artificial Intelligence and Internet of Things in Rehabilitation and Health Recovery

Lea Masnec, Vjeran Strahonja
Modeling and Construction of Circular Economy Digital Twins Based on System Dynamics

The Potential of Using a Small PV Battery System to Overcome Electricity Supply Disruptions

Zvonimir Šimić, Danijel Topić, Marina Dubravac, Nemanja Mišljenović
Josip Juraj Strosssmayer University of Osijek
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Kneza Trpimira 2B, 31000 Osijek, Croatia
{zvonimir.simic, danijel.topic, marina.dubravac, nemanja.misljenovic}@ferit.hr

Abstract. *The potential of using small PV battery systems to satisfy electricity demand during disruptions in the electricity supply is investigated in this paper. The analysis was done for the year 2023 which is characterized with more disruptions caused by extreme weather in Croatia. The day-ahead scheduling is done according to the hourly forecast data considering planned and potential interruptions. Evaluation of day-ahead scheduling is done through simulation according to hourly actual data considering all interruptions. According to results, the PV battery system is capable to meet electricity demand during 85% of the time in the cable power grid and during 83% of the time in the overhead power grid when the main grid is not available due to grid interruptions.*

Keywords. Disruptions, PV battery system, day-ahead scheduling, electricity demand, reliability indices

1 Introduction

Modern society is highly dependent on electricity, and even short electricity supply disruptions cause discomfort. In the first place, it is thermal discomfort due to lack of heating or cooling (Hachem-Vermette and Yadav, 2023). In addition, people are unable to charge phones, watch movies, or prepare dinner, which is the main willingness to pay, as it is not possible to objectively assess financial loss in households during disruptions in electricity supply (Carlsson et al., 2021). Extreme weather conditions, such as thunderstorms or strong winds, are the main cause of unplanned interruptions (Alpay et al., 2020). However, grid interruptions can often be planned to ensure a safe environment for electricians during regular electrical network maintenance (Crognier et al., 2021). The reliability of the distribution system is evaluated with the following index values (Chitumodhu and Pilly, 2024):

- System Average Interruption Frequency (SAIFI)
- System Average Interruption Duration (SAIDI)
- Customer Average Interruption Duration (CAIDI)
- Energy Not Supplied (ENS)

Existing solutions for electricity supply during interruptions are Uninterrupted Power Supplies (UPS) and backup generators. The drawbacks of the mentioned systems are the potentially high running and maintenance costs. Furthermore, the use of backup generators powered by fossil fuels is inconsistent with the decarbonization of the energy sector.

Researchers are focused on the implementation of carbon-free technologies to ensure electricity supply during interruptions. The electric vehicle (EV) can meet electricity demand during interruptions while ensuring enough battery capacity for the next day (Zhang et al., 2023). Vehicle-to-home (V2H) technology is suitable for providing backup power for 12 hours in mild climate conditions, the resilience depends on the start time of interruption and daily EV requirements (Liu et al., 2025). The island operation of grid-tied photovoltaic (PV) systems is not allowed in EU countries (Miao, 2024). On the other hand, the island operation of hybrid PV battery storage (BS) microgrids during grid interruptions is possible (Grant and Clark, 2024). There are practical cases of hybrid microgrid applications in Spain and Portugal. To meet a regulatory framework, physical disconnection of the microgrid from the main grid is required (Giorgi, 2025). The main advantages of hybrid PV battery systems are modularity that provides a better match with electricity demand and lower investment and running costs than with a backup generator (Chatterji et al., 2021). The contribution of this paper is summarized as follows:

- Generation of random interruptions through 50 different cases in cable and overhead power grid
- Each case represents one year with random frequency, durations and start times of interruptions
- Consideration of planned interruptions and extreme weather forecast in day-ahead scheduling
- Evaluation of day-ahead scheduling through simulation with consideration of all grid interruptions

This paper is organized as follows. In Section 2, a summary of the power reliability indices is given, while in Section 3, methodology and input data are described. In Section 4, results are presented and discussed, while the last section is the conclusion of the paper.

2 A Summary of the Power Reliability Indices in Croatia

The decrease in the SAIFI and SAIDI value in Croatia is noticeable between 2011 and 2022 (HEP-ODS, 2024). In 2011 the SAIFI was 4.84 and in 2022 it was 2.46, as can be seen in Fig. 1. The SAIDI was reduced from 9.32 h in 2011 to 4.31 h in 2022 as shown in Fig. 2. The power reliability indices are improved due to significant investments in power infrastructure.

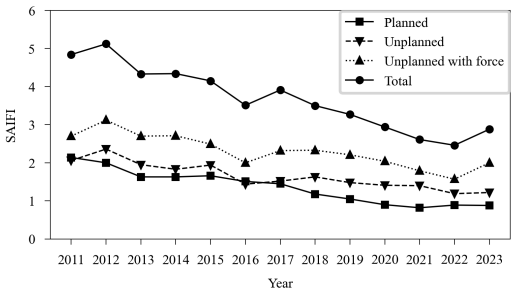


Figure 1. SAIFI in Croatia from 2011 to 2023

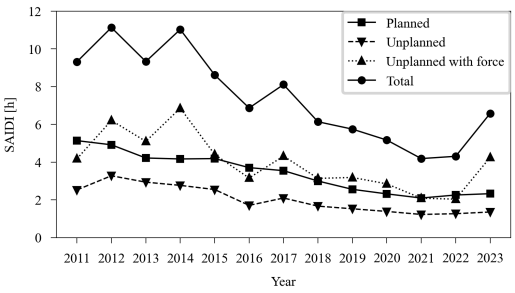


Figure 2. SAIDI in Croatia from 2011 to 2023

In 2023, the SAIFI value increased to 2.88, or 0.42 compared to the previous year. The SAIDI value is increased to 6.58 h in 2023 which is 2.27 h more than in 2022. The main reason for this increase was the strong storms during the summer of 2023. The damage to the infrastructure was significant and a long time was required to restore the power supply (HEP-ODS, 2024).

The general power reliability standard in Croatia (*Pravilnik o uvjetima kvalitete opskrbe električnom energijom*, 2022) defines 3 long interruptions per customer for the cable power grid and 6 long interruptions per customer for the overhead power grid. The duration of interruptions per customer is 400 minutes or 6.67 h during a year in the case of cable power grid and 700 minutes or 11.67 h during a year in the case of the overhead power grid.

The Croatian distribution system operator (HEP-ODS) is divided into 21 distribution areas. In 2023 the SAIFI standard for the cable power grid is not met in 2 while the SAIFI standard for the overhead power grid is not met in 7 distribution areas according to Fig. 3. According to Fig. 4 in 2023 the SAIDI standard for the cable power grid is not met in 1 while the SAIDI standard for the overhead power grid is not met in 9 distribution areas (*Izvještaj o kvaliteti opskrbe za 2023. godinu*, 2024). It is expected as the cable power grid is protected from extreme weather conditions, unlike the overhead power grid.

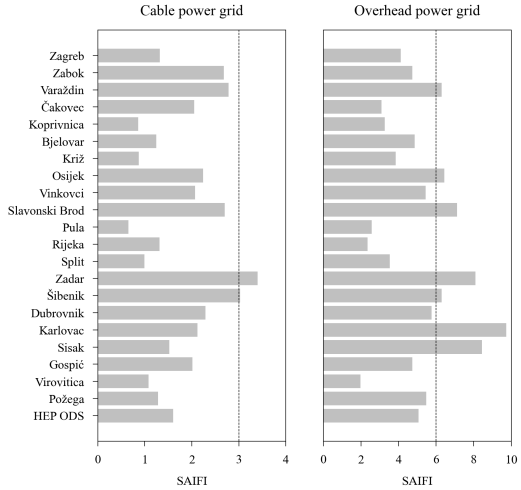


Figure 3. SAIFI for each distribution area in 2023

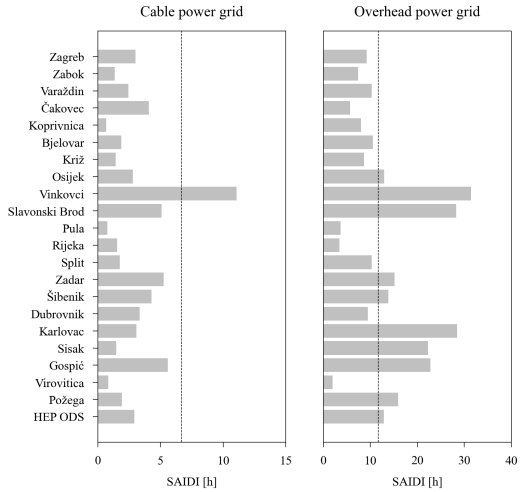


Figure 4. SAIDI for each distribution area in 2023

3 Methodology and Input Data

Although power reliability indices are improved over years, extreme weather conditions can lead to increase of power interruptions. PV battery systems can provide electricity supply even during grid interruptions, but day-ahead scheduling is required due to limited battery capacity. The flow chart of methodology applied in this work is presented in Fig. 5.

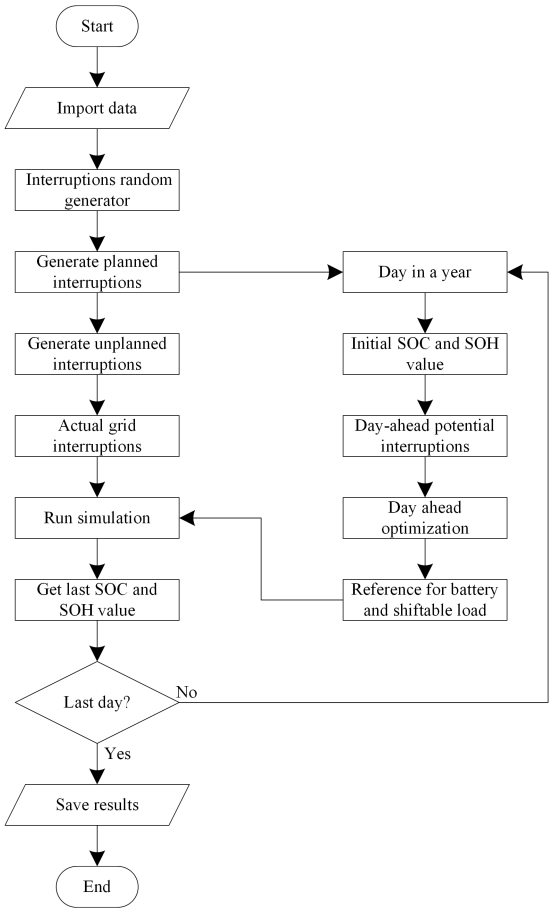


Figure 5. Flowchart of applied methodology

Input data required for simulation is weather and electricity consumption data while required input parameters are reliability indices for the subject distribution area and PV battery system parameters. First step is generation of random grid interruptions while satisfying average power reliability indices.

A day-ahead optimization is done with forecast weather and load hourly data with considered planned interruptions and possible interruptions due to extreme weather. Results of optimization are referent battery charging and discharging currents and start times of two time-shiftable appliances. Then simulation is done with actual weather and load hourly data with all interruptions considered. In normal operation battery is charging and discharging according to reference from day-ahead optimization. During grid interruption battery charging and discharging depends on the amount of electricity generation from PV system and power demand. Last battery state of charge (SoC) and battery state of health (SoH) values are initial SoC and SoH values for next day optimization. The algorithm is repeated for each day during a year and when last day is simulated, results for entire year are saved.

The methodology is applied to a household with a PV battery system located in continental part of Croatia in distribution area "Osijek".

3.1 Weather Data

The weather data is downloaded from Open-Meteo (Zippenfenig, 2023) open-source weather API for location Osijek with the coordinates 45.55°N, 18.69°E for year 2023. Two datasets are downloaded, the first is forecast weather data required for day-ahead planning and the second is actual weather data from satellite.

For the estimation of the electricity generation from the PV system solar irradiation and temperature data are required. Electricity consumption of air conditioning is estimated according to temperature data. Fig. 6 represents forecast and actual hourly solar irradiation data for one summer week while Fig. 7 represents forecast and actual hourly temperature data for the same summer week.

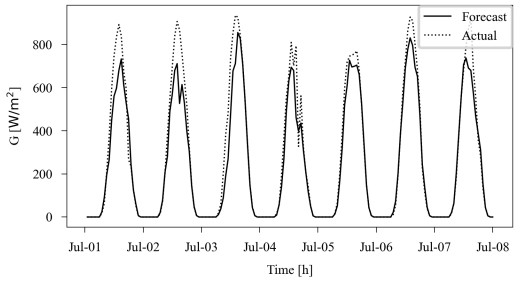


Figure 6. Forecast and actual hourly solar irradiation data from July 1 to July 7

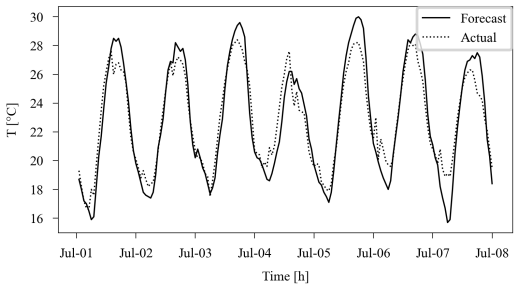


Figure 7. Forecast and actual hourly temperature data from July 1 to July 7

According to (*Izvještaj o kvaliteti opskrbe za 2023. godinu*, 2024) 39% of the unplanned interruptions in 2023 are associated with extreme weather conditions that can be predicted and included in day-ahead planning. Thunderstorms are predicted according to convective available potential energy (CAPE) day-ahead forecast while strong winds are predicted according to wind gusts day-ahead forecast. Fig. 8 represents day-ahead hourly forecast of CAPE and wind gusts during 2023. Hours with CAPE value above 1 kJ/kg and wind gusts above 40 km/h are considered as hours with potential grid interruptions in day-ahead planning.

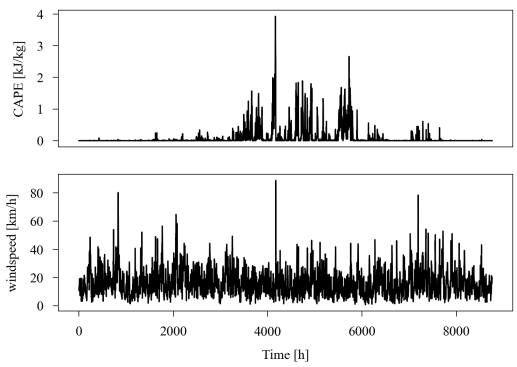


Figure 8. The forecast of CAPE and wind gusts for each hour during 2023

3.2 Electricity Consumption Data

Electricity consumption data is obtained from measured electricity consumption of separate appliances. Electricity consumption of different home appliances is measured with VOLTCRAFT SEM5000 energy meter with integrated data logger. In the continental part of Croatia, electricity is not dominant heating source for households and heating requirements are not included in estimation of electricity consumption. On the other hand, cooling requirements during summer are considered. Forecast and actual hourly electricity consumption during one summer week is presented in Fig. 9.

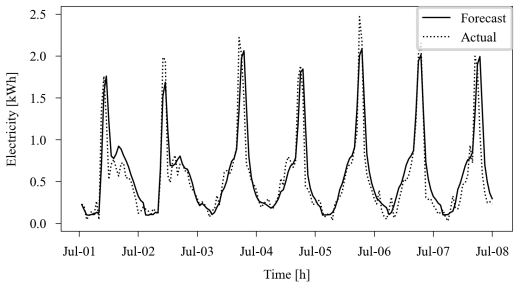


Figure 9. Forecast and actual hourly electricity consumption from July 1 to July 7

3.3 Distribution Area Reliability Indices

The total number of electricity customers at the end of 2022 in distribution area "Osijek" is 154 948. According to the type of grid interruption, SAIFI is 1.59 for planned and 2.69 for unplanned interruptions while SAIDI is 4.07 h for planned and 3.68 h for unplanned interruptions. CAIDI value is 2.56 h for planned and 1.37 h for unplanned interruptions. Reliability indices according to the type of power grid are presented in Table 1. Values for total interruptions for cable and overhead power grid are from (*Izvještaj o kvaliteti opskrbe za 2023. godinu*, 2024) and values according to grid and interruption type are estimated.

Table 1. Reliability indices according to grid type and interruption type in 2023

Type	SAIFI	SAIDI [h]	CAIDI [h]
Cable	2.24	2.8	1.25
Planned	0.83	1.47	1.97
Unplanned	1.41	1.33	0.94
Overhead	6.43	12.95	2.01
Planned	2.39	6.80	2.85
Unplanned	4.04	6.15	1.52

3.4 Small PV Battery System

A small PV battery system consists of 4 kW PV power plant and 5 kWh battery storage which are connected to electricity consumption and the main grid over the bidirectional power converter, as can be seen in the schematic diagram presented in Fig. 10. The main technical specifications of PV modules and battery, according to (Bluesun, 2025) and (Fuji, 2023), are presented in Table 2.

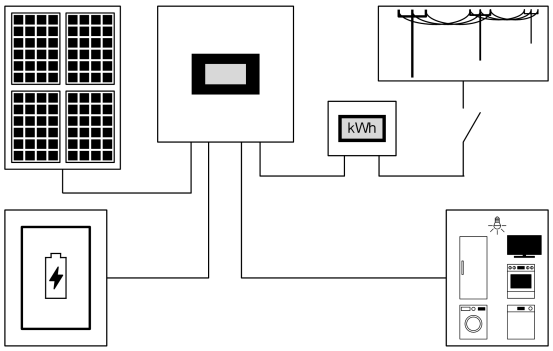


Figure 10. Schematic diagram of residential PV battery system

Table 2. Specifications of PV module and battery

PV module	
Technology	PERC
Rated power	500 W
Efficiency	0.213
Length	2.056 m
Width	1.140 m
NOCT	42.3 °C
Temp. coeff.	0.34 %/°C
Battery	
Technology	$LiFePO_4$
Rated capacity	100 Ah
Rated voltage	51.2 V
Max. DoD	80%
Max. current	60 A
Cycle life (0.2C)	6000
Cycle efficiency	0.92

4 Results

In this section results are presented for cable and overhead power grid. The number of interruptions is conditioned with SAIFI values according to the grid and interruption type in Table 1. The average durations are conditioned with the CAIDI values in Table 1. The hours in which interruptions occur depend on the type of interruption. Planned interruptions are limited to workdays, without rain and strong winds, and to work time between 08:00 and 15:00. The start times of unplanned interruptions associated with extreme weather conditions are defined based on events with reported damage to the electrical network, while the start times of other unplanned interruptions are random.

4.1 Results for cable power grid

The grid interruptions for 50 cases are generated and presented in Fig. 11. The total duration of all grid interruptions is 132 h and divided by 50 gives 2.64 h per case. The duration of planned and unplanned interruptions is 66 h total or 1.32 h per case. These values are similar to the SAIDI values in Table 1.

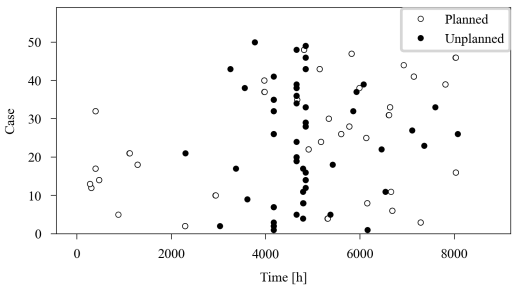


Figure 11. Occurrence of planned and unplanned interruptions in cable power grid

The capability of a PV battery to meet demand during interruptions is shown in Fig. 12. During planned interruptions, electricity demand is covered in all cases, which is expected as planned interruptions are considered in the day-ahead optimization.

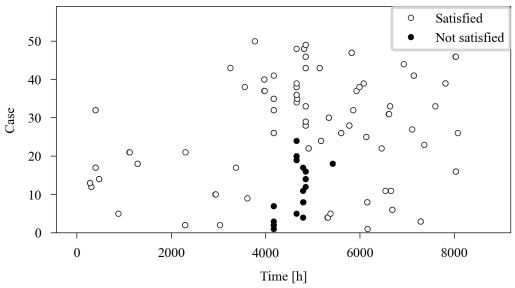


Figure 12. The capability of PV battery system to satisfy electricity demand during interruptions in cable power grid

However, during unplanned interruptions, the electrical supply from the PV battery is provided for 46 h or 70% of time when the main grid is not available due to unplanned interruptions. During all interruptions, the PV battery can meet the electricity demand for 112 h or 85% of the total time without the main grid. In a single case, the demand requirements are satisfied during 2.24 h from 2.64 h without the main grid.

4.2 Results for overhead power grid

The occurrence of planned and unplanned grid interruptions is shown in Fig. 13. The total duration of all interruptions is 640 h. When this total time is divided by 50, it is 12.8 h in a single case. The duration of planned interruptions is 335 h or 6.7 h in a single case, while the duration of unplanned interruptions is 305 h or 6.1 h in a single case. The values obtained are close to the SAIDI values in Table 1.

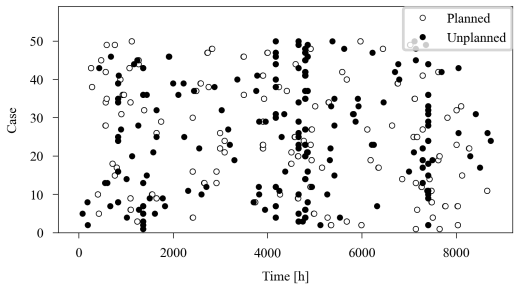


Figure 13. Occurrence of planned and unplanned interruptions in overhead power grid

The capability of a PV battery to provide electrical supply during grid interruptions is presented in Fig. 14. Demand requirements are met during all planned interruptions as expected. However, during unplanned interruptions, electricity demand is satisfied for 198 h or 65% of time without the main grid. During all interruptions, the demand requirements can be met for 533 h or 83% of time without the main grid. In a single case, the PV battery is capable of satisfying demand during 10.66 h from 12.8 h without the main grid.

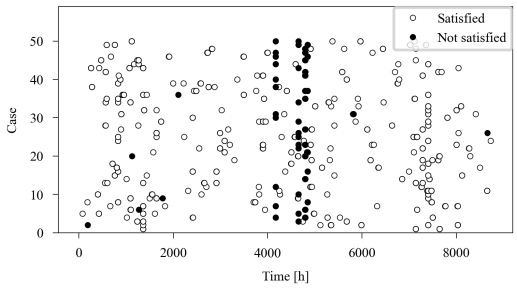


Figure 14. The capability of PV battery system to satisfy electricity demand during interruptions in overhead power grid

5 Conclusion

The PV battery has great potential to overcome electricity supply disruptions. The day-ahead planning is necessary to ensure electricity supply even when the main grid is not available during grid interruptions. Planned interruptions can be included in day-ahead planning, as well as potential disruptions associated with extreme weather conditions.

The results show that the PV battery is capable of satisfying the demand for electricity during all planned interruptions, while during unplanned interruptions, the demand for electricity is satisfied for 70% of the time in the cable power grid and 65% of the time in the overhead power grid. Most unplanned interruptions when demand is not satisfied occur in summer during thunderstorms. Although thunderstorms are considered in day-ahead planning, interruptions occur in the afternoon when the electricity demand in households usually increases and due to limited battery current and capacity, demand is not satisfied.

To ensure more electricity supply during interruptions, the increase of battery capacity and implementation of demand response should be considered in future work. There is also a potential for using EVs during grid interruptions through V2H concept due to the large available battery capacity in EVs.

Acknowledgments

This work was supported by the Croatian Science Foundation under the project "Young researchers' career development project – training of doctoral students" (DOK-2021-02-1257) and by the European Union's Horizon Europe Framework Programme: HORIZONWIDERA-2023-ACCESS-04, Pathways to Synergies – Coordination and Support Actions - under the project name SynGRID - Creating synergies in Widening countries on the topic of low-voltage grid management (grant number 101160145). The funding bodies had no involvement in the preparation of the manuscript.

References

Alpay, B. A., Wanik, D., Watson, P., Cerrai, D., Liang, G., & Anagnostou, E. (2020). Dynamic modeling of power outages caused by thunderstorms. *Forecasting*, *2*(2), 151–162.

Bluesun. (2025). *Bsm500pm5-72sb datasheet*. https://www.enfsolar.com/pv/panel-datasheet/crystalline/48199 (Accessed: 26.04.2025).

Carlsson, F., Kataria, M., Lampi, E., & Martinsson, P. (2021). Past and present outage costs – a follow-up study of households' willingness to pay to avoid power outages. *Resource and Energy Economics*, *64*, 101216.

Chatterji, E., Anderson, K., & Bazilian, M. D. (2021). Planning for a resilient home electricity supply system. *IEEE Access*, *9*, 133774–133785.

Chitumodhu, B., & Pilly, G. (2024). Reliability evaluation of radial distribution system considering common mode failures and different weather effects. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, *7*, 100394.

Crognier, G., Tournebise, P., Ruiz, M., & Panciatici, P. (2021). Grid operation-based outage maintenance planning. *Electric Power Systems Research*, *190*, 106682.

Fuji. (2023). *Residential ess lithium-ion battery 5100*. https://fuji-solar.com/wp-content/uploads/2023/05/FujiCFE-5100-brochure.pdf (Accessed: 26.04.2025).

Giorgi, M. (2025). *When the grid fails: Microgrids for competing with intelligence and energy autonomy*. https://strategicenergy.eu/microgrids-for-blackout/# (Accessed: 21.05.2025).

Grant, E., & Clark, C. E. (2024). Hybrid power plants: An effective way of decreasing loss-of-load expectation. *Energy*, *307*, 132245.

Hachem-Vermette, C., & Yadav, S. (2023). Impact of power interruption on buildings and neighborhoods and potential technical and design adaptation methods. *Sustainability*, *15*(21).

HEP-ODS. (2024). *Godišnje izvješće o sigurnosti opskrbe u distribucijskom sustavu za 2023. godinu s projekcijom za 2024. godinu*.

Izvještaj o kvaliteti opskrbe za 2023. godinu. (2024).

Liu, S., Vlachokostas, A., & Kontou, E. (2025). Leveraging electric vehicles as a resiliency solution for residential backup power during outages. *Energy*, *318*, 134613.

Miao, M. (2024). *How to achieve anti-islanding in inverters with energy storage solutions*. https://luxpowertek.com/blog/how-to-achieve-anti-islanding-in-inverters-with-energy-storage-solutions/ (Accessed: 21.05.2025).

Pravilnik o uvjetima kvalitete opskrbe električnom energijom. (2022).

Zhang, W., Wang, Y., Xu, P., Li, D., & Liu, B. (2023). Evaluation and improvement of backup capacity for household electric vehicle uninterruptible power supply (ev-ups) systems. *Energies*, *16*(12).

Zippenfenig, P. (2023). *Open-meteo.com weather api*. https://open-meteo.com/ (Accessed: 08.04.2025).

Use of Piezoelectric Technology for Fluid Valves as an Energy Saving Alternative

Milan Daňo, Martin Juhás, Patrik Gašparovič, Bohuslava Juhásová, Fedor Burčiar
Slovak University of Technology in Bratislava, Faculty of Materials Science and Technology in Trnava
Institute of Applied Informatics, Automation and Mechatronics
Jána Bottu 2781/25, 917 24 Trnava, Slovakia
{milan.dano, martin_juhas, patrik.gasparovic,bohuslava.juhasova, fedor.burciar}@stuba.sk

Abstract*. Fluid control valves are an integral part of production equipment. With the development of new technologies such as piezoelectric actuators, new control possibilities are opening up. Piezoelectric actuators are characterized by low power consumption and noise with very fast and precise movement. In light of the new machine building philosophies of Industry 4.0, machine building components must be more versatile and consume less energy. The standard technology of so-called solenoid valve types is limiting in its design, operating principle and power consumption. Piezoelectric actuators offer wider valve control possibilities and have lower energy consumption. This paper describes the piezoelectric phenomenon itself, piezoelectric materials and their properties, actuators and the possibility of using them in fluidic control valves, indicating their main advantages over conventional valves.*

Keywords. piezoelectric phenomenon, piezoelectric materials, piezoelectric drives, fluid control valves, energy consumption

1 Piezoelectric Phenomenon and its Properties

The piezoelectric phenomenon was discovered and confirmed by the brothers Pierre and Jacques Curie in 1880 in experiments with materials such as Siegnett's salt or tourmaline. The basis of this phenomenon is the ability of so-called piezoelectric materials to electrically polarize - to generate an electric charge - when stressed by pressure, tension, bending, or twisting. Experiments have confirmed both the direct piezoelectric phenomenon and the indirect (inverse) piezoelectric phenomenon.

In direct piezoelectric phenomenon, the mechanical stress on the piezoelectric material results in the displacement of ions in the crystals of piezoelectric materials, resulting in the generation of an electric charge. This property of piezoelectric materials is used in practice to sense physical values such as pressure and deformations, or to provide energy for various devices in automation or biomedicine. "Over the past few decades, the functionality of energy harvesters based on piezoelectric energy from mechanical vibrations has been extensively investigated in energy technologies. In the case of low power electronic devices, energy harvesting is generally in the mW or μW range. In addition, the majority of these applications involve equipment that has the ability to operate both indoors and outdoors in such a way that it does not have a significant dependence on climatic conditions. Energy harvesting is a promising area that enables Internet of Things (IoT) devices to generate electricity by absorbing energy from the environment" (Bobic et al.2023).

The indirect (inverse) piezoelectric phenomenon manifests itself as a deformation of the piezoelectric material that occurs on the crystals when they are connected to an electric voltage or when they are placed in an electric field. This phenomenon is used in practice as a drive for precise and fast positioning such as focusing camera lenses. The use of the indirect piezoelectric phenomenon has also been elaborated and used to actuate fluidic valves. It can also be used to produce sound and ultrasound generators (Tamburrano et al., 2020).

The generation of charge on the surface of crystals, or the deformation of the material when a voltage is applied to the material, is called the piezoelectric properties of the material. Piezoelectric properties can be seen on acentric crystals, crystals without central symmetry. Therefore, when deformed in the vertical direction, they change the distance of positive and negative charge. The charges move away from each other and their centres of gravity no longer match - an electric dipole is formed (Fig. 1) (Budkova, 2022), (What is Piezoelectricity? 2024).

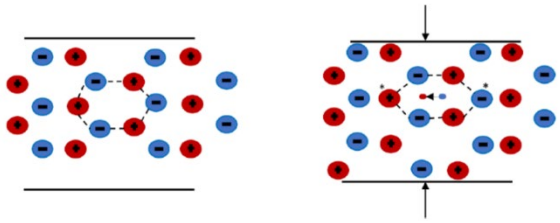


Figure 1. Structure of a piezoelectric material (Budkova, 2022)

The correct function of a piezoelectric material is also influenced by the so-called Curie temperature (Tc). If this temperature is exceeded the material loses its typical piezoelectric properties. At temperatures higher than Tc, piezoelectric materials undergo significant changes in their internal structure - the arrangement of atoms in the crystal lattice is deformed. Grains with a slightly deflected central ion give rise to symmetrical grains in which no electric dipoles are formed and spontaneous polarization of the material occurs (Budkova, 2022).

The size of the piezoelectric phenomenon can be described by piezoelectric constants. Important parameters of piezoelectric materials include:

- Piezoelectric charge constant d_{ij} which gives the magnitude of the generated charge per area with respect to the applied mechanical stress and the area ($C.N^{-1}$). For indirect piezoelectric phenomena, this constant is equal to the mechanical deformation of the area when a voltage is applied ($m.V^{-1}$). This constant is very important in the selection of a suitable material for actuators (Budkova, 2022), (Piezoelectric charge constants, 2024).
- Piezoelectric voltage constant g_{ij} which indicates the magnitude of the electric field generated by the piezoelectric material with respect to the applied mechanical stress. Materials with high voltage constant have large output voltage and are mainly used for sensors (Budkova, 2022), (Piezoelectric charge constants, 2024).

2 Piezoelectric Materials

In order for a material to be labelled as piezoelectric, it must have a specific structure - a non-symmetrical arrangement of elements. The piezoelectric phenomenon is not the same for all materials. Some like natural materials it is very small. For this reason, research has focused on making materials that have a larger piezoelectric effect after modification.

To enhance the properties of the piezoelectric phenomenon, it is necessary for some materials to arrange (polarize) the randomly acentric grains in the crystal structure. The poling is done by placing them in an electric field in which the temperature is raised and the structure is subsequently cooled. The result is a material with permanent piezoelectric properties. It is an electrical reorientation and alignment of electric dipoles (Fig. 2) (Budkova, 2022), (Pinto et al., 2017).

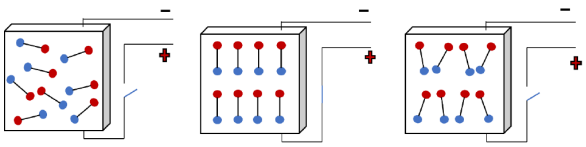


Figure 2. Feroelctrical ceramics before, during and after polarisation (Budkova, 2022)

2.1 Monocrystalline Structures

Materials such as quartz, tourmaline and Siget salt are natural single crystalline materials with an already built-up dipole structure. In the rest state, these materials are electrically neutral. When they are mechanically stressed (squeezed, stretched, twisted, bent), ions are displaced and thus charged - forming an electric dipole. Quartz is one of the best-known natural materials with an asymmetric structure that is alternately composed of positive charges (silicon ions) and negative charges (oxygen ions). Its great advantage is the small dependence of the piezoelectric coefficient on temperature, unlike other crystalline substances, which have non-negligible pyroelectric properties. In addition to this property, the stability of its lower form up to 573°C also predisposes it to work at higher temperatures (Budkova, 2022), (Plass, 2009), (Overview of minerals and rocks, 2024)

2.2 Ceramic Piezoelectric Materials

During the 1950s, new ceramic piezoelectric materials were developed, enabling their use in several applications such as sensors, actuators and piezoelectric motors. These are polycrystalline ceramic materials that are paraelectric in the normal state. They acquire piezoelectric properties only after polarisation in a strong electric field. These ceramic materials include, e.g., barium titanate $BaTiO_3$, lead zirconate $PbZrO_3$ or the most widely used material is the PZT ceramic $Pb(Zr_xTi_{1-x})O_3$ (Plass, 2009).

2.3 Organic Piezoelectric Polymers

The third group of piezoelectric materials are organic piezoelectric polymers, the most widely used are polyvinylidene-fluoride PVDF and difluoro-polyethylene PVF2. They are characterised by a piezoelectric coefficient up to several times higher than quartz crystals. They are used as sensors and most often as sensitive acoustic and vibration sensors. Their production is also considerably simpler than e.g. ceramic piezoelectric materials (Plass, 2009).

3 Piezoelectric Actuators

In industry and especially in scientific fields such as precision machining, precision measurement, biological and genetic engineering, micro/nano positioning systems with micro/nano precision are very important. Actuators refer to functional devices that can perform specific motions such as linear, or rotational motion (Cheng, Li, 2022). Actuators commonly used in industry, such as pneumatic, hydraulic or electric actuators, have the ability to perform longer strokes with greater forces, but have larger size and power consumption, but mainly lower positioning accuracy and speed. For applications that require more precise positioning, high speed and low power consumption, piezoelectric actuators are more efficient. Piezoelectric actuators are characterized by compact size, light weight, high accuracy, fast response, good control characteristics, high energy density, low power consumption, and are free from magnetic field interference (Cheng, Li, 2022).

Based on their working principle, piezoelectric actuators can be divided into direct-driving piezoelectric actuators and stepping piezoelectric actuators. In direct-driving types, the deformation of the piezoelectric element is directly used to move the output mechanism. The working stroke of these actuators is small. Stepping types use a stepping motion mode whereby larger strokes can be realized. For the purpose of controlling pneumatic valves, directly controlled piezoelectric actuators have the best prerequisites, which can be further divided into actuators with and without a amplification mechanism as can be seen in the Fig. 3 (Cheng, Li, 2022).

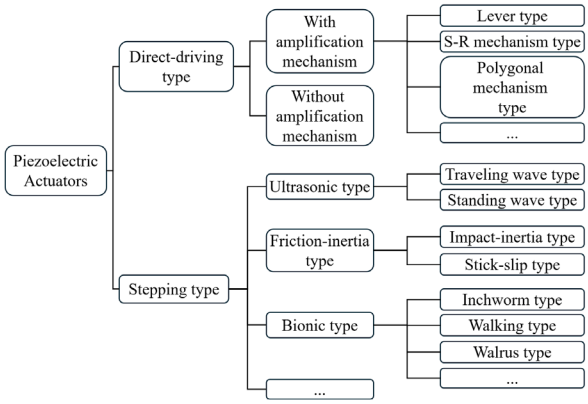


Figure 3. Piezoelectric actuator types (adapted from Cheng, Li, 2022)

4 Piezoelectric Composite Structures

In applications that use direct-driving piezoelectric actuators, piezoelectric materials form single or multilayer structures. From a mechanical point of view it can be either a membrane, a built-in beam fixed on one side (bender), or multilayered microactuators (Budkova, 2022), (Plass, 2009).

4.1 Unimorphic Structure

This structure is made up of a metal plate on which a layer of piezoelectric material is deposited and has power connections (Fig. 4). They are most commonly used in electroacoustic transducers and sensors of non-electrical quantities (Plass, 2009), or in precision small-volume dispensers (Hradil, 2011).

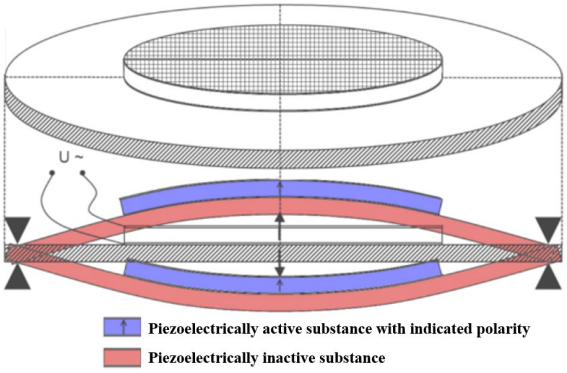


Figure 4. Unimorphic membrane (Hradil, 2011)

4.2 Bimorphic Structure

This type of structure consists of two piezoelectric layers that can be combined with other materials such as thin metallic sheets or plastics. According to the way the voltage is applied, bimorph structures are divided into serial and parallel types. Parallel bimorph structures have twice the deflection (deformation) of serial bimorph structures at the same voltage. In the serial circuit, the supply voltage is applied to the edges of two connected piezoelectric wafers and in the parallel circuit, one pole is connected to a metal electrode sandwiched between two piezoelectric wafers and the other pole is connected to the edges of these wafers (Fig. 5). For bimorphic structures, the stroke is in the range of hundreds to thousands of microns and the bending force in Newtons (Plass, 2009). Bimorphic structures are suitable for the design of bending beams.

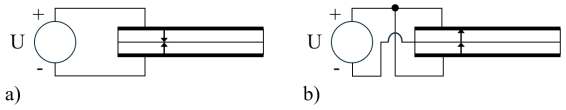


Figure 5. a) Serial bimorphic structure, b) Parallel bimorphic structure (Plass, 2009)

4.3 Polymorphic Structure

These structures consist of more than two piezoelectric layers, thereby increasing efficiency and reducing

electrical voltage at the same power. Movement occurs in the direction of the layered piezoelectric elements (Fig. 6). These structures have working deflections in tens of microns and are capable of exerting a force in hundreds up to thousands of Newtons (Plass, 2009). Polymorphic structures are suitable for the design of layered micro actuators.

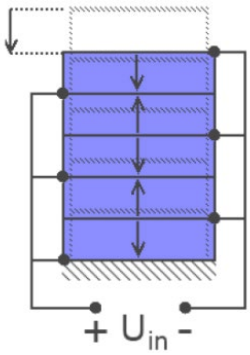


Figure 6. Polymorphic structure (Hradil, 2011)

5 Use of Piezoelectric Actuators to Control Fluidic Valves

Fluidic valves are widely used in all parts of our life. Starting with medical equipment thru daily used gadgets up to the industry. The most used technology because of its versatility is pneumatics (compressed air). If we want to control pneumatical systems we need different types of valves. Directional (way) valves and pressure or flow control valves.

The role of pneumatic way valves is to connect the individual working connections and thus ensure the flow of compressed air to the different parts of the equipment at the specified time. They can be controlled in a number of ways (manually, mechanically, by air, etc.). In modern automation equipment an electrical control is the most used type of actuation. Regardless of the method of actuation, pneumatic valves are divided into directly controlled and indirectly controlled valves.

In directly controlled valves, the force generated by the control is used directly to move the internal parts of the valve. This type of valve requires a control that is capable of generating a higher force, which in the case of solenoid coils means a larger design and therefore a higher power consumption.

In the case of indirectly controlled valves, the force generated by the control does not act directly on the internal parts of the valve, but uses other energy (e.g. compressed air) to increase the actuating force. This type of design allows the use of smaller components (solenoid coils) and thus reduces energy consumption.

All three types of piezoelectric composite structures can be used to control pneumatic control valves, each with its own advantages and disadvantages.

For biomedical applications where the application requires the smallest possible valve size, a diaphragm (unimorphic) type of piezoelectric actuator is ideal. For microfluidic biomedical systems, this type is the most efficient because it has a very simple design (contains a minimum of moving parts) thus significantly reducing the risk of failure (Fig. 7) (Durasiewicz et al., 2021).

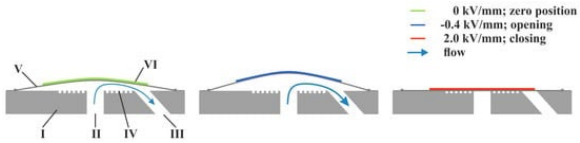


Figure 7. Membrane piezoelectric valve (Durasiewicz et al., 2021)

A different approach is to control a manifold servo valve using a layered piezoelectric micro actuator, i.e. an actuator with a polymorphic structure (Fig. 8). This method of actuation is fast and accurate, but the ratio between the piezoelectric actuator and the way valve itself is not ideal. The piezoelectric actuator does not achieve the working stroke that is directly required to override the way valve. For this reason, it uses a lever mechanism which increases its working stroke but also its design size, making the piezo actuator with the stroke increasing design in this case as large as the body of the way valve itself. Because of the mentioned ratio between actuator and valve size, this principle is not suitable for structurally larger valves, but also not for small valves or micro-valves used in biomedical applications. It can be used in applications where the size of the actuator is not required (Park et al., 2021).

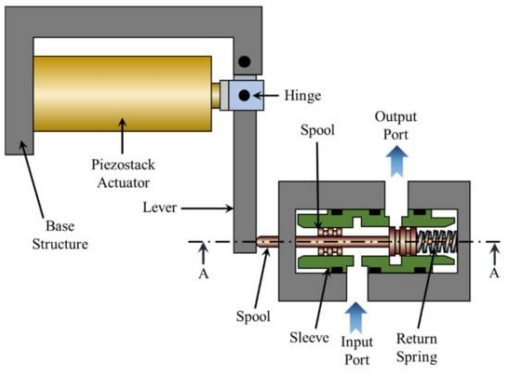


Figure 8. Way valve with piezostack actuator (Park et al., 2021)

If the application requires flow or pressure control, proportional valves (servo valves) must be used. For these types of valves, the bending-beam design (Fig. 9) is best suited. Compared to proportional solenoid valves, it has up to 95% lower power consumption (depending on the valve construction), is noiseless and does not generate heat (Easy integration of piezo valves for gas handling, 2021). The energy

consumption is ensured by the fact that after polarization of the piezo valve, it remains in the position to which it was set even after the power supply is disconnected. By not having to constantly power the valve a very large amount of energy is saved. Although up to 310V is required to operate these valves, the corresponding current is very small (max. 5 mA) (Valves with piezotechnology, 2018).

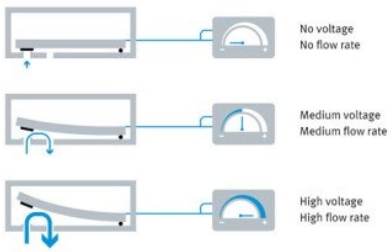


Figure 9. Proportional valve with bending beam actuator (Valves with piezotechnology, 2018)

For applications with higher flow requirements, this technology can also be applied to indirectly controlled valves.

6 Comparison of Power Consumption of Proportional Valves

Market offers a wide range of proportional valves with different constructions. The aim for the machine builders is to use a valve with low power consumption so they can reduce the operating costs for their machines thus reducing operating costs and ensuring sustainable production. Power consumption of a proportional valve depends on its construction of the control circuit, so called pilot stage. The main working stage is in this case not important, because it only multiplies the controlled pressure value to higher flows and is almost the same for any type of the proportional valve. So measuring was pointed on the control circuit (pilot stage) of the valves.

For the test itself, proportional pressure regulating valves from company Festo were selected. They have a different construction of main working stage, but as mentioned before this issue has no effect on the overall electrical consumption of the valve.

One of the valves was a proportional valve type VPPE-3-1/8-6-010 with common on/off technology that is widely used for standard proportional valves (Fig. 10). This type of valve is an indirectly operated valve (piloted valve) which uses two 2/2-way valves to pressurise or depressurise the main membrane of the valve and thus controls the output pressure of the main working part.



Figure 10. Common proportional pressure regulating valve (Proportional pressure-regulator VPPE, 2025)

Second type was a proportional valve with piezoelectrical actuator VEAB-L-26-D9-Q4-V1-1R1 that directly controls the output of the valve (Fig. 11). This construction refers to a directly operated valves, but could be also used as a pilot stage instead of 2/2-way valves in common designs for larger valves with higher flow rates.



Figure 11. Piezoelectrical proportional pressure regulating valve (Proportional pressure-regulator VEAB, 2025)

Both valves have the same parameters regarding the control (Table 1). The difference is in construction and its nominal flow, but this fact has no influence, because tests were focused on the electrical energy consumption of the control circuit of the valves.

Table 1. Technical parameters

	VPPE type	VEAB type
Type of pilot	indirect	direct
Inlet pressure	6 – 8 bar	0 – 6,5 bar
Outlet pressure	0,15 – 6 bar	0,03 – 6 bar
Setpoint value	0 – 10 V	0 – 10 V
Accuracy AO	2,25% FS	2% FS
Nominal voltage	24 V DC	24 V DC
Nominal flow of main working stage	850 l/min	20 l/min

7 Measuring Setup

Measuring setup consisted of a PLC, two multimeters, proportional valve and a pressure sensor (Fig. 12).

PLC was used to control the analogue value. It was programmed to change the analogue setpoint for the proportional valve in time. Sequence was from 0V. After every 10 seconds the setpoint value was raised by 25% of full scale of the output (1,5 bar, 2,5 V) until it reached 100%. After that the setpoint value was set again to 0V. Time delay of 10 seconds was used for the valve to stabilise the output pressure to desired value.

Multimeters were used as measuring devices to check the analogue value from the PLC and also to measure the current that was used by the proportional valves during operation. Multimeter for current measurement was connected to PC to collect the measured data.

Pressure sensor was a visual feedback for the measurement to check the value of pressure at the outlet of the proportional valve.



Figure 12. Measuring setup

8 Results

The aim of the measurement was to collect information about the overall power consumption of the proportional valves during their operation.

Basic power consumption P refers to electrical energy E per unit of time t (eq. 1).

$$P = \frac{E}{t} \tag{1}$$

Electrical energy E is represented by current I floating thru a component connected to voltage V during time t (eq. 2).

$$E = U * I * t \tag{2}$$

As both proportional valves work with the same nominal voltage and the same testing procedure is used for all measurements, it is not necessary to calculate the power consumption. The ratio of power consumption of both valves can be represented by the current they used during operation.

Results of the measurement (Fig. 13) show the values of current consumption for each of the proportional valves. These values were calculated from 50 measurements as an average value.

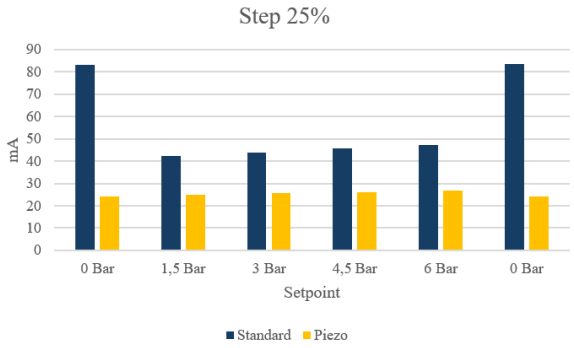


Figure 13. Results of measurements

Displayed values are a total consumption of the valves including electronics necessary to drive the valves and also the actuating stage (2/2-way valves or piezo valves).

9 Conclusion

Measured values of current consumption clearly show the potential in energy saving (Table 2). Energy saving potential is calculated as a percentage of overall energy consumption of a common solution with solenoid valves that can be saved using proportional technology instead.

Table 2. Energy saving potential

Setpoint (bar)	Analogue input (V)	Current cons. VPPE (mA)	Current cons. VEAB (mA)	Energy saving potential (%)
0	0	83,19	24,20	70,91
1,5	2,5	42,44	25,01	41,08
3	5	43,99	25,60	41,81
4,5	7,5	45,58	26,17	42,58
6	10	47,41	26,73	43,62
0	0	83,68	24,22	71,06

Energy saving potential changes in different parts of the working cycle. Maximum savings can be seen in standby state without regulation. In this state the savings go up to 70% which is given by the working principles of both valves. Standard on/off proportional valve uses energy to power the electronics and the coils

to maintain the pressure of 0 bar. On the other hand, piezo proportional valve uses the energy only to power the electronics.

By setting the pressure to a certain value the consumption of standard on/off valve drops, because it deenergises one of the coils. Consumption of piezo valve rises by a small value but is still lower than the consumption of standard on/of valve creating an energy saving potential higher than 40%.

Use of proportional technology in pneumatic control valves can clearly reduce the power consumption of these components creating a huge space to save the energy. In our test, we saved between 40 and 70% of energy depending on the duty cycle. When used in larger quantities on automation equipment, this technology can significantly reduce the energy consumption of the output stages and thus also reduce the cost of production. Valves using piezo technology also have a wider variability in their application, enabling more flexible production processes and allowing us to move closer to the essence of Industry 4.0 or even wider. The cost of valves with piezo technology is higher than standard valves with coils, but their lower consumption and variability allows more sustainable and economical plant operation.

Acknowledgments

This publication was written with the financial support of the VEGA agency in the frame of the project 1/0176/22 “Proactive control of hybrid production systems using simulation-based digital twin”.

References

Bobic, J., Ilic, N., Despotovic, Ž., Džunuzovic, A., Grigalaitis, R., Stijepovic, I., Stojanovic, B., Petrovic, M.V. (2023). Properties and Potential Application of Lead-Free ($BaZr_{0.2}Ti_{0.8}O_3$) and Lead-Based ($PbZr_{0.52}Ti_{0.48}O_3$) Flexible Thick Films. Retrieved from https://doi.org/10.3390/cryst13081178

Budkova, A. (2022). Testing of piezoelectric modules for powering a biomedical applications. *Czech Technical University in Prague*.

Durasiewics, C.P., Güntner, S.T., Maier, P.K., Hölzl, W., Schrag, G. (2021). Piezoelectric Normally Open Microvalve with Multiple Valve Seat Trenches for Medical Applications. Retrieved from https://doi.org/10.3390/app11199252

Easy integration of piezo valves for gas handling (2021). Retrieved from https://www.festo.com/de/en/e/about-festo/blog/innovation/easy-integration-of-piezo-valves-for-efficient-gas-handling-id_1621987/

Hradil, A. (2011). Design of microactuators using SMART materials. *Brno University of Technology*.

Overview of minerals and rocks (2024). E-learning of Faculty of Science – Masaryk University Retrieved from https://is.muni.cz/elportal/estud/pedf/js07/mineraly/materialy/mineraly/oxidy_kremen.html

Park, Z.J., Kim, B.G., Jeon, J.C., Jung, D., Choi, S.B. (2021). The Effect of Spool Displacement Control to the Flow Rate in the Piezoelectric Stack-Based Valve System Subjected to High Operating Temperature. *Actuators*. Retrieved from https://doi.org/10.3390/act10090239

Piezoelectric Charge Constants (2024). Retrieved from https://www.americanpiezo.com/knowledge-center/piezo-theory/piezoelectric-constants.html

Pinto, R., Rashmi, K.R., Jayarama, Y., Navine Bappalige, N. (2017). A Review on Vibration Based Piezoelectric Energy Harvesters. Retrieved from https://www.researchgate.net/publication/339052286

Plass, P. (2009). PiezoLegs motor in biomedical applications. *Brno University of Technology*.

Proportional pressure-regulator VEAB (2025). Retrieved from https://www.festo.com/de/en/a/8046300/?q=veab%7E%3AfestoSortOrderScored

Proportional pressure-regulator VPPE (2025). Retrieved from https://www.festo.com/de/en/a/539639/?q=vppe%7E%3AsortByCoreRangeAndNewProduct

Tamburrano, P., Plummer, A.R., De Palma, P., Distaso, E., Amirante, R. (2020). Novel Servovalve Pilot Stage Actuated by a Piezo-electric Ring Bender: A Numerical and Experimental Analysis. Retrieved from https://doi.org/10.3390/en13030671

Tinghai Cheng, Jianping Li (2022). Introductory chapter: Piezoelectric actuators. doi:10.5772/intechopen.104232

Valves with piezotechnology (2018). Retrieved from https://www.festo.com/de/en/e/about-festo/blog/innovation/valves-with-piezo-technology-id_1721683/

What is piezoelectricity? (2024). Retrieved from https://onscale.com/piezoelectricity/what-is-piezoelectricity/

Comparative Analysis of Bandpass IIR Filters on Microphone PDM Signal for FPGA Implementation

Lukrecia Vulić, Ivan Aleksi, Tomislav Matić
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Kneza Trpimira 2B, 31000 Osijek, Croatia
{lukrecija.vulic, ivan.aleksi, tomislav.matic}@ferit.hr

Abstract. *In this paper, a comparative analysis of quantisation effects on different types of Infinite Impulse Response (IIR) filters is performed. This research provides a basis for future applications in Field-Programmable Gate Array (FPGA) implementations. Data processing and filter evaluation are carried out using MATLAB. Butterworth, Chebyshev Type I, Chebyshev Type II and elliptical filters are selected as IIR filters. The experiments assess how different types of quantisation approximate the double-precision floating-point signal used as a reference. Several quantisation formats are considered, including Q7.24, Q2.16, and Q4.12. The metric used to compare quantisation levels and filter types is Mean Square Error (MSE) in combination with the Hamming window around frequencies of interest. The conclusions are based on the performance of the quantised version in comparison to the double-precision version and how well this format is suited to the intended hardware. The study finds that, for this scenario, the optimal ratio of resource consumption to minimum quantisation error occurs with a Butterworth filter at a quantisation level of Q2.16.*

Keywords. sound signal processing, MEMS microphone, PDM, IIR, quantisation, FPGA

1 Introduction

Digital filters are fundamental components in modern signal processing that enable the manipulation and analysis of discrete-time signals. They are used in various applications, including telecommunications, audio processing and biomedical engineering Oppenheim, 1999. In this paper, the processing of audio signals is performed using a Micro-Electro-Mechanical Systems (MEMS) microphone that outputs a pulse density modulation (PDM) digital signal, with implementation targeted on a Field-Programmable Gate Array (FPGA) platform. Due to the nature of the PDM format, suitable filtering is required to reconstruct the original signal. As the FPGA does not provide native floating-point support, the use of fixed-point arithmetic is required, which necessitates the quantisation of the filter parameters.

Infinite Impulse Response (IIR) filters are particularly appreciated because they achieve the desired frequency responses with fewer coefficients compared to their Finite Impulse Response (FIR) counterparts Hamming, 1989. Among the different IIR filter structures, Direct Form I is characterised by its straightforward implementation, where the filter's difference equation into a system diagram Oppenheim, 1999. This structure offers advantages in terms of simplicity and reduces the risk of numerical overflow in fixed-point arithmetic Oppenheim, 1999. When implementing Direct Form I filters on FPGAs to process PDM microphone signals, several aspects need to be considered, including resource utilisation, numerical stability and real-time performance.

This study examines the effect of quantisation on the following Direct Form I filter types: Butterworth, Chebyshev Type I, Chebyshev Type II and elliptical. The quantisation levels chosen reflect both common standards and forms that are suitable for the hardware to which they can be applied. The suitability of the individual quantisation forms is determined by comparing the double precision floating-point and fixed-point response to the step function and the response to the sample recording of an audio signal.

The rest of the paper is organised as follows. Section 2 gives an overview of related works, while section 3 describes the theoretical background of signal type, filter design and implementation form. Section 4 gives an overview of the methodology and Section 5 presents the experimental evaluation with the results and discussion. Conclusions and future remarks can be found in Section 6.

2 Related research

The comparative analysis of filters for specific applications is not new. In Ramdoss and Pl, 2013, for example, the authors conducted a comparative study to determine the best filtering method for audio signal processing, more specifically speech signal processing. In this work, the authors evaluated the effectiveness of each filter by observing how well chosen filters perform con-

tinuous speech segmentation. Filtering this type of signal allows for the removal of unwanted background noises and other interferences. The experiments were conducted using the OGI-MLTS corpus Muthusamy et al., 1992 and a sample voice signal to evaluate the performance of the different filters.

In Carletta et al., 2003 a framework for the implementation of IIR filters in fixed-point arithmetic on an FPGA was developed. The aim of the paper was to determine the bit size of the filter coefficients while maintaining accurate performance. The methodology is used in a magnetic bearing control system where a compensator is generated according to the selected bit width. In that way, the authors ensured that there is no overflow in registers. It was also shown that the approach creates a balance between computational accuracy and efficient utilisation of hardware resources.

The authors in Zelmat et al., 2023 address both IIR filter design and precision by applying multiobjective optimisation for both aspects. Similarly, Hormigo and Caffarena, 2021 proposes an emulator to accelerate word length optimisation in FPGAs. The method requires the user to reconfigure the FPGA only once, incorporating precision limiters into the design itself. In that way, it is possible to test different word length combinations on the circuit without the need for reconfiguration.

The implementation of IIR filters in fixed-point arithmetic is considered in Pinheiro et al., 2010. In the research, filter types such as: Butterworth, Chebyshev Type I and II and elliptical are evaluated based on the quality factor parameters and the influence of cut-off frequencies on the number of significant bits needed to represent the coefficients. It becomes clear that Direct Form I structures are preferable for FPGA implementation as they have lower latency and handle parallel operations better. The results also show that most filters achieve acceptable performance with only 10–12 coefficient bits, making them suitable for FPGAs with limited resources.

These studies collectively demonstrate that the choice of filter type, coefficient precision, and implementation structure significantly affects both performance and hardware efficiency in fixed-point FPGA designs. Building on this foundation, the present work examines the viability and challenges of implementing IIR filters in fixed-point arithmetic for FPGA applications, with an emphasis on minimising quantisation error. This approach enables the identification of filter–structure combinations that achieve the best trade-off between numerical accuracy and hardware efficiency.

3 Theoretical background

3.1 Pulse density modulation

Literature describes PDM as a 1-bit digital audio signal. A digital microphone captures this type of signal. This step is done via an analog preamplifier and a PDM modulator. The analog signal is initially captured by a microphone component and immediately amplified by the preamplifier. The signal is then sampled at a high frequency and quantised in the PDM modulator. The modulator generates a single-bit output signal that has also undergone a noise shaping process. Finally, the digital microphone interface receives the master clock and passes on the sampled bit stream. The device using the digital microphone provides the master clock, the rate of which is determined by the sampling rate of the system and the bit transmission rate Evans, n.d.

3.2 Filter Design

The choice of digital filter type is crucial for any signal processing task. This choice is limited by the given task and by the computing power and resources available for the task, in this case an FPGA board. There are two major divisions of digital filters: the non-recursive filters and the recursive filters. Non-recursive filters are commonly referred to as FIR filters and recursive filters are known as IIR filters Hamming, 1989. FIR filters are not recursive and therefore only use the previous and current inputs when calculating the filter output. On the other hand, IIR filters take current and past inputs as well as the past outputs when calculating the filter output. This distinction between the filters leads to several notable differences. The loop architecture of IIR filters requires fewer coefficients for sharper transitions, in contrast to FIR filters, which require a higher order of coefficients to achieve similar results National Instruments, 2023a. This indicates that IIR has superior computational performance and resource efficiency. The recursive properties of IIR can lead to potential drawbacks. If not constructed correctly, they can become unstable, whereas FIR structures maintain inherent stability Oppenheim, 1999.

Considering all these aspects, the selected filter type for this paper is the IIR filter, due to the limitations of the hardware platform planned for the implementation. The aim of this paper is to find the optimal IIR version for implementation on the FPGA platform. This paper focuses on the following IIR filter coefficient-approximation methods:

1. Butterworth approximation

   Butterworth filters are particular types of digital filters that are characterised by a flat, magnitude response in the passband. This characteristic requires a compromise characterised by a wide transition bandwidth, which is poor compared to other classical approximations. To achieve a narrow

transition bandwidth, a higher order of the Butterworth filter is required Lai, 2003.

2. Chebyshev approximation

   Chebyshev filters are calculated using Chebyshev polynomials, which are defined by the filter order. Chebyshev polynomials are used to shape the filter frequency response to minimise the maximum deviation of the passband. The error curve generated by this approximation has equal waves. Chebyshev filters are divided into two categories: Type 1 and Type 2 Lai, 2003. The Chebyshev Type 1 filter permits a ripple in the passband, whereas the Type 2 filter permits a ripple in the stopband Hamming, 1989.

3. Elliptic function

   Eliptical filters are characterised by the fact that they have ripples in both the passband and stopband. Their design is based on the theory of eliptical functions Hamming, 1989.

3.3 Direct Form I

Direct Form I is a digital filter implementation technique for IIR filter realization. Other common forms include Direct Form II, cascade form, parallel form, and lattice-based structures, each offering different trade-offs in memory usage, numerical properties, and suitability for fixed-point or floating-point implementation. The Direct Form I structure can be described as two FIR filters, one forward and one reversed. The forward FIR filter comprises numerator or *b* coefficients, and the reversed one uses denominator or *a* coefficients Inc., 2008. Its most valuable attribute is found in fixed-point arithmetic, where there is no possibility of filtering internal register overflow. This advantage is due to the single summation point in the filter structure, as well as the characteristic fixed-point wraparound behavior. One notable downside, compared to other types of implementation, is the double amount of delay elements required Smith, 2007. Another important consideration is the possibility of filter stability being affected by roundoffs needed for fixed-point arithmetic. This issue is solved by splitting the filter's transfer function into second-order sections (SOS) or biquads whose structure is shown in Fig.1 Smith, 2007. The biquads can be arranged in a cascaded and parallel structure. For the purposes of this paper, a cascaded structure is observed. These structures are used for larger filter orders and represent a product of the number of biquads Inc., 2008. The SOS bit-width parameter between each SOS has to be adjusted to handle the influence of input and output gains. These widths have to be adjusted in order not to lose data and retain a level of precision Mauerer, 2018.

3.4 Qm.n quantisation format

In digital signal processing, numerical data is usually represented in either floating-point or fixed-point formats. This choice depends on the constraints of the application, such as accuracy, dynamic range and hardware resources. Floating-point representation is defined by the IEEE 754 standard and represents numbers with one sign bit, one exponent and mantissa "IEEE Standard for Floating-Point Arithmetic", 2019. This enables a large dynamic range and high accuracy, making it suitable for general calculations Oppenheim, 1999. In resource-constrained environments, however, a fixed-point representation is used due to the lower hardware complexity and deterministic performance Meyer-Baese, 2014. The $Qm.n$ format is a fixed-point convention that divides a binary word into m integer bits containing a sign and n decimal bits. This enables efficient encoding of real numbers without the need for floating-point units. Converting a floating-point number to $Qm.n$ format involves scaling the value by 2^n, then rounding or truncating to the nearest integer and implementing overflow or saturation management to maintain numerical integrity Texas Instruments, 2001. This conversion reduces numerical accuracy and dynamic range while significantly increasing execution speed and resource efficiency, making Q-format arithmetic particularly advantageous in applications such as audio processing and real-time control Widrow and Kollár, 2008.

4 Methodology

The methodology in this paper focuses on evaluating the impact of the quantisation process on filter performances, focusing on digital audio signals. A key challenge in this process is that the target FPGA lacks native support for floating-point arithmetic, which makes it incompatible with the floating-point filter coefficients generated by MATLAB. Furthermore, the optimal filter type and corresponding quantisation format is not immediately obvious. To overcome these limitations,

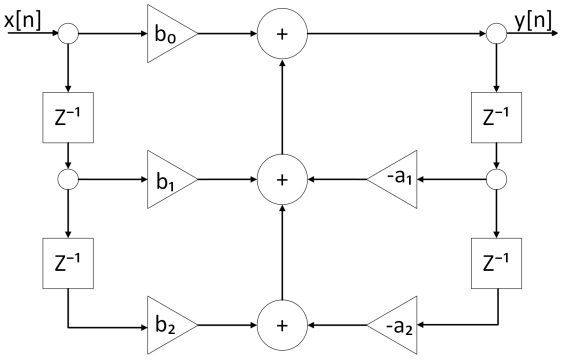


Figure 1. Block diagram of a Direct Form I IIR filter SOS National Instruments, 2023b

the filters are first designed in MATLAB using the Direct Form I structure. These filters are then converted into a fixed-point representation using a Python-based tool chosen for its flexibility and open-source availability. A simulation framework Mauerer, 2018 in Python is then used to evaluate the impact of the quantisation process on filter accuracy compared to the one in double precision. This evaluation is performed using a step function and a PDM audio recording captured by an FPGA device.

4.1 Sound production

The test samples are designed to mimic the real conditions that the planned system would encounter. The sound samples are generated artificially, using pre-recorded ambient noise and a sine wave. Human, non-speech sounds and outdoor urban noises were selected as background sounds. A bursting sine wave is a sine wave that occurs for a certain number of periods and then remains inactive. The frequency of the sine wave is set to 11.025 kHz due to the limited range of available MEMS microphones Analog Devices, n.d. and the value of the sound sampling frequency in MATLAB. When generating or analysing a sinusoidal signal in digital systems, it necessary the sine wave frequency is an integer divisor of the sampling frequency Chassaing and Reay, 2011. This exact periodicity ensures that the sampled signal repeats perfectly and without phase drift, resulting in a clean frequency domain analysis without spectral leakage in the FFT. For example, using a sampling frequency of 44.1 kHz and a sine wave frequency of 11.025 kHz results in exactly four samples per sine wave period. The frequency range of this artificial sound can be seen in Fig.2.

The dataset used for the environmental sounds is named ESC-50 and can be found in Piczak, 2015. These test samples may be used as individual sounds and as combinations. In contrast to the dataset, these sounds were cut to a duration of 4 seconds due to hardware limitations of the FPGA. The BRAM memory of the selected FPGA device can only hold up to 4 seconds of samples.

4.2 Hardware setup

The test setup consists of a loudspeaker and a MEMS microphone connected to an FPGA. The loudspeaker is used to reproduce ambient sounds and the bursting sine wave signal. The FPGA board used is the Nexys A7 from Digilent Digilent Inc., n.d. and the onboard MEMS microphone is used Analog Devices, n.d.

The FPGA is programmed to capture data from the MEMS microphone at a rate of 1 MHz and store a 4-second PDM sample in the block RAM memory. This sample is later transferred to a PC via a UART connection. This process of recording PDM data on an FPGA is designed as a Finite State Machine (FSM), see Fig.3.

The FSM is designed to record a number of N samples, each of which is stored in memory as soon as it is recorded. After N samples have been recorded and stored in memory, they are read from memory one by one and sent via a UART connection as long as there are samples to be read from memory and a UART connection is available.

4.3 Evaluation procedure

The evaluation procedure is separated into a number of steps. Firstly, the focus is on choosing the correct filter type as well as making sure that the filter is stable and is of minimum order. This is done through a Signal Processing Toolbox in MATLAB MathWorks, Inc., 2025. The filters are represented in Direct Form I structure. The generated parameters are exported to an open-source tool that is specialised for filter coefficient quantisation and testing filters with quantised co-

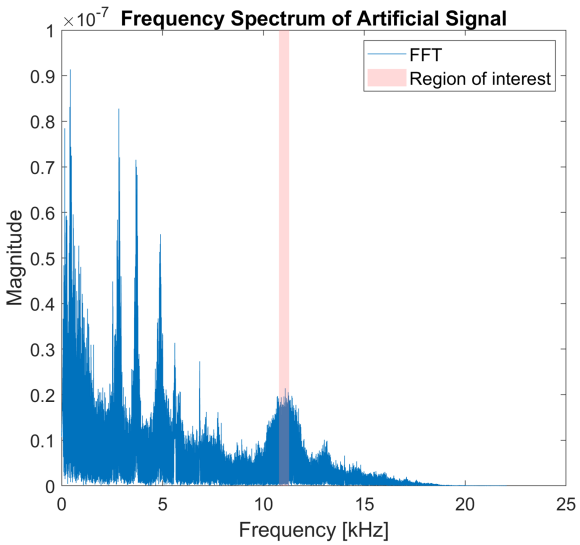


Figure 2. FFT analysis of the test sound used for the sine extraction at 11.025 kHz with a sampling frequency of 44.1 kHz

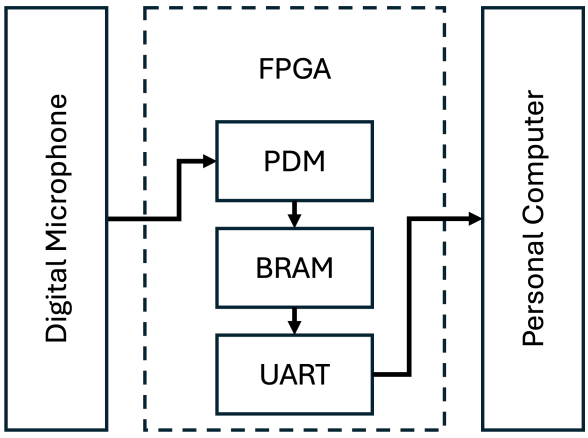


Figure 3. Digital MEMS microphone and PDM signal processing chain on an FPGA board using FSM

efficients. The stability of the quantised coefficients is determined based on the pole position in the complex plane. The next step is focused on observing the double precision floating-point and quantised filter's response to a step function. The unit step function is a signal that has a constant value of zero for all time before a specified starting point, and a constant value of one for all time after and including that point Shmaliy, 2007. Using this activation signal the filter type and quantisation settings can be vetted through observing the error rates between the quantised and floating-point representation responses. Finally, the filter is tested on the PDM signal described in subsection 4.1. The validity of the filter is vetted by observing how it filters out a PDM audio recording with a known frequency included. This recording is filtered and processed by Fast Fourier Transform (FFT).

The steps are numbered and explicitly stated in the following paragraph:

1. designing the filter using MATLAB with desired specifications
2. exporting filter coefficients to an open-source Python tool VIIRF Mauerer, 2018
3. quantise filter coefficients to the chosen fixed-point format and confirm filter stability in this format
4. setting the value of the SOS bit size
5. simulate and evaluate the filter using test signals
   (a) compute MSE of step response
   (b) compute MSE of FFT of a PDM signal.

5 Experimental evaluation

5.1 Observed parameters

Four filter types were chosen for this evaluation process: eliptical, Butterworth, Chebyshev Type I and Chebyshev Type II. These filters were chosen due to their popularity and their varying effects as in Pinheiro et al., 2010. A bandpass filter was chosen with a center frequency of 11.025 kHz and a bandwidth of 250 Hz on each side, resulting in a total passband ranging from 10.775 kHz to 11.275 kHz. The *Apass* parameter represents the maximum allowed variation in the passband Proakis and Manolakis, 2006, the chosen values for this parameter can be seen in Table 1.

It was also important to look at how each filter reacts to different quantisation parameters, as they all have varying tolerances for this process. As mentioned, this comparative analysis is used to determine which filter would best be suited for FPGA fixed-point implementation.

Calculations on the FPGA are done using Digital Signal Processing (DSP) slices, which have a certain data width that they handle. In cases where this width is exceeded, calculations have to be transferred to a new slice. It is important to keep calculations in a certain bit range to minimize the use of the available DSP slices. The SOS bit-size parameter refers to the register size in between SOS of the filter, where the output of each section in stored and used for further calculations. This parameter depends on the size of filter input, the value of the gain and the user's input on additional bits. Depending on the gain value for each section there may be need for larger register size at the section output. The user can also impact this parameter if there is need for extremely precise calculations.

Errors in the quantisation process are also observed, primarily in the Mean Square Error (MSE) format after testing the filters. The MSE is calculated based on a fixed and double precision floating-point step response, as well as the Fourier analyses of these two formats. The floating-point value is taken as a reference point to see how the conversion and round off error is manifested in fixed-point format response. A Hamming window is applied and centered on the bandpass region for the MSE FFT calculation. This approach prioritises accuracy in the passband by assigning higher weights to the error in that region while suppressing the contributions outside of it. Stability and usability of the filter after the quantisation process were also looked at in terms of pole placement in a complex plane and the value of the filter output. The parameters used in this study are presented in Table 1. The *Apass* values were selected to investigate the impact of varying passband ripple levels on quantisation effects. The effects of different *Apass* values are shown in Fig.4. The Q2.16, Q4.12, and Q7.24 formats were chosen to align with common register widths in fixed-point hardware implementations, thereby enabling a comparative evaluation of their performance. The SOS bit-width values were varied between 17 and 28 bits to examine how the precision of intermediate results influences the accuracy of the final output.

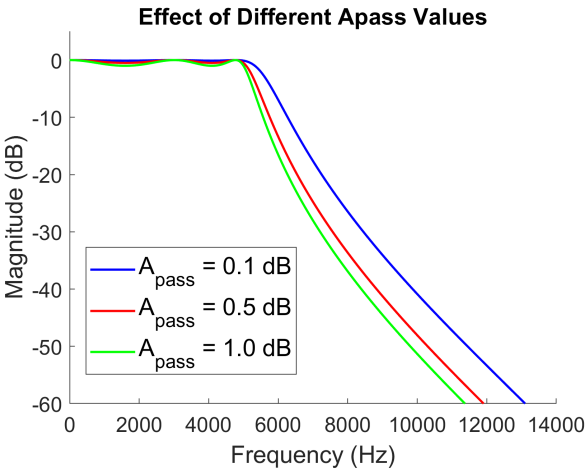


Figure 4. Example of the effects of different Apass values

Table 1. EXPERIMENT PARAMETERS

Apass [dB]	Q format	SOS bit-size
0.10	Q2.16	17–28
0.25	Q4.12	17–28
0.50	Q7.24	17–28

5.2 Results and discussion

The results are obtained iteratively by repeating the procedure described in subsection 4.3 for each parameter combination. Fig. 5(a) shows the effects of changing the parameters of the Butterworth filter. It is immediately apparent that the calculated error for the FFT MSE of the Q4.12 format is higher compared to others for all Apass values, and they are represented by the color green. All step response errors approach zero which indicates a good fixed-point approximation. The FFT MSE for Q2.16 and Q7.24 also approaches zero as the SOS bit size increases for all Apass values.

The graph for Chebyshev Type I filter, as seen in Fig. 5(b), has different scaling of the x axis. This is due to the need for accommodating larger values for cases when the output gain is lesser than one Mauerer, 2018. From the graph it is evident that values on the step response MSE are approaching zero. The FFT MSE values are reducing as the SOS bit size goes up, in all cases but one. For the Q7.24 format this value approaches zero, while Q2.16 preforms with somewhat lesser precision, its stagnant values being as high as 0.1. The worst performing format is once again Q4.12 as its error rate ranges from 0.3 to 0.6.

Similar results are noticed in Fig. 5(c) for Chebyshev Type II filter design. Step responses for all Q formats approach zero. On the other hand, the MSE for the FFT depicts a significant error. Despite the increase of SOS bit size, the MSE value stagnates in range of 0.1 to 0.2 for Q2.16 and Q7.24 formats. The Q4.12 format gives subpar values compared to other formats.

The elliptical filter design, as seen in Fig. 5(d), produces such filters that SOS bit size has no impact on the MSE of any Q formats. The values are constant and it is yet again apparent that calculations using Q4.12 format produce higher error rates. The error in the step response once again approaches zero, and the FFT error rate for the other Q formats for all Apass values is less than 0.1.

For the final part of this analysis, the viability of each filter type and Q format is observed. As previously mentioned, it is important to take into consideration both the error of calculation and the constrictions of the FPGA. As mentioned already, the targeted FPGA DSP slices allow calculations in the format of 25x18 bits and therefore to achieve minimal resource usage it is imperative to limit the SOS bit size parameter to be up to 25 bits in length, and coefficient size to be up to 18 bits. The SOS bit size across all filter types is seen in Table 1. The Q formats overall bit size is set to be either 16, 18, or 31 in size.

Table 2 shows the best results obtained thru this experiment. For each of the four filter types the best result of each Q format with the corresponding SOS bit size is extracted, the best result being a minimal error of the FFT comparison between double precision and fixed-point. MSE of step response is also included in Table 2 and shows that the minimal error for most cases is achieved with Q7.24 format. In this table the best results are bolded and it is clear that for three types of filters the smallest error for FFT is produced using Q7.24 format which is to be expected. With an extended fraction bit size precision in calculation is enhanced and a better approximation of the floating-point format is achieved. Upon further review, Q2.16 format appears to be a clear second choice. The clear outlier remains the Q4.12 format. The SOS bit size in t Table 2 is within the bounds of 25 bits and therefore does not impact the choice of filter and format.

Taking into consideration the FPGA constrains mentioned and the MSE of the FFT analysis and step response chosen as a metric, it is therefore concluded that the best approach to implementation is to use the Butterworth filter with the Q2.16 format. This Q format would ensure that a minimal number of DSP slices is utilised, as the SOS bit width is also in the correct bound. The MSE metric is also the smallest of any in this format.

Table 2. MSE EXPERIMENTAL RESULTS

Filter	Apass	Q format	SOS bit size	MSE step	MSE FFT
Butterworth	0.10	Q2.16	22	3.45e-08	3.57e-03
	0.25	Q4.12	22	5.46e-06	2.14e-01
	0.50	Q7.24	22	**4.37e-10**	**7.58e-07**
Chebyshev I	0.10	Q2.16	23	1.54e-07	1.69e-02
	0.10	Q4.12	22	2.17e-06	3.40e-01
	0.50	Q7.24	23	**3.51e-10**	**5.40e-07**
Chebyshev II	0.50	Q2.16	19	**2.12e-06**	1.20e-01
	0.50	Q4.12	18	5.14e-06	1.96e-01
	0.50	Q7.24	18	2.56e-06	**1.19e-01**
Elliptic	0.25	Q2.16	17	**4.93e-07**	3.18e-02
	0.50	Q4.12	17	1.63e-06	2.66e-01
	0.50	Q7.24	17	2.10e-06	**1.86e-02**

6 Conclusion

The paper demonstrates the viability and challenges of choosing an implementation format of IIR filters – Butterworth, Chebyshev Type I and II, and elliptical – in fixed-point arithmetic for FPGA applications. The design of the filters in floating-point and its validity was verified using MATLAB, while the effects of quantisation on filter stability and errors in approximation caused by this process were observed in order to determine the best-suited combination for FPGA implementation. The specific FPGA implementation sup-

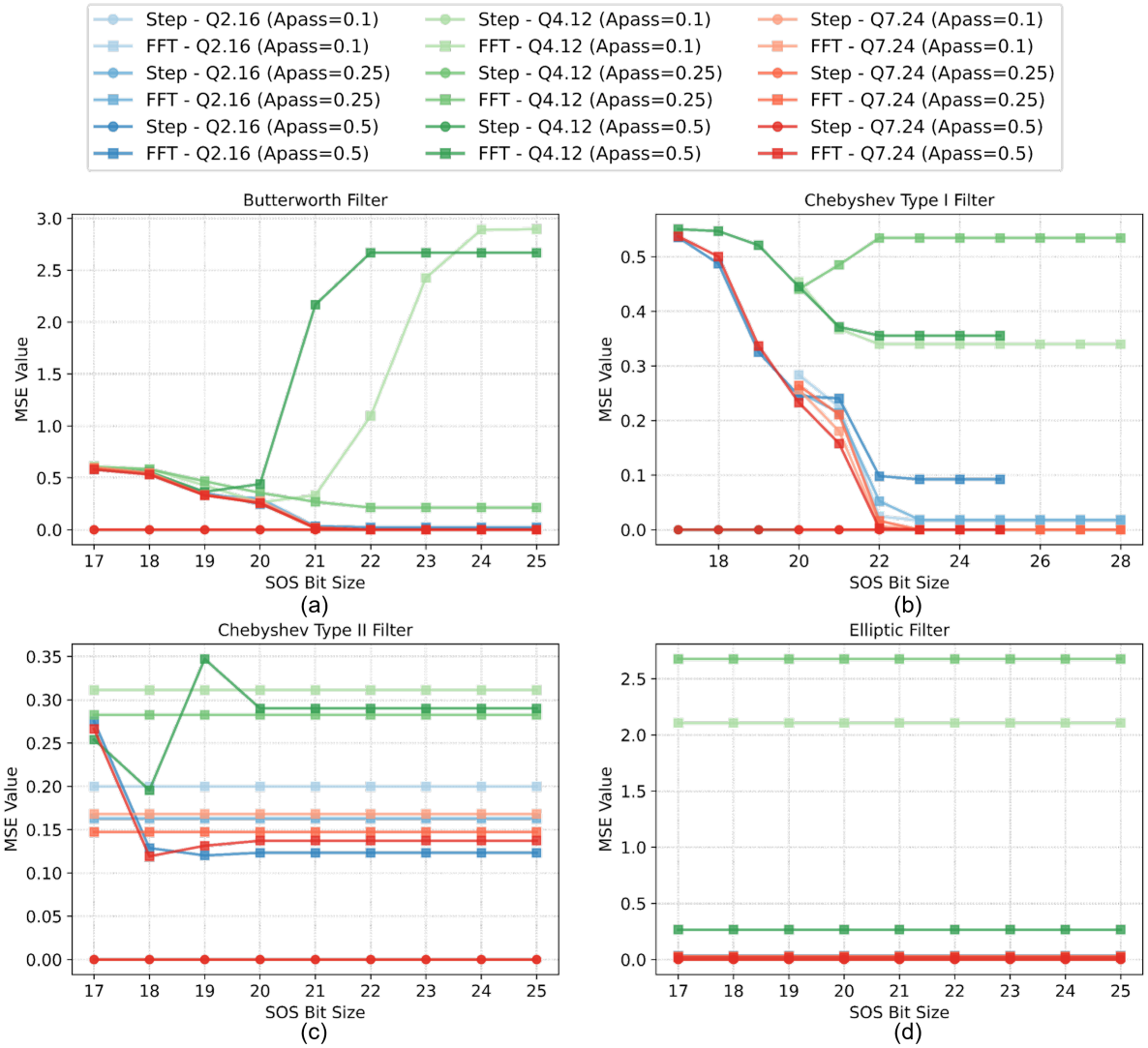


Figure 5. Comparison of different filter types and quantisation formats (a) Butterworth filter (b) Chebyshev Type I filter (c) Chebyshev Type II filter (d) Eliptical filter

posed is Direct Form I, which is a straightforward IIR implementation using SOS that additionally reduces the effects of quantisation. The coefficient stability is checked after the quantisation takes place, after which both quantised and floating-point coefficient are used to filter a step signal and a audio recording. In this way quantisation responses are measured up to the reference captured by the floating-point responses.

All the filter types and corresponding quantisation levels have produced stable filters. Regarding the comparison, across most filter types the minimal quantisation error was produced using the Q7.24 format. However, taking into consideration the resource management needed for low-level implementation, the authors conclude that the best ratio of error management and resource consumption is achieved with the Q2.16 format. The minimum error for this format is achieved with the Butterworth filter type.

This research opens up several directions for future investigation. The obvious one is the implementation and real-time testing of the designed filters on the FPGA platform, using VHDL to verify synthesis, timing and resource utilisation. Future work could also incorporate dynamic range analysis and noise sensitivity, especially for audio and sensor applications. Finally, extending the analysis to adaptive filtering algorithms or multi-channel systems, such as beamforming arrays is also a possibility.

Acknowledgement

The authors would like to thank Infineon Technologies Austria AG for their support in this research in the framework of the Important Project of Common European Interest (IPCEI) PhD program on Microelectronics.

References

Oppenheim, A. (1999). *Discrete-time signal processing*. Pearson Education. https://books.google.hr/books?id=geTn5W47KEsC

Hamming, R. (1989). *Digital filters (3rd ed.)* Prentice Hall International (UK) Ltd.

Ramdoss, A., & Pl, C. (2013). A comparative study on various types of filters in audio signal processing.

Muthusamy, Y. K., Cole, R. A., & Oshika, B. T. (1992). The ogi multi-language telephone speech corpus. *Proceedings of the 2nd International Conference on Spoken Language Processing (ICSLP 1992)*, 895–898. https://doi.org/10.21437/ICSLP.1992-276

Carletta, J., Veillette, R., Krach, F., & Fang, Z. (2003). Determining appropriate precisions for signals in fixed-point iir filters. *Proceedings of the 40th Design Automation Conference (DAC)*, 656–661. https://www.cecs.uci.edu/~papers/compendium94-03/papers/2003/dac03/pdffiles/38_3.pdf

Zelmat, M., Lamini, E.-S., Tagzout, S., Belbachir, H., & Belouchrani, A. (2023). Multi-objective approach for iir filter design and bit-width optimization. *Circuits, Systems, and Signal Processing*, *42*(8), 4836–4867. https://doi.org/10.1007/s00034-023-02334-1

Hormigo, J., & Caffarena, G. (2021). Fpga acceleration of bit-true simulations for word-length optimization. *2021 IEEE 28th Symposium on Computer Arithmetic (ARITH)*, 119–122. https://doi.org/10.1109/ARITH51176.2021.00033

Pinheiro, E., Postolache, O., & Silva Girão, P. (2010). Fixed-point implementation of infinite impulse response notch filters. *Metrology and Measurement Systems*, *17*. https://doi.org/10.2478/v10178-010-0019-3

Evans, B. L. (n.d.). Understanding pdm digital audio [Course materials, The University of Texas at Austin, Accessed: May 16, 2025].

National Instruments. (2023a). Fir and iir filters [Updated 2023-02-21; accessed 2025-09-05].

Lai, E. (2003). 7 - infinite impulse response (iir) filter design. In E. Lai (Ed.), *Practical digital signal processing* (pp. 145–170). Newnes. https://doi.org/https://doi.org/10.1016/B978-075065798-3/50007-2

Inc., X. (2008, March). Infinite impulse response filter structures in xilinx fpgas. https://www.xilinx.com/support/documentation/white_papers/wp330.pdf

Smith, J. O. (2007). *Introduction to digital filters with audio applications* [online book].

Mauerer, M. (2018). Viirf: Versatile iir filter implementation [Accessed: 2025-04-17]. https://github.com/MauererM/VIIRF

National Instruments. (2023b). IIR SOS Specifications - LabVIEW Digital Filter Design Toolkit API Reference [Accessed: 2025-05-26]. https://www.ni.com/docs/en-US/bundle/labview-digital-filter-design-toolkit-api-ref/page/lvdfdtconcepts/iir_sos_specs.html

Ieee standard for floating-point arithmetic. (2019). *IEEE Std 754-2019 (Revision of IEEE 754-2008)*, 1–84. https://doi.org/10.1109/IEEESTD.2019.8766229

Meyer-Baese, U. (2014). *Digital signal processing with field programmable gate arrays* (4th ed.). Springer. https://doi.org/10.1007/978-3-319-00584-1

Texas Instruments. (2001). *Tms320c55x dsp programmer's guide* [Literature Number: SPRU376B]. Texas Instruments. Dallas, TX. https://www.ti.com/lit/ug/spru376a/spru376a.pdf?ts=1747302209728&ref_url=https%253A%252F%252Fwww.google.com%252F

Widrow, B., & Kollár, I. (2008). *Quantization noise: Roundoff error in digital computation, signal processing, control, and communications*. Cambridge University Press.

Analog Devices. (n.d.). ADMP421: High-performance, low power, digital output MEMS microphone [Datasheet, Accessed: May 16, 2025]. https://www.analog.com/media/en/technical-documentation/obsolete-data-sheets/ADMP421.pdf

Chassaing, R., & Reay, D. (2011). *Digital signal processing and applications with the tms320c6713 and tms320c6416 dsk*. Wiley. https://books.google.hr/books?id=LO_whVmhXE8C

Piczak, K. J. (2015). ESC: Dataset for Environmental Sound Classification. *Proceedings of the 23rd Annual ACM Conference on Multimedia*, 1015–1018. https://doi.org/10.1145/2733373.2806390

Digilent Inc. (n.d.). *Nexys a7 reference manual* [Accessed: May 16, 2025]. https://digilent.com/reference/programmable-logic/nexys-a7/reference-manual

MathWorks, Inc. (2025). Digital and analog filters [Accessed: 2025-09-05]. https://www.mathworks.com/help/signal/digital-and-analog-filters.html

Shmaliy, Y. (2007, September). *Continuous-time systems*. https://doi.org/10.1007/978-3-642-14325-0_1

Proakis, J. G., & Manolakis, D. G. (2006). *Digital signal processing: Principles, algorithms, and applications* (4th). Pearson Prentice Hall.

Deep-Learning and Stereo Camera Based System for Real-Time Virtual Safety Boundaries

Juraj Repka, Juraj Ďuďák
Slovak University of Technology
Faculty of Materials Science and Technology in Trnava
Jána Bottu 25, Trnava, Slovakia
{juraj.repka, juraj.dudak}@stuba.sk

Abstract. *We propose a real-time, cost-effective system for implementing a virtual safety zone that enables monitoring of a person's position relative to user-defined virtual boundaries. Our approach employs a hybrid algorithm that combines depth mapping and disparity calculation using deep learning models applied to stereo image pairs. Using a stereo camera setup, the system estimates human keypoint positions, which are then compared against the predefined zones. When a zone is violated, a corresponding trigger is generated. Although the system has shown its effectiveness, it also faces challenges in achieving high accuracy, making it less suitable for applications requiring precise localization.*
Keywords. Virtual protective zone, deep learning, stereo vision, human tracking.

1 Introduction

The rise of automation in industrial environments has significantly improved production efficiency, product quality, and workplace safety. Machines can now operate 24/7, reduce human error, and perform tasks that may be too dangerous for workers. Despite these advantages, automation presents serious challenges, particularly in ensuring safe interaction between humans and machines (Rybski et al., 2012).

Many industrial systems still rely on older technologies that lack contextual awareness and dynamic safety mechanisms. This means that even a minor mistake, such as an operator unintentionally entering a hazardous zone, can lead to serious accidents if machines are not equipped to detect such situations. Although newer industrial equipment may include built-in safety logic, a large portion of existing systems continues to rely on manual safety protocols (Mohammadi Amin et al., 2020; Rybski et al., 2012).

This paper is motivated by the need for a flexible, intelligent safety system that can automatically monitor safety zones and identify when a person violates them. The primary goal is to design and implement a virtual safety zone system that uses 3D computer vision and deep learning techniques to detect people in real time and assess their position relative to dangerous areas. Compared to traditional methods such as light curtains or infrared sensors, 3D vision allows for greater adaptability and accuracy, especially in complex environments (Mosberger et al., 2014; Zhou et al., 2022).

In this work, we created a working prototype capable of detecting individuals in a monitored 3D space and determining whether they are within a customizable danger zone. The system is also designed to provide visual feedback and alerts in real time to the operator.

We are utilizing stereo vision and deep learning models to create a cost-effective system for human position monitoring. We developed our own stereo system but also used an OAK-D-PRO commercial one. We applied multiple depth estimation techniques and also signal processing techniques such as smoothing algorithms.

2 Related Work

Recently, there has been a growing interest in improving safety within industrial environments using advanced technologies such as computer vision and sensor-based systems. Traditional solutions, such as light curtains or mechanical interlocks, often lack the flexibility and context awareness required to adapt to dynamic workspaces. As a result, research has increasingly focused on intelligent systems that can respond in real-time to environmental changes and reduce the risk of human error.

One significant approach involves the use of 3D cameras to detect human presence near hazardous machinery. Cheng Zhou and his colleagues developed a system that monitors the area surrounding robotic arms using 3D visual data (Zhou et al., 2022). These systems enable cameras to be placed at a greater distance compared to traditional light barriers, which must be positioned close to the machine. This configuration reduces false positives caused by non-human objects and improves overall reliability.

Sotiris Makris and his team introduced a more ad-

vanced concept, which proposed dynamic safety zones instead of static ones (Makris, 2021). These zones adjust in real-time on the basis of operational requirements and human-robot interaction. Such an approach allows for closer, yet safe collaboration between humans and machines, thereby improving productivity without compromising safety.

For mobile robotic platforms that navigate freely throughout the facility, the safety challenge becomes even more complex. Hyunjoong Cho et al. addressed this problem by developing a system that detects objects of interest and generates protective zones around them based on contextual factors, such as robot speed (Cho et al., 2022). The faster the robot moves, the larger the protective zone becomes, providing a dynamic buffer that minimizes the risks of collision.

Juraj Slovák et al. proposed a hybrid vision and RTLS-based safety system that adapts robot behavior according to the proximity and identity of nearby objects, including humans (Slovák et al., 2021). Their approach combines depth camera data with real-time location systems to distinguish between authorized (e.g., supply trolleys) and unauthorized (e.g., human) entries into safety zones, reducing unnecessary downtimes without compromising safety.

Additionally, Kozamernik et al. introduced a visual quality and safety monitoring system for human–robot collaboration that combines stereo and depth cameras with deep learning-based object and posture recognition (Kozamernik et al., 2023). Their system supports both safety (e.g., hand detection and posture analysis) and quality assurance (e.g., inspection of final assemblies), demonstrating that safety and performance feedback can be achieved with minimal hardware overhead.

All of these approaches share a common foundation in leveraging 3D computer vision to detect humans and define intelligent safety mechanisms. They demonstrate a shift toward autonomous systems capable of adapting to varied and evolving industrial environments. These insights have informed the development of our proposed solution, which aims to detect breaches of safety zones in a cost-effective and highly adaptable manner suitable for wide deployment.

3 System Architecture

The proposed virtual security zone system is made up of three main modules: a stereo vision module, a processing module, and a visualization module. These components operate in real time to detect human presence, estimate 3D position using stereo cameras and deep learning, and determine whether individuals are within a predefined virtual safety zone. The system includes hardware-based depth estimation, keypoint detection models, and a web-based visualization interface for interactive zone management (see Fig. 1).

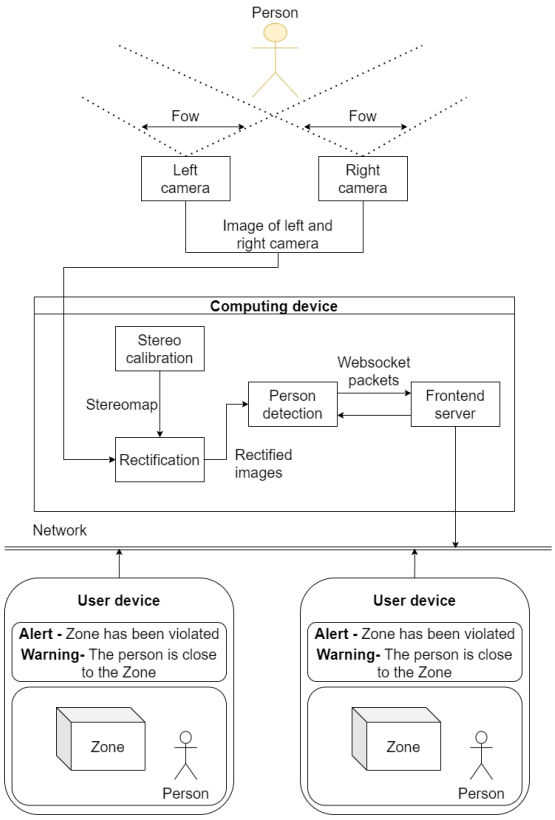


Figure 1. System architecture diagram

3.1 Stereo Vision Module

The stereo vision module is responsible for capturing the real-world 3D scene and generating a depth map. Using stereo vision technology, it emulates human depth perception by using two spatially separated cameras (see Fig. 2). These cameras are offset by a fixed baseline, a critical parameter that influences depth accuracy. A wider baseline improves depth estimation for distant objects by increasing disparity, whereas a narrower baseline is preferable for nearby objects to avoid excessive disparity and reduce matching errors.

Depth estimation is performed via triangulation, a geometric method that calculates the 3D position of a point based on its disparity in the horizontal shift between corresponding points in the left and right images. Larger disparities indicate closer objects; smaller disparities suggest objects are farther away. The accuracy of the depth calculation depends on factors such as the baseline length, image resolution, and stereo calibration quality.

For effective stereo vision, the camera pair must be calibrated and the images corrected so that the corresponding points lie on the same horizontal scan lines (epipolar lines). This rectification step is essential for robust disparity computation and is a fundamental part of our stereo vision software implementation.

We used two different stereo vision camera implementations in our research. The first is our custom-built system (see Fig. 3), also used in our previous

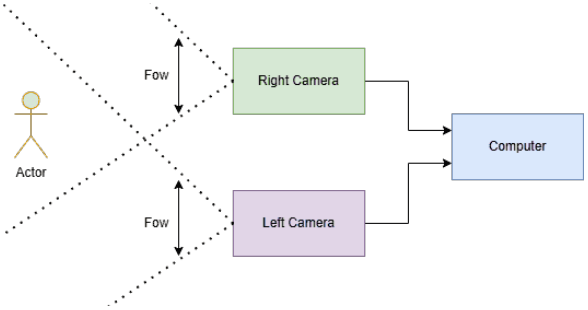


Figure 2. Abstraction of stereovision

work (Sládek et al., 2024). This setup consists of two standard webcams mounted on a rigid fixed baseline. Fixated positioning is essential to maintain consistent calibration, which is handled by our software. This calibration process estimates both the intrinsic and the extrinsic parameters of the stereo camera system, which are crucial for accurate depth estimation.



Figure 3. Homemade stereo vision camera

The second implementation uses the commercially available OAK-D Pro camera developed by Luxonis (see Fig. 4). This device integrates both stereo vision sensors and an onboard processor, enabling on-device depth estimation and neural inference. This hardware acceleration allows us to offload computationally intensive tasks to the device, significantly improving overall processing speed.



Figure 4. OAK-D-PRO camera

All algorithms were implemented and tested using both stereo vision systems. However, in this paper, we focus exclusively on the implementation using the OAK-D Pro. This decision is based on the superior performance of the OAK-D Pro, which benefits from its own hardware-accelerated stereo matching pipeline and optimized internal processing. In addition, the advanced camera specifications contributed to its improved results. For these reasons, we selected the OAK-D Pro as the primary stereo vision device for our system.

3.2 Processing Module

The processing module is the next component in our system's data flow. It receives images from the stereo camera, along with the necessary camera parameters for depth calculation. This module is responsible for detecting person keypoints, computing depth information, and forwarding the processed data to the Visualization module.

The processing module incorporates two independent processing algorithms that operate on stereo images acquired from a stereo vision module.

The two implemented algorithms are as follows: (i) a depth map-based algorithm and (ii) a disparity-based algorithm utilizing keypoint detections from stereo image pairs. Both algorithms are designed to function independently and work differently within the processing pipeline.

Depth Map-based Algorithm

This algorithm operates directly on the real-time depth map computed by the stereo device's onboard processor. It follows the depth-lookup strategy demonstrated in Luxonis' OAK-D calc-spatials-on-host reference implementation (Luxonis Inc., 2023). The primary steps are:

1. The stereo camera computes a depth map directly on the device using its onboard processor.
2. A deep learning model, specifically MediaPipe Pose (Bazarevsky et al., 2020), is applied to detect the human subject and extract keypoint coordinates from a single image frame.
3. For each detected keypoint, the corresponding depth value is retrieved from the depth map using its pixel coordinates.

The depth map used in this algorithm is calculated by the internal OAK-D-Pro processor, allowing faster performance compared to a disparity-based algorithm.

Disparity-based Algorithm

We developed a disparity-based algorithm that runs entirely on the system's main processing unit. This algorithm performs keypoint detection independently on

both the left and right stereo images. The main steps of our method are as follows:

1. Two independent instances of the MediaPipe Pose model are applied, one for the left image and one for the right image, due to the model's internal state management, which improves detection stability but prevents reusability across frames.
2. The key points are detected independently in both images.
3. For each pair of corresponding key points, the disparity is calculated as the horizontal difference of pixels between their positions in the left and right images (see Fig. 5).
4. The depth is then estimated on the basis of the computed disparity using stereo vision geometry.

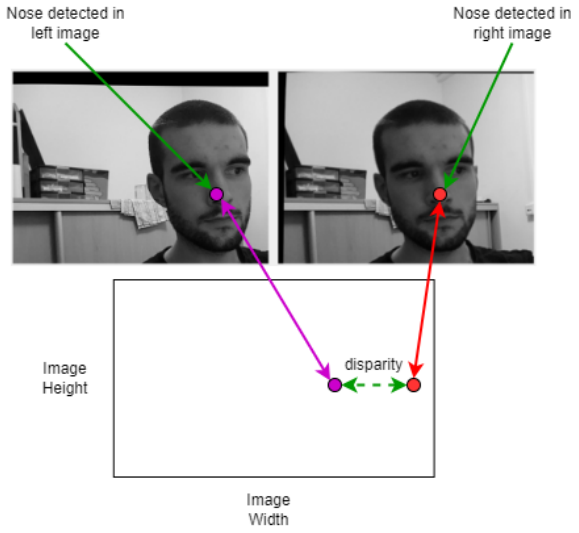


Figure 5. Disparity visualization

Although this approach requires running two models in parallel, the computational load remains manageable, as MediaPipe Pose is optimized for low-latency performance and can be executed efficiently even on edge devices.

3.3 Visualization Module

The Visualization module is responsible for displaying the detected 3D keypoints received from the processing module, as well as the boundaries of user-defined zones. It also allows users to define and modify zone parameters interactively.

Implemented as a web-based application, the module communicates with the processing module via WebSockets to receive real-time data. Display 3D keypoints using React for front-end development and Three.js for 3D rendering (see Fig. 6).

In addition to visualization, the module includes logic to determine whether detected keypoints fall

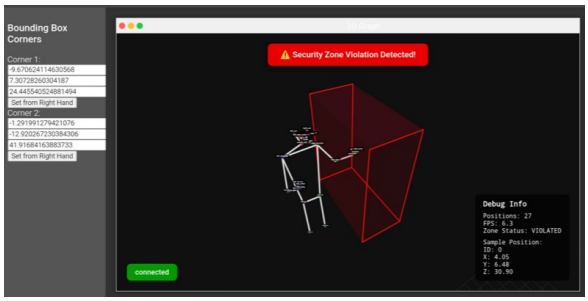


Figure 6. Frontend visualization

within predefined zones. If a person enters a defined zone, an on-screen alarm is triggered. The alarm is deactivated once the person exits the zone.

4 Experiments

Stereo vision systems often face two primary challenges that can degrade performance: poor lighting conditions and low image contrast.

To evaluate the robustness of our algorithms under these conditions, we conducted four experiments using two distinct environments: a residential room and an office room. These settings were selected because of their differing background contrasts; the residential room featured higher contrast compared to the office environment.

The experiments were conducted as follows:

- Low-light residential room.
- Normal-light residential room.
- Low-light office room.
- Normal-light office room.

Each algorithm was tested in all four scenarios, enabling us to isolate and analyze the impact of lighting and contrast on stereo vision performance.

The **Disparity-based Algorithm** exhibited significant performance degradation under low light conditions. Poor lighting adversely affected the accuracy of person detection by deep learning models, which in turn caused higher errors in disparity estimation. These errors manifested as incorrect or noisy depth values. However, this algorithm performed comparably in both residential and office settings, suggesting that it is relatively robust to variations in image contrast.

The **Depth Map-based Algorithm**, on the other hand, was sensitive to both low light and low contrast environments. This method relies on stereo matching, which attempts to find corresponding points between the left and right camera images to compute depth. Low contrast made this matching process more difficult, while low light further reduced the image quality. However, the contrast issue was somewhat mitigated

Table 1. Quantitative performance of the depth-map and disparity pipelines under four illumination/contrast scenarios.

Scenario	Algorithm	MAE (cm)	RMSE (cm)
Normal-light – residential	Depth-map	2.9 ± 0.4	4.3 ± 0.6
	Disparity	3.5 ± 0.5	5.1 ± 0.7
Normal-light – office	Depth-map	4.4 ± 0.5	6.5 ± 0.7
	Disparity	4.1 ± 0.6	6.0 ± 0.8
Low-light – residential	Depth-map	7.0 ± 0.8	10.3 ± 1.2
	Disparity	5.6 ± 0.7	8.2 ± 1.0
Low-light – office	Depth-map	8.6 ± 1.0	12.5 ± 1.4
	Disparity	7.2 ± 0.9	10.9 ± 1.3

by using the OAK-D Pro built-in infrared (IR) projector, which projects an active pattern on the scene. This projection enhances texture in otherwise low-contrast regions, improving stereo correspondence and depth estimation.

We also identified two additional faults that are not related to environmental conditions, but stem from system-level issues.

The first issue was detected in the depth map–based algorithm, which occasionally produced incomplete depth maps. Specifically, some regions within the map contained anomalous values or failed to compute depth altogether (see Fig. 7). These inaccuracies pose a significant problem during depth extraction, as the resulting data is unreliable.

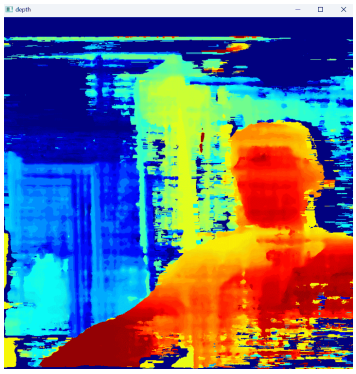


Figure 7. Depth map before postprocessing

To address this issue, we first applied an outlier removal step using local statistics. For each pixel, we computed the mean and standard deviation within a neighborhood $n \times n$ and zeroed any pixel values that significantly deviated from the local mean, based on a predefined threshold. We then used the Telea inpainting algorithm (Telea, 2004) to reconstruct the missing (zero) values by interpolating from nearby valid pixels. Finally, a bilateral filter was applied to smooth the depth map while preserving important edge details (see Fig. 8).

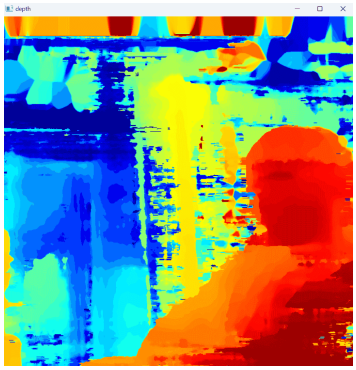


Figure 8. Depth map after postprocessing

These steps improved the overall smoothness and continuity of the depth map. However, they did not fully resolve the underlying issue and, in some cases, introduced new artifacts.

The second fault was observed in the disparity-based algorithm. This issue was traced to the deep learning–based object detectors. Since these detectors do not operate with 100% accuracy, occasional misdetections led to unstable disparity estimations. As a result, the depth calculated would sometimes fluctuate dramatically, jumping to abnormally low or high values, which in turn caused false activations of the safety zone (see Fig. 9).

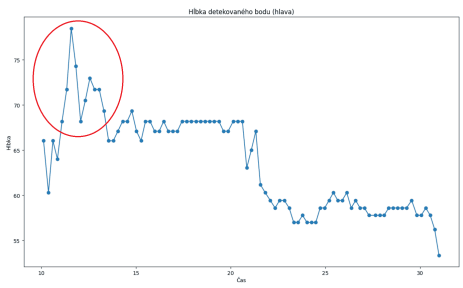


Figure 9. Temporal variation in depth due to detection anomalies

To mitigate this, we experimented with both Kalman filtering and exponential smoothing to suppress sudden peaks and stabilize the depth output. After testing, we found that exponential smoothing yielded better results. Although Kalman filtering effectively reduced noise, it also overly smoothed rapid intentional movements, such as a person walking quickly, causing a loss of responsiveness. In contrast, exponential smoothing maintained stability while adapting more gracefully to fast motion (see Fig. 10).

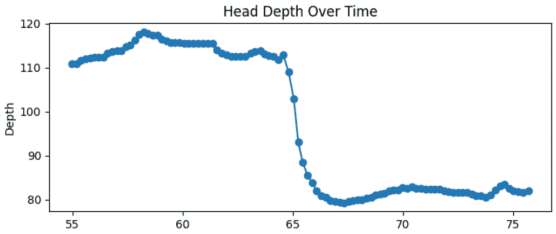


Figure 10. Depth stabilized using exponential smoothing

As before, this approach improved performance but did not completely resolve the problem. Therefore, we implemented a final hybrid solution that combines both depth map-based and disparity-based algorithms, leveraging the strengths of each to enhance robustness and reliability.

5 Hybrid approach

Based on the experiments conducted with our system, we developed a final solution that combines two algorithms into a hybrid approach. This methodology leverages the strengths of each algorithm while mitigating their respective weaknesses. The proposed approach is divided into two main stages. Depth Calculation and Post-processing.

Depth Calculation

Our depth calculation process integrates both algorithms. Initially, we perform person detection independently in both left and right images captured by the stereo camera. This detection yields the corresponding coordinates for keypoints in each image, enabling depth computation through a triangulation based on disparity.

A crucial improvement in our method is the constraint applied during the matching phase: we only accept corresponding points that lie within a 10-pixel threshold of their respective epipolar lines. This constraint significantly reduces the number of erroneous detections.

Subsequently, we employ a depth-map-based algorithm with specific enhancements. Instead of directly using the depth value from a single pixel, we extract a 5×5 pixel neighborhood surrounding the detected keypoint and calculate the average depth within this window. As a result, we obtain two distinct depth measurements:

1. Depth derived from triangulation (d_{tri})
2. Depth derived from the depth map average (d_{map})

The final depth is computed using a weighted combination of these measurements:

$$w_{\text{tri}} = \min\left(1.0, \frac{\text{disparity}}{\text{disparity_threshold}}\right) \quad (1)$$

$$w_{\text{depth}} = 1.0 - w_{\text{tri}} \quad (2)$$

$$D_{\text{final}} = d_{\text{tri}} \cdot w_{\text{tri}} + d_{\text{map}} \cdot w_{\text{depth}} \quad (3)$$

Where:

- w_{tri}: weight for triangulation depth (increases with disparity)
- w_{depth}: complementary weight from depth map
- d_{tri}: depth from stereo triangulation
- d_{map}: average depth around keypoint in depth map
- D_{final}: final depth estimate

The disparity threshold defines the point at which the system begins relying solely on depth calculated from disparity-based algorithm. In our implementation, we used a threshold value of 30 pixels. However, we did not conduct a systematic evaluation of alternative threshold values, so there may be opportunities for further accuracy improvements through parameter tuning.

This weighting scheme ensures that when a person is located further from the camera (where disparity calculations tend to be more error-prone), we rely more heavily on the depth map-based measurement. In contrast, closer objects benefit from more accurate disparity calculations.

Post-processing

The post-processing stage is designed to further refine the calculated depth and is divided into two steps:

1. **Outlier Removal**: We utilize the Z-score method to identify and discard anomalous depth points, significantly improving the reliability of our measurements.
2. **Smoothing**: Depth coordinates are accumulated within a temporal window. If a stationary state is identified across five consecutive windows, robust smoothing is applied:

   $$\text{position} = 0.98 \cdot \text{last_stable} + 0.02 \cdot \text{current}$$

   In dynamic scenarios, median filtering followed by exponential smoothing is applied to effectively minimize noise.

Post-processing Hyperparameters. The following hyperparameters are used during the post-processing stage:

- **Z-score threshold:** A threshold of 2.5 is applied to detect outliers.
- **Depth window:** A 5×5 pixel mean filter is computed around each landmark.
- **Epipolar tolerance:** Vertical mismatches are tolerated up to 10 pixels.
- **Temporal median:** A history of $N = 7$ frames is used for temporal median filtering.
- **Stationary lock:** Activated after 5 consecutive "stable" detection windows.
- **Locked smoothing:** When locked, smoothing is applied as $0.98\times$ previous $+0.02\times$ current value.
- **Dynamic smoothing:** An exponential smoothing factor α in the range $[0.80, 0.90]$ is used.

This dual-step smoothing strategy considerably reduces anomalies in the measured depth of human keypoints, thereby minimizing false activations of the safety zone.

6 Conclusion and Future Work

In this work, we addressed the challenge of implementing virtual safety zones using 3D computer vision. Our main objective was to design and develop a system capable of detecting individuals in a monitored physical space and determining whether they have entered a predefined virtual boundary. Upon zone violation, the system generates a real-time alert.

We proposed a stereo vision–based system integrated with deep learning for human detection. Two primary depth estimation approaches were evaluated: a triangulation-based method using independent keypoint detections from stereo image pairs, and a depth map–based method relying on a single-image keypoint detection with depth values retrieved from a stereo generated map. Each approach demonstrated strengths and weaknesses. The triangulation method provided accurate results for nearby objects, but suffered from occasional detection anomalies. The depth map approach was more stable in keypoint detection, but less reliable in poorly lit or low-texture environments, with incomplete depth maps.

To address these limitations, we implemented several post-processing techniques, including outlier removal, image inpainting, and smoothing using Kalman filtering and exponential smoothing. Although these techniques improved overall stability, they also introduced trade-offs such as delayed system response and increased computational load.

Ultimately, we proposed a hybrid approach that combines both triangulation and depth map–based estimates. This solution balances the strengths of each method, producing more accurate and reliable results at the cost of higher computational demands. In parallel, we developed an interactive visualization interface for defining and monitoring virtual safety zones, detecting zone violations, and visualizing detected keypoints.

Despite the system meeting its core objectives, several limitations remain, most notably in depth estimation accuracy and the stability of stereo keypoint correspondence. Therefore, we propose several avenues for future work:

- **Improved Neural Architectures:** Develop or integrate deep learning models that leverage both stereo image inputs while respecting epipolar constraints to improve correspondence matching and keypoint consistency.
- **Multi-camera Integration:** Extend the system to support multiple stereo camera setups for increased spatial coverage and accuracy, especially in larger or occluded environments.
- **Multi-person Tracking:** Implement advanced tracking mechanisms to enable simultaneous monitoring of multiple individuals and reduce false zone violations.

Our findings indicate that combining stereo vision and deep learning offers a viable path toward intelligent safety zone monitoring. However, due to its current limitations, we do not yet recommend this approach for safety-critical environments requiring high-precision tracking, until the proposed enhancements are implemented and validated.

Privacy and ethics

The proposed system is currently in the proof-of-concept stage. In its current version, it is not deployed in an industrial environment. When processing data, the source data is deleted after the camera data has been read. The system does not archive any data that could be attributed to a specific person.

Acknowledgments

This work was supported by KEGA through the Teaching and Development of methodology for the use of microcontrollers in automation using practical examples and laboratory exercises for engineering students, under Grant 024STU-4/2024.

References

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T. L., Zhang, F., & Grundmann, M. (2020). Blazepose: On-device real-time body pose tracking. *ArXiv*, *abs/2006.10204*. https : / / api . semanticscholar.org/CorpusID:219793039

Cho, H., Lee, K., Choi, N., Kim, S., Lee, J., & Yang, S. (2022). Online safety zone estimation and violation detection for nonstationary objects in workplaces. *IEEE Access*, *10*, 39769–39781. https://doi.org/10.1109/ACCESS.2022.3165821

Kozamernik, N., Zaletelj, J., Kosir, A., Suligoj, F., & Bracun, D. (2023). Visual quality and safety monitoring system for human-robot cooperation. *The International Journal of Advanced Manufacturing Technology*, *128*, 685–701. https://doi.org/https://doi.org/10.21203/rs.3.rs-2409100/v1

Luxonis Inc. (2023). Calculate spatial coordinates on the host [DepthAI Examples repository (sample code and README explain depth-lookup for arbitrary key-points); accessed 28 Jun 2025.].

Makris, S. (2021). Dynamic safety zones in human robot collaboration. In *Cooperating robots for flexible manufacturing* (pp. 271–287). Springer International Publishing. https://doi.org/10.1007/978-3-030-51591-1_14

Mohammadi Amin, F., Rezayati, M., van de Venn, H. W., & Karimpour, H. (2020). A mixed-perception approach for safe human–robot collaboration in industrial automation. *Sensors*, *20*(21). https://www.mdpi.com/1424-8220/20/21/6347

Mosberger, R., Andreasson, H., & Lilienthal, A. J. (2014). A customized vision system for tracking humans wearing reflective safety clothing from industrial vehicles and machinery. *Sensors*, *14*(10), 17952–17980. https://www.mdpi.com/1424-8220/14/10/17952

Rybski, P., Anderson-Sprecher, P., Huber, D., Niessl, C., & Simmons, R. (2012). Sensor fusion for human safety in industrial workcells. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3612–3619. https : / / doi . org / 10 . 1109 / IROS.2012.6386034

Sládek, I., Gašpar, G., Repka, J., & Budjac, R. (2024). Stereo vision based 3d positioning for real-time control applications. *Proceedings of 35th International scientific conference CECIIS*.

Slovák, J., Melicher, M., Šimovec, M., & Vachálek, J. (2021). Vision and rtls safety implementation in an experimental human—robot collaboration scenario. *Sensors*, *21*(7). https : / / www . mdpi . com / 1424-8220/21/7/2419

Telea, A. (2004). An image inpainting technique based on the fast marching method. *Journal of Graphics Tools*, *9*(1), 23–34. https://doi.org/10.1080/10867651.2004.10487596

Zhou, C., Ren, D., Zhang, X., Yu, C., & Ju, L. (2022). Human position detection based on depth camera image information in mechanical safety. *Advances in Mathematical Physics*. https://doi.org/10.1155/2022/9170642

Application of Artificial Intelligence and Internet of Things in Rehabilitation and Health Recovery

Ivan Kolarić, Karlo Pepelnik, Dino Vlahek, Dijana Oreški,
Jelena Gusić Munđar, Bojan Žugec, Nikola Kadoić
University of Zagreb Faculty of organization and informatics, Pavlinska 2
42000 Varaždin, Croatia
{ikolaric, kpepelnik, dvlahek, dijoresk, jelgusic, bzugec, nkadoic}@foi.hr

Abstract. *Both Artificial Intelligence (AI) and the Internet of Things (IoT) optimize different aspects of human life. This is also the case with human health. This research focuses on the application of AI and IoT in rehabilitation and health recovery, with a particular emphasis on physical recovery and kinesiotherapy. The primary goals of this literature review are twofold. First, to analyse the literature related to the application of AI and IoT in rehabilitation and health recovery, focusing on the addressed health issues, applied AI solutions, IoT and robotics solutions, technology readiness levels of AIoT systems, and publications and their authors (years of publishing, journals, conferences, countries of origin). Second, to analyse papers related to physical therapy and kinesiotherapy AI and IoT to determine how those they are integrated. The literature review was implemented using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach. In total, 77 papers were analysed. They are mainly journal papers. The number of publications has increased over the years. When analysing the AIoT solutions related to physical therapy, most systems are in prototype or pilot stages and show promising results towards better recovery and quality of life for patients.*

Keywords: AI, IoT, sensors, health, physical rehabilitation

1 Introduction

This paper has been prepared under the project "Research and Development of an Integrated Artificial Intelligence – Internet of Things (AIoT) Solution for the Personalization of Physical Therapy and Kinesitherapy – AIoT4Therapy", which aims to develop a comprehensive AIoT solution for physical therapy and kinesiotherapy, integrating innovative sensor systems, multi-criteria decision-making, advanced AI algorithms to optimize the therapeutic process, and specially trained language models to advise patients and experts. The solution will be available to end users in the form of an intuitive, easy-to-use digital tool. The goal of the project is to enable more personalized therapeutic procedures, ensure continuous system improvement based on patient feedback, and provide a platform for more effective education of experts and patients, particularly through large language models (LLMs). When implemented, the solution aims to significantly enhance the effectiveness of existing therapeutic procedures, reduce the incidence of musculoskeletal diseases through everyday use, and improve the overall quality of life for patients.

The goal of this paper is to analyse the current state of the application of AI and IoT in health recovery and rehabilitation. The research is a literature review, conducted using a PRISMA approach (Page et al., 2021) on the scientific articles databases, Web of Science (WoS) and Scopus. The analyses were focused on identification of (1) health recovery issues that are analysed using AI and IoT, (2) AI methods, approaches and solutions that were used to deal with the rehabilitation and health recovery issues, (3) IoT devices that were used to deal with the rehabilitation and health recovery issues, (4) levels of application of AIoT healthcare solutions for rehabilitation and recovery issues, and (5) descriptions of AIoT solutions primarily related to physical therapy;and the determination of (1) the distribution of the number of papers considering the year of publication, (2) the distribution of the number of papers per countries (based on authors affiliations' countries), (3) the number of papers considering the publications and the types of publications.

The paper is organised as follows: Section 2 describes the potential of AI and IoT devices in healthcare. Section 3 presents the research methodology. Section 4 presents the results of the research. Section 5 concludes the paper.

2 AI and IoT in Healthcare

The following two paragraphs present claims found in the existing literature and do not reflect the original theses of this paper. Artificial Intelligence (AI) and the Internet of Things (IoT) are transformative technologies that are reshaping industries worldwide. Below is an explanation of these technologies and their

applications across various sectors. AI refers to the simulation of human intelligence in machines that can perform tasks such as learning, reasoning, problem-solving, and decision-making. It encompasses technologies like machine learning (ML), natural language processing (NLP), and computer vision, which allow systems to analyse data, recognize patterns, and make predictions or decisions. IoT involves a network of interconnected devices that collect, share, and analyse data through the internet. These devices include sensors, wearables, and smart appliances that enable real-time monitoring, automation, and data-driven decision-making.

AI and IoT technologies are integrated across diverse sectors, including healthcare, smart cities, manufacturing, and agriculture, utilizing sensor data and machine learning algorithms for various applications. The combination of AI and IoT enhances their individual capabilities: IoT provides massive amounts of real-time data from connected devices. AI processes this data to generate actionable insights, enabling smarter automation, predictive analytics, and improved decision-making. For example, in healthcare, IoT devices collect patient data while AI analyses it for early diagnosis. In manufacturing, IoT sensors monitor equipment while AI predicts maintenance needs. Both technologies are driving innovation across industries by improving efficiency, reducing costs, and enhancing user experiences.

Some studies report applications in healthcare, smart cities, industrial manufacturing, robotics, and automation (Ma et al., 2019). In healthcare, Mahalakshmi (2019) and Oniani et al. (2020) describe systems that use machine learning, support vector machines, k-nearest neighbours, and other algorithms in combination with medical sensors and Internet of Medical Things devices (Oniani et al., 2021; S Mahalakshmi, 2019). These studies detail diagnostic tools, patient data analysis, predictive disease management, and robotic surgery enhancements. Other studies illustrate urban and industrial uses. Ma et al. (2019) and Mali (2019) note smart public services and intelligent manufacturing that rely on IoT sensor data and AI analytics for tasks such as traffic monitoring and process optimization. G Tzafestas (2018) discusses the integration of AI with IoT to improve robotics, automation, and industrial operations. Additional application areas include smart homes, transportation, agriculture, and marketing, each mentioned in one study (G Tzafestas, 2018; Mali, 2019).

Innovation in medical treatment relies on formal R&D processes, clinical practice feedback, and enabling factors like public-private partnerships and regulatory frameworks (Kaitin et al., 2011; Kanavos et al., 2010). Formal research and development processes support innovation in medical treatment (Cheng et al., 1996; Garber et al., 2014; Lo et al., 2022). Studies document a shift from disorganized beginnings toward refined, targeted approaches in pharmaceutical and anticancer drug development (Liu et al., 2017). For example, analyses reveal an increased focus on priority products even as overall approval rates decline, while historical reviews note a move toward targeted and immune-related therapies (Mahoney, 2011).

Clinical practice feedback loops also drive innovation. Investigations into medical device development and mechanism-based pain diagnosis show that input from end users helps refine and direct product design (Woolf et al., 2001). Technology transfer through product development partnerships facilitates the adoption of non-medical technologies for medical use. Moreover, collaborative public–private partnerships, evolving regulatory frameworks that both constrain and hasten approval, and robust funding structures emerge as key enablers in shaping a dynamic, non-linear innovation process (Kanavos et al., 2010).

Many studies report that integrating AI and IoT (or IMU (Inertial Measurement Unit)) in healthcare can improve remote monitoring, diagnostics, and overall clinical care. Chintala (2024) describes a remote patient monitoring system in which IoT devices combined with AI algorithms yield better clinical outcomes, lower costs, and higher patient satisfaction (Chintala, 2024). Dwivedi et al. (2021) and Ejiyi et al. (2023) detail layered and edge-based Internet of Medical Things (IoMT) systems that support telehealth, self-monitoring, rehabilitation, and chronic disease management while enhancing patient outcomes and cost-effectiveness (Dwivedi et al., 2022; Ejiyi et al., 2023). Ghode et al. (2024) presents a healthcare analytics framework that increases diagnostic accuracy and enables personalized care through real-time data analysis. One review of IoT applications outlines use cases in patient monitoring and supply chain management without detailing specific outcomes (Akoh Atadoga et al., 2024). Common themes include real-time insights for early diagnosis and rapid intervention, along with challenges in data privacy, interoperability, and system integration that may affect implementation (Chintala, 2024; Ghode et al., 2024).

The most promising AI and IoT technologies in healthcare today are AI-powered diagnostics, robotic surgery, personalized medicine, virtual health assistants, drug discovery, internet of medical things, ambient clinical intelligence, precision imaging and predictive analytics for emergency care.

3 Methodology of the Research

In this research, we defined several research questions (RQ) to achieve the research goals:

RQ1. Which rehabilitation and health recovery issues are analysed using AI and IoT?
RQ2. Which AI solutions, methods and approaches are used to deal with rehabilitation and health recovery issues?

RQ3. Which IoT devices are used to deal with rehabilitation and health recovery issues?
RQ4. What are the levels of application (or technology readiness levels, TRLs) of IoT-AI-healthcare solutions for rehabilitation and health recovery reported in the literature?
RQ5. What is the distribution of papers considering the year of publication?
RQ6. What is the geographical distribution of the authors of the papers related to analysing rehabilitation and health recovery issues using AI and IoT?
RQ7. What is the type of publication where the papers are published (conferences and/or journals)?
RQ8. What are the strengths and weaknesses of AIoT solutions created for optimize physical therapy and kinesiotherapy?

Literature review using PRISMA approach was applied to answer the previous research questions. The PRISMA approach is a widely recognized framework designed to enhance the transparency, quality, and reproducibility of systematic reviews and meta-analyses. It provides a structured methodology for reporting the process and findings of such reviews.

The databases that were searched in are Web of Science (WoS) and Scopus.

The keywords that were searched are:

- AI IOT Therapy,
- AI IOT Kinesiotherapy,
- AI IOT Physical therapy
- AI IOT Rehabilitation,
- AI IOT Recovery Health,
- AI IMU Therapy,
- AI IMU Kinesiotherapy,
- AI IMU Physical therapy,
- AI IMU Rehabilitation.

Inside each item in the list, there is an operator AND, and among all items in the list, there is an operator OR. Fig. 1 presents the flow diagram of PRISMA steps implemented in the research considering the research questions.

After the initial search, there were 111 records in WoS, and 178 records in Scopus. Many of the records were found in both databases (the number of duplicated records was 105). After duplicates removal, there were 184 records. In a subsequent step, an open-access filter was applied, and 79 records remained for analysis. During the access to the full papers, two papers were not found, so 77 papers remained for the analysis. Those 77 papers cover all medical categories since rehabilitation is associated with different medical states.

Furthermore, it was decided that RQs 1 to 7 will be analysed across the entire set of 77 papers to describe the state of AIoT application in all healthcare categories, and RQ 8 will be analysed in the sample that remains after applying the filter health category = physical disease, 12 papers since the AIoT4Therapy project is related to physical therapy.

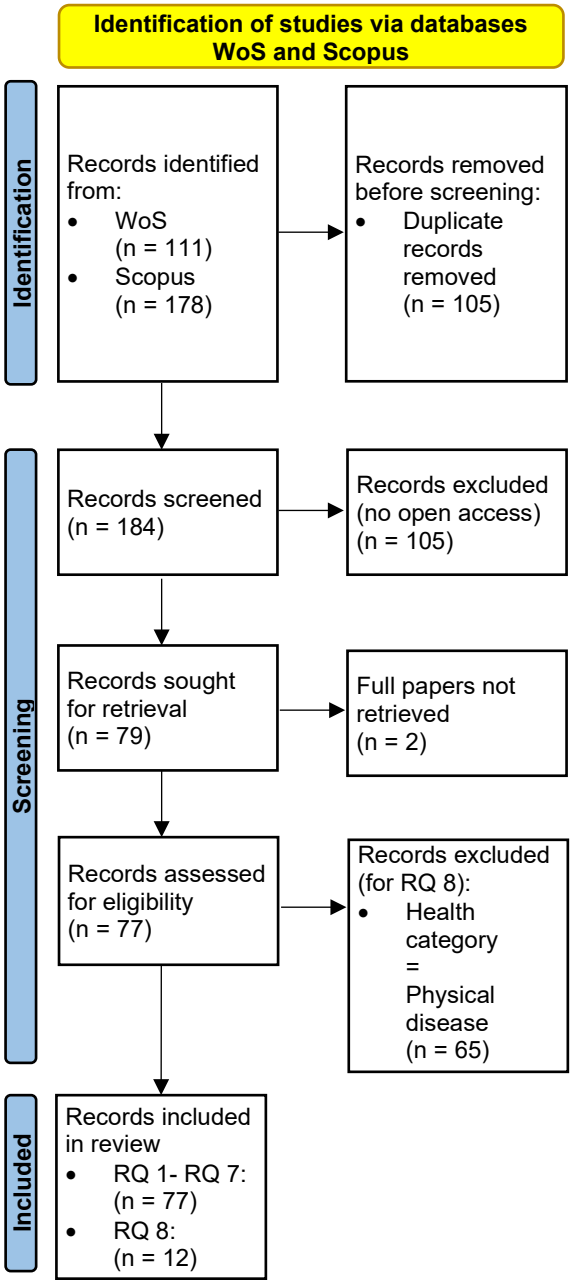


Figure 1. PRISMA flow diagram

4 Results

4.1 AI and IoT in rehabilitation and health recovery

The distribution of the number of papers respecting the health category is presented in Table 1. The highest number of papers discuss the general application of AI and IoT in healthcare. Considering concrete diseases, the highest number of papers are related to physical disease (12), followed by stroke rehabilitation (7), cancer (6) and mental health (6). The following tables are primarily sorted alphabetically (Tables 1–4 and 6). Table 5 is sorted chronologically by year.

Table 1. Number of papers per health category (RQ1)

Health category	Number of records
Cancer	6
Chronic disease	5
Healthcare in general	30
Heart disease	5
Mental health	6
Not healthcare	*3*
Ophthalmology	1
Physical disease	12
Stroke rehabilitation	7
Traumatic brain injury	2
Grand total	77

Table 2. Number of papers per AI solutions, methods and approaches (RQ2)

AI solutions	Number of records
AI (general)	70
Artificial Neural Network	8
Augmentation methods	14
Big Data Analytics	45
Convolutional Neural Network	14
Deep Learning	35
Edge computing	16
Machine Learning	55
Natural Language Processing	3
Simulations	23
Simultaneous Localization and Mapping	1

The distribution of the number of papers with respect to AI solutions, methods and approaches that were used or discussed in the context of rehabilitation and health recovery is presented in Table 2. As can be seen in the table, not all AI solutions, methods, and approaches are at the same level. Convolutional Neural Networks belong to the category of Artificial Neural Networks, which are a subset of deep learning. Furthermore, deep learning is a subset of machine learning, and all methods share a common umbrella of AI.

The distribution of IoT devices which are used to deal with rehabilitation and health recovery issues is presented in Table 3. The most commonly used IoT devices are sensors (some papers specify the type of sensors, so we identified GPS sensors, vital signs sensors, optical motion sensors, among others), followed by wearables and Wi-Fi devices.

Furthermore, we analysed all papers and categorised them based on the level of development of AIoT devices. Even though there are commonly accepted technological readiness levels (TRLs), it was hard to distinguish between levels of some AIoT solutions in papers, so we introduced a simpler scale for this purpose. As can be seen in Table 4, we differ (1) papers that are related to the review of trends (that theoretically discuss the possibilities of AI and IoT application in rehabilitation and health recovery, without proposing any concrete solution; some of those papers were literature reviews), (2) papers that present the ideas how certain AIoT can be introduced (ex. Ergen and Belcastro (2019) presented the concept of AIoT integration in medicine), (3) papers that present the designs of AIoT systems (ex. Gomez-Valiente et al. (2023) presented an IoT architecture to support a smart IoT that focuses on addressing the challenges of scheduling, resource allocation, and state management of Smart IoT systems), (4) the papers that describe the prototypes of the systems and (5) papers that present the results of the application of prototypes in the real environment (ex. Moghbelan et al. (2024) proposed a smart motor rehabilitation system which leverages wearable sensors to monitor patients' vital signs and edge computing to detect and estimate motor routines). The highest number of papers is related to the category of the review of trends, followed by the prototype applications and designs of the AIoT systems. Some examples are described in Section 4.2.

Table 3. Number of papers per IoT devices (RQ3)

IoT devices	Number of records
AR/VR	9
Bed mat	2
Cameras (2D/3D)	22
Doppler/Ultrasound	2
Floor sensors	3
GPS Sensors	2
IMU	12
Optical motion sensor (Kinect)	2
Robots	11
Sensors	63
Triboelectric nanogenerator (TENG) sensors	2
Vital signs sensor	28
Wearables	40
Wi-Fi-enabled Microscopic Device	35

Table 4. Number of papers per level of AIoT system development (RQ4)

Level of AIoT system development	Number of records
AIoT idea	2
AIoT system design	11
Prototype description	1
The application of prototypes in a real environment	29
The review of trends (including literature reviews)	34
Grand total	77

Considering the year of publication and the type of publication, the results are presented in Table 5. We

can see that the first ideas of applying AI and IoT for rehabilitation and health recovery were documented in 2019, and that the number of papers increases each year.

The papers were mainly published in journals, in more than 50 journals. The most popular journals are Applied Sciences (MDPI), Sensors (MDPI) and IEEE Access (IEEE), Computers (MDPI), Healthcare (MDPI), IEEE Internet of Things Journal (IEEE) and Information (MDPI). Other journals published only one paper.

Table 5. Number of papers per year and publication type (RQ5)

Year	Journal	Conference	Total
2019	2		2
2020	3		3
2021	9		9
2022	11	1	12
2023	19		19
2024	27	2	29
2025	3		3
Grand total	74	3	77

Table 6. Number of papers per country (RQ6)

Country	Number of papers	Country	Number of papers
Australia	9	Norway	2
Bangladesh	3	Pakistan	4
Belgium	1	Poland	1
Brazil	2	Portugal	1
Canada	6	Saudi Arabia	10
China	15	Serbia	1
Ecuador	1	Singapore	3
Egypt	2	South Korea	6
Germany	1	Spain	5
India	16	Sweden	1
Iran	4	Taiwan	6
Israel	1	Thailand	1
Italy	6	Tunisia	1
Japan	1	Turkey	2
Lithuania	1	United Arab Emirates	1
Malaysia	3	United Kingdom	4
Nepal	1	United States of America	10
Nigeria	1		

Table 6 presents the distribution of papers per country, where the authors of papers come from. The highest number of papers is authored by Chinese and also Indian authors, followed by authors from Saudi Arabia, the United States of America and Australia. Authors from Asia dominate in the field.

4.2 AIoT-based solutions for physical therapy

In this subsection, we present descriptions of AIoT papers that belong to the health category of physical disease (RQ8). The last step from the PRISMA methodology finished with 12 papers, and they are analysed further.

Qayyum et al. applied machine learning to evaluate a herbal remedy (linseed, psyllium, and honey pessary) versus standard antibiotics for pelvic inflammatory disease. A clinical trial showed the herbal treatment is a safe and effective alternative, with the treatment group achieving over 70% improvement in key clinical outcomes. Patients on the herbal therapy had slightly greater reductions in symptoms (e.g. abnormal discharge reduced by 87% vs. 78% in controls). ML classifiers (decision tree, random forest, etc.) could only moderately distinguish the two groups (~53–62% accuracy), indicating comparable efficacy. High patient compliance (~98%) was observed in both groups, supporting the herbal AIoT approach as a viable therapy option (Qayyum et al., 2023).

Zhang et al. developed wearable triboelectric nanogenerator sensors for gait and waist motion tracking in lower-limb rehabilitation. Four textile-based sensors on a belt capture waist movement for real-time robot control and AR/VR exercises, while two insole sensors monitor gait. Using an ANN, the system achieved 98.4% accuracy in identifying individuals from gait signals (tested on five users). The integrated system, including a rehab robot and immersive game, performed well in user recognition and motion monitoring, enabling gaming-aided training and demonstrating potential for IoT-based smart rehabilitation. This low-cost, energy-efficient AIoT solution is presented as a prototype for future smart healthcare (Zhang et al., 2022).

Rahman et al. explored upper-limb rehabilitation by classifying elbow joint angles using surface EMG signals. They recorded biceps muscle activity at five fixed elbow angles and extracted time-domain features. A k-NN classifier achieved high accuracy, especially with k-fold cross-validation (mean ~89.7% vs. 82.5% with leave-one-out). This result demonstrates that a wearable EMG sensor can discern elbow flexion angles with nearly 90% accuracy, which could aid in developing advanced assistive devices and remote monitoring of rehabilitation exercises (Rahman et al., 2021).

Lin and Wai proposed an AIoT-based fall evaluation and prevention program for community-dwelling older adults. Their system uses gait and balance gesture data (based on sarcopenia criteria) with adaptive training courses reviewed bi-weekly by experts. In a 3-month field intervention, participants significantly improved their gait speed (mean increase of ~30% in males and 34% in females). The study demonstrates that AI-guided rehabilitation can improve mobility in sub-healthy seniors, and it

provided insights into how gait and gesture patterns relate to stability, enabling targeted interventions to minimize fall risk. This pilot highlights effective health promotion through AIoT in long-term care (Lin et al., 2021).

Nadian-Ghomsheh et al. introduced a hierarchical IoT architecture for remote hand rehabilitation that emphasizes privacy. The system uses camera-based range-of-motion (ROM) tracking for hand/wrist exercises, with edge processing (including federated learning for skin detection) to avoid sending sensitive images to the cloud. A machine-learning module immediately assesses patient exercises and extracts ROM data to track progress. In tests, the vision-based method (using simple colored markers and standard RGB cameras) measured wrist/finger ROM with high accuracy, comparable to a goniometer and outperforming a depth camera (Leap Motion) in precision. The solution achieves real-time, accurate exercise assessment in a privacy-preserving manner, demonstrating an effective tele-rehabilitation tool that addresses data protection concerns in AIoT healthcare (Nadian-Ghomsheh et al., 2021).

Moghbelan et al. proposed the I-TROPHYTS framework, integrating wearable IoT sensors, edge computing, and a humanoid robot for group motor rehabilitation exercises. The system monitors patients' vitals and movements (via IMUs on the wrist and elbow) and uses an ontology-driven approach to model therapy sessions. The humanoid robot acts as an adaptive physiotherapist, demonstrating and adjusting exercises in real-time based on patient status. A pilot with volunteers doing shoulder exercises achieved nearly 100% accuracy in automatically distinguishing different movements (shoulder abduction vs. flexion) using machine learning on the IMU data. These results validate the framework's ability to automatically detect exercise types with high accuracy, supporting personalized and sustainable rehabilitation through AIoT and robotics (Moghbelan et al., 2024).

Shefa et al. designed a smart ankle–foot orthosis (AFO) system for gait rehabilitation, combining IoT sensors and deep learning. The AFO device is outfitted with a surface EMG sensor (to read muscle activity) and an IMU (to track leg motion), streaming data via fog computing for analysis. Multiple AI models were tested to classify walking patterns; a Transformer model performed best, achieving 98.97% accuracy in recognizing gait phases (normal vs. aberrant). This high accuracy enables the system to precisely detect abnormal gait events and quantify improvements in walking patterns. The platform can then provide personalized recommendations on AFO usage (duration and intensity) tailored to the patient's recovery needs, with clinicians reviewing the data for validation. This AIoT approach shows effectiveness in optimizing gait training and supporting data-driven orthotic therapy (Shefa et al., 2024).

Yao et al. (2023) developed a multifunctional smart heating pad to aid rehabilitation and pain relief, focusing on usability for older adults. The device uses an IoMT architecture with an onboard microcontroller and Bluetooth, employing fuzzy logic at the edge to regulate temperature without cloud support. The team designed a custom temperature sensor circuit with high stability (achieving ±0.3 °C accuracy in the therapeutic range). Experimental results confirmed the system's feasibility: the IoT controller self-calibrates and maintains target temperature ranges for different therapy modes using fuzzy rules, and achieves stability on par with expensive industrial controllers while using low-cost components. The heating pad's multiple modes (for various body areas or conditions) and auto-calibration make it user-friendly for seniors, illustrating how AIoT edge computing can improve traditional therapeutic devices (Yao et al., 2022).

Petrovic et al. conducted an experimental study using IoT force and EMG sensors to assess the safety of repetitive pushing/pulling tasks (simulating industrial cart handling) in relation to workers' pain and mental stress. Two 3-axis force sensors (on cart handles) and six EMG electrodes (on major upper-body muscles) collected data from participants with and without musculoskeletal pain or high psychological stress. Analysis revealed clear differences: subjects with pain or elevated stress showed significantly higher exertion (force integral) and muscle activation (EMG mean values) during the task. The study is the first to jointly evaluate how pain, spinal mobility, and psychological state affect task execution, concluding that these factors significantly degrade task performance and likely raise injury risk. The authors suggest using these AIoT measurements for personalized risk assessment in the future, with plans to incorporate computer vision (posture tracking) and EEG-based cognitive load monitoring to improve the system (Petrovic et al., 2022).

Alsareii et al. demonstrated a pilot IoT framework for remotely monitoring post-surgical patients using machine learning and ultra-reliable communication. Their system optimizes IoT network settings to ensure low-latency, stable transmission of patients' vital signs and activity data, even for critical patients. In simulations, the proposed network achieved reduced delay and packet loss compared to standard low-latency networks, improving reliability of data delivery. Additionally, a gradient-boosting regression model was employed to predict patient vital signs locally. This AI model could accurately forecast both slow-changing and fast-changing vital signals, effectively acting as a "virtual sensor" to fill in data gaps during connectivity outages. The results indicate that AIoT can enable robust, continuous monitoring for recovery patients, ensuring timely detection of issues despite network disruptions. This pilot underscores the feasibility of combining optimized IoT communications with ML prediction to enhance post-operative care (Alsareii et al., 2022).

Al-Hababi et al. introduced a novel non-contact sensing testbed using wireless signals and AI to

monitor rehabilitation exercises in post-surgery patients. Instead of wearables or cameras, the system employs a Universal Software Radio Peripheral (USRP) to capture WiFi channel state information (WCSI) as patients perform prescribed movements (e.g. a weight-lifting exercise). Machine learning (a fine-tuned k-NN classifier) analyses the wireless signal perturbations to classify whether the patient executed the movement correctly or incorrectly. The experimental setup (with data from spinal surgery patients lifting weight) achieved 99.6% classification accuracy in distinguishing correct vs. incorrect form. This highlights an innovative AIoT approach for physical therapy: leveraging ubiquitous wireless signals to passively monitor exercise compliance and technique. Such a system could remind patients to maintain proper form (reducing complications) without requiring cameras or attached sensors, though it remains at a prototype stage (Al-Hababi et al., 2020).

Hasan et al. conducted a scoping review of digital health interventions for musculoskeletal pain, focusing on systems that use AI, IoT, and ML to improve outcomes. The reviewed evidence indicates that AIoT-driven solutions (including smart exercise coaching apps, remote monitoring tools, and even brain-computer interfaces) can help reduce pain and improve functional impairment in chronic conditions like low back pain and osteoarthritis. A key benefit is improved adherence to prescribed exercise therapy: intelligent systems provide reminders, feedback, and motivation, helping patients perform rehab exercises regularly to relieve pain. Overall, the review found that such digital therapies have received largely favourable feedback from patients and can offer convenient, cost-effective care. However, some users reported dissatisfaction, and current computer-aided rehab systems still "lack flexibility and robustness" in fully tailoring to individual needs. The authors conclude that AIoT-based interventions show substantial promise in musculoskeletal rehabilitation, but wider adoption and more research are needed to generalise their use and maximise patient outcomes (Hasan et al., 2023).

The analysis of 12 studies related to AIoT in rehabilitation is upgraded with the analysis of the advantages and limitations presented in Table 7.

The integration of Artificial Intelligence (AI) and the Internet of Things (IoT) into physical rehabilitation has marked a significant shift toward more personalized, data-driven, and accessible healthcare solutions. Current AIoT applications range from wearable sensors and smart orthoses to non-contact monitoring systems and robotic assistants, offering real-time feedback, precise motion analysis, and remote supervision. These technologies have demonstrated promising results in enhancing therapy adherence, optimizing treatment outcomes, and supporting clinicians with objective performance data. Emerging trends include privacy-preserving

Table 7. Analysis of advantages and limitations

Study	Advantages	Limitations
Qayyum et al. (2023)	Clinically effective herbal alternative; high patient adherence; non-antibiotic option	Limited diagnostic classification power of Machine Learning (~62% accuracy); Pelvic Inflammatory Disease-specific study
Zhang et al. (2022)	High identification accuracy (98.4%); immersive rehab gaming; low-power wearable design	Only tested on 5 users; limited generalizability
Rahman et al. (2021)	High angle classification accuracy (~90%)	Only elbow joint tested; simple movement classification
Lin et al. (2021)	Improved gait metrics in elderly; low-cost fall risk monitoring; real-world intervention study	Short duration (3 months); limited to Taiwan community setting
Nadian-Ghomsheh et al. (2021)	Privacy-preserving architecture; high Range of Motion measurement precision; real-time edge processing	Marker setup required; Range of Motion-focused, no muscle feedback
Moghbelan et al. (2024)	High exercise recognition accuracy (~100%); adaptive group therapy via robot; ontology-based logic	Prototype stage; complex hardware integration
Shefa et al. (2024)	High gait phase accuracy (~99%), personalized feedback	Focus on gait; needs Ankle–Foot Orthosis hardware
Yao et al. (2022)	Accurate self-calibrating temperature control; low-cost; older adult friendly	Passive therapy only; limited to heating/pain relief
Petrovic et al. (2022)	Multimodal data fusion (pain, stress, spinal mobility); real work-task safety assessment	Assessment-focused; no rehabilitation feedback or therapy
Alsareii et al. (2022)	High resilience in patient monitoring; local Artificial Intelligence prediction during outages	Simulation results only; no clinical deployment
Al-Hababi et al. (2020)	Very high form classification accuracy (99.6%); no wearables or cameras needed	Specific use-case; early prototype, small study size
Hasan et al. (2023)	Broad evidence, improved adherence, cost-effective	Limited robustness/flexibility in current systems

architectures, edge computing, and immersive rehabilitation environments using augmented or virtual reality. Despite the promising advances, many solutions remain in prototype or pilot stages, with limited clinical validation and scalability. Looking forward, the continued development of robust, interoperable, and user-centred AIoT systems is expected to expand access to rehabilitation services, especially for ageing populations and remote communities, while promoting a shift toward preventive and personalised care models.

5 Conclusion

The number of studies dealing with the application of AIoT for rehabilitation purposes increases from year to year. They are related to several health categories, mostly to healthcare in general and physical. The most active authors come from Asia (India and China). The papers are published in a large number of journals. Most of the papers present the developed prototypes and their applications, which often correspond to the TLR 4 and higher.

AIoT-based systems in physical therapy and kinesiotherapy demonstrate significant potential for improving rehabilitation through continuous monitoring, personalized feedback, and intelligent intervention. Studies show high accuracy in detecting patient states and positive health outcomes, including enhanced mobility, pain reduction, and better adherence to therapy. AI-driven personalization ensures treatments adapt to individual needs, while advanced sensing technologies enable real-time progress tracking.

Despite promising results, challenges remain, such as limited large-scale validation, data privacy concerns, and integration into clinical workflows. Many systems are still in prototype stages, requiring further refinement for usability, reliability, and wide-spread adoption. Future research should focus on long-term clinical trials and robust security measures to enhance trust and effectiveness.

With continued advancements, AIoT solutions are poised to transform rehabilitation by making therapy more accessible, data-driven, and effective, ultimately improving patient recovery and quality of life.

Acknowledgments

This paper was supported by the project “AIoT4Therapy - Research and development of an integrated AIoT solution for the personalization of physical therapy and kinesiotherapy” funded by European Regional Development Fund, IP.1.1.03; Strengthening strategic partnerships for innovation in the process of industrial transition (Integrated Territorial Program 2021-2027). The project was approved by the Croatian Agency for Small Business, Innovations and Investments (HAMAG BICRO) and the Ministry of Regional Development and European Union Funds in the Republic of Croatia.

References

Al-hababi, M. A. M., Khancheng, M. B., Al-Turjman, F., Zhao, N., & Yang, X. (2020). Non-Contact Sensing Testbed for Post-Surgery Monitoring by Exploiting Artificial-Intelligence. Applied Sciences, 10(14), 4886. https://doi.org/10.3390/app10144886

Alsareii, S. A., Raza, M., Alamri, A. M., AlAsmari, M. Y., Irfan, M., Khan, U., & Awais, M. (2022). Machine Learning and Internet of Things Enabled Monitoring of Post-Surgery Patients: A Pilot Study. Sensors, 22(4), 1420. https://doi.org/10.3390/s22041420

Cheng, Y.-T., & Van de Ven, A. H. (1996). Learning the Innovation Journey: Order out of Chaos? Organization Science, 7(6), 593–614. doi: 10.1287/orsc.7.6.593

Chintala, S. (2024). Iot and Ai Synergy: Remote Patient Monitoring for Improved Healthcare. 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM), 1–6. doi: 10.1109/ICIPTM59628.2024.10563530

Dwivedi, R., Mehrotra, D., & Chandra, S. (2022). Potential of Internet of Medical Things (IoMT) applications in building a smart healthcare system: A systematic review. Journal of Oral Biology and Craniofacial Research, 12(2), 302–318. doi: 10.1016/j.jobcr.2021.11.010

Ejiyi, C., Qin, Z., Ejiyi, M. B., Nneji, G. U., Monday, H. N., Agu, F. A., Ejiyi, T. U., Diokpo, C., & Orakwue, C. O. (2023). The internet of medical things in healthcare management: a review. Journal of Digital Health, 30–62. doi: 10.55976/jdh.22023116330-62

Ergen O, Belcastro KD. (2019.) Ai Driven Advanced Internet Of Things (Iotx2): The Future Seems Irreversibly Connected in Medicine. Anatol J Cardiol. 22(Suppl 2):15-17. doi: 10.14744/AnatolJCardiol.2019.73466. PMID: 31670717.

G Tzafestas, S. (2018). Synergy of IoT and AI in Modern Society: The Robotics and Automation Case. Robotics & Automation Engineering Journal, 3(5). doi: 10.19080/RAEJ.2018.03.555621

Garber, S., Gates, S. M., Keeler, E. B., Vaiana, M. E., Mulcahy, A. W., Lau, C., & Kellermann, A. L. (2014). Redirecting Innovation in U.S. Health Care: Options to Decrease Spending and Increase Value. Rand Health Quarterly, 4(1), 3.

Ghode, S., Suneel, S., Yashwanth Reddy, A., Isabels, K. R. & Ramkumar A. (2024.) A Holistic Approach to Healthcare Analytics Using IoT and AI Technologies. Nanotechnology Perceptions, 67–76. doi: 10.62441/nano-ntp.vi.3439

Gómez-Valiente, P., Benedí, J. P., Lillo-Castellano, J. M. and Marina-Breysse, M. (2023) Smart-IoT Business Process Management: A Case Study on Remote Digital Early Cardiac Arrhythmia Detection and Diagnosis. IEEE Internet of Things Journal, vol. 10, no. 19, pp. 16744-16757, 1 doi: 10.1109/JIOT.2023.3269820.

Hasan, F., Mudey, A., & Joshi, A. (2023). Role of Internet of Things (IoT), Artificial Intelligence and Machine Learning in Musculoskeletal Pain: A Scoping Review. Cureus. doi: 10.7759/cureus.37352

Kaitin, K. I., & DiMasi, J. A. (2011). Pharmaceutical Innovation in the 21st Century: New Drug Approvals in the First Decade, 2000–2009. Clinical Pharmacology & Therapeutics, 89(2), 183–188. doi: 10.1038/clpt.2010.286

Kanavos, P., & Sullivan, R. (2010). The role of funding and policies on innovation in cancer drug development. Ecancermedicalscience. doi: 10.3332/ecancer.2010.164

Lin, K. C., & Wai, R. J. (2021). A Feasible Fall Evaluation System via Artificial Intelligence Gesture Detection of Gait and Balance for Sub-Healthy Community- Dwelling Older Adults in Taiwan. IEEE Access, 9, 146404–146413. doi: 10.1109/ACCESS.2021.3123297

Liu, Z., Delavan, B., Roberts, R., & Tong, W. (2017). Lessons Learned from Two Decades of Anticancer Drugs. Trends in Pharmacological Sciences, 38(10), 852–872. doi: 10.1016/j.tips.2017.06.005

Lo, A. W., & Thakor, R. T. (2022). Financing Biomedical Innovation. Annual Review of Financial Economics, 14(1), 231–270. doi: 10.1146/annurev-financial-031721-081537

Ma, Y., Ping, K., Wu, C., Chen, L., Shi, H., & Chong, D. (2019). Artificial Intelligence powered Internet of Things and smart public service. Library Hi Tech, 38(1), 165–179. doi: 10.1108/LHT-12-2017-0274

Mahoney, R. T. (2011). Product Development Partnerships: Case studies of a new mechanism for health technology innovation. Health Research Policy and Systems, 9(1), 33. doi: 10.1186/1478-4505-9-33

Mali, A. D. (2019). Recent Domain-Specific Applications of Artificial Intelligence Using IoT. International Journal on Artificial Intelligence Tools, 28(07), 1930003. doi: 10.1142/S0218213019300035

Moghbelan, Y., Esposito, A., Zyrianoff, I., Spaletta, G., Borgo, S., Masolo, C., Ballarin, F., Seidita, V., Toni, R., Barbaro, F., Di Conza, G., Quartulli, F.

P., & Di Felice, M. (2024). A Smart Motor Rehabilitation System Based on the Internet of Things and Humanoid Robotics. Applied Sciences (Switzerland), 14(24). doi: 10.3390/app142411489

Nadian-Ghomsheh, A., Farahani, B., & Kavian, M. (2021). A hierarchical privacy-preserving IoT architecture for vision-based hand rehabilitation assessment. Multimedia Tools and Applications, 80(20), 31357–31380. doi: 10.1007/s11042-021-10563-2

Oniani, S., Marques, G., Barnovi, S., Pires, I. M., & Bhoi, A. K. (2021). Artificial Intelligence for Internet of Things and Enhanced Medical Systems (pp. 43–59). doi: 10.1007/978-981-15-5495-7_3

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ, n160. doi: 10.1136/bmj.n160

Petrovic, M., Vukicevic, A. M., Djapan, M., Peulic, A., Jovicic, M., Mijailovic, N., Milovanovic, P., Grajic, M., Savkovic, M., Caiazzo, C., Isailovic, V., Macuzic, I., & Jovanovic, K. (2022). Experimental Analysis of Handcart Pushing and Pulling Safety in an Industrial Environment by Using IoT Force and EMG Sensors: Relationship with Operators' Psychological Status and Pain Syndromes. Sensors, 22(19). doi: 10.3390/s22197467

Qayyum, S., Sultana, A., Heyat, M. B. Bin, Rahman, K., Akhtar, F., Haq, A. ul, Alkhamis, B. A., Alqahtani, M. A., & Gahtani, R. M. (2023). Therapeutic Efficacy of a Formulation Prepared with Linum usitatissimum L., Plantago ovata Forssk., and Honey on Uncomplicated Pelvic Inflammatory Disease Analysed with Machine Learning Techniques. Pharmaceutics, 15(2). doi: 10.3390/pharmaceutics15020643

Rahman, S. M., Ali, M. A., & Al Mamun, M. A. (2021). The Use of Wearable Sensors for the Classification of Electromyographic Signal Patterns based on Changes in the Elbow Joint Angle. Procedia Computer Science, 185, 338–344. doi: 10.1016/j.procs.2021.05.043

S Mahalakshmi. (2019). Artificial Intelligence with the Internet of Things on Healthcare systems: A Survey. International Journal of Advanced Trends in Computer Science and Engineering, 8(6), 2847–2854. doi: 10.30534/ijatcse/2019/27862019

Shefa, F. R., Sifat, F. H., Uddin, J., Ahmad, Z., Kim, J. M., & Kibria, M. G. (2024). Deep Learning and IoT-Based Ankle–Foot Orthosis for Enhanced Gait Optimization. Healthcare (Switzerland), 12(22). doi: 10.3390/healthcare12222273

Woolf, C. J., & Max, M. B. (2001). Mechanism-based Pain Diagnosis. Anesthesiology, 95(1), 241–249. doi: 10.1097/00000542-200107000-00034

Yao, Y. C., Wu, H. Te, Shu, L. F., & Lu, C. Y. (2022). Developing a Multifunctional Heating Pad Based on Fuzzy-Edge Computations and IoMT Approach. Journal of Internet Technology, 23(7), 1519–1525. doi: 10.53106/160792642022122307007

Zhang, Q., Jin, T., Cai, J., Xu, L., He, T., Wang, T., Tian, Y., Li, L., Peng, Y., & Lee, C. (2022). Wearable Triboelectric Sensors Enabled Gait Analysis and Waist Motion Capture for IoT-Based Smart Healthcare Applications. Advanced Science, 9(4). doi: 10.1002/advs.202103694

Modeling and Construction of Circular Economy Digital Twins Based on System Dynamics

Lea Masnec, Vjeran Strahonja
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{lmasnec, vjeran.strahonja}@foi.unizg.hr

Abstract. *This paper examines system dynamics (SD) in the context of building digital twins (DTs), particularly within circular economy systems. It begins with an overview of SD. The strengths and limitations of SD are analysed in comparison with complementary simulation methods, including Agent-Based Modeling (ABM), Petri Nets, and Discrete Event Simulation (DES), to identify potential for integration. The analysis focuses on how these methods can address specific gaps in SD, such as individual-level behaviour, discrete events, and operational workflows and enhance the accuracy, detail, and flexibility of DT development. The paper explores methodological approach for combining SD with complementary approaches and highlights the importance of selecting modeling techniques based on system characteristics and modeling goals. The feasibility of implementing such models is further examined through the analysis of relevant modeling tools, with particular attention to hybrid platforms like AnyLogic that support SD, ABM, DES, and real-time data integration in DT applications. The paper discusses future directions for simulation-based DT development in circular economy contexts, emphasizing the potential for multi-method modeling to support strategic and operational decision-making.*

Keywords. system dynamics, agent-based modeling, digital twin, petri nets, discrete-event simulation

1 Introduction

In modern organizations, value chains and business ecosystems there is a strong emphasis on increasing agility and efficiency, optimizing operations, reducing disruptions and other risks (Sabri & Sahri, 2010). This requires real-time analytical and predictive capabilities. Increasingly and in various problem areas, the solution is sought in building a DT of the system being managed (Akhtar et al., 2022). A DT is a simulation model of a physical system that reflects its function, behavior, or other properties of interest with sufficient accuracy, and is coupled to that system in such a way that it receives data about the states and changes of its original in real time. A physical system and its corresponding DT as a cyber representation form a cyber-physical system (CPS) (Januario et al., 2019).

The choice of the method and technique of simulation modelling, which is the basis for building a DT, is critical in its construction. The focus of this paper is on the research of properties of SD in the development of a simulation model, and later a DT in the field of circular economy. SD is a method used to understand the behavior of complex systems over time. Its core strength lies in its ability to model interactions among different system components and analyze how changes in one part of the system affect others (Ding et al., 2018; Gejo-García et al., 2022). SD can be limited in predicting unexpected behaviors and external influences that can significantly affect the long-term behavior of a system. Because of these limitations, the combination of SD with other modeling methods is becoming increasingly popular (Sterman, 2018). For example, ABM allows for more detailed simulations of individual behaviors within a system (Ding et al., 2018), while DTs provide real-time data and updates (Liu et al, 2021) that can improve the accuracy of the model.

The aim of the research, the results of which are presented in this paper, is to analyze the suitability of SD, independently and in combination with other methods, for simulation modeling and building DTs of physical circular economy systems.

The circular economy, which is based on the principles of waste reduction, reuse and recycling, requires models to analyze the flows of materials and energy through the system (Chenet al., 2021). Modeling circular economy processes faces challenges, like nonlinear flows of materials and energy, uncertainties in supply chains, heterogeneity of actors, and the need to integrate short-term operational with long-term strategic decisions. Precisely because of these challenges, it is important to combine multiple modeling approaches. SD is suitable for this purpose because of its ability to model complex interrelationships and feedback loops. Optimizing resource use is essential to achieving the goals of the circular economy. SD can help analyze the

effectiveness of different resource management strategies, such as reducing energy or water consumption, and assess their long-term effects on the system. This method allows testing different scenarios to identify the most effective measures to conserve resources (Akhtar et al., 2022).

The paper is structured into five chapters. Chapter 2 discusses SD and the possibility of its complementary application with other modeling methods. SD has been a subject of scientific research for decades, as well as the similarities, differences, overlaps, complements and possibilities of integration of different methods, such as ABM, Petri Nets and DES. Advantages of separate application of SD and other methods, and to use the complementarity of methods and synergistic effects on accuracy, usability, reliability and other properties of the simulation model and DT in an integrated environment (Choi et al., 2023; Ding et al., 2018; Garcia, 2020). Building a DT based on a SD and complementary methods, including the methodological framework and decision-making in the requirements definition and system design phase, is the subject of Chapter 3. After comparative analysis, followed the process of selecting methods complementary to SD in simulation modeling and possibly creating a DT of the physical system in the circular economy. Chapter 4 analyzes the possibility of applying tools for the development of DTs based on SD and complementary methods. Conclusion and future research summarize the analysis and highlight future directions for the study of DTs based on SD and complementary methods.

2 System dynamics and complementary modeling methods

As SD is a system modeling method based on stocks and flows, feedback loops (positive and negative), time delays, as well as variables and parameters that define relationships and influence system behavior over time (Bala et al., 2017; Garcia, 2020). SD is more of a set of methods and tools than a single method (Bala et al., 2017). It involves creating models that represent systems as a set of interconnected components, using simulations to predict the behavior of the system over time, and analyzing the feedback loops that govern the dynamics of the system. Data for SD models usually come from actual measurements, historical data, expert assessments, and scientific research. The process involves identifying key variables and the relationships between them and quantifying them in a way that allows for simulation and analysis (Bala et al., 2017).

SD is used in modeling and simulation of short-term and long-term dynamics. It helps to understand how changes in a system act in the short and long term, including identifying causes and effects that may not manifest immediately (Garcia, 2020). It is suitable for modelling linear and nonlinear relationships within a system, which is crucial for understanding complex systems (Garcia, 2020). In addition to predictive analytics at the operational and tactical levels, SD is used to analyze and design policies and strategies for managing loosely coupled systems, including value chains and ecosystems in the circular economy, which is a particular subject of interest in this paper (Malbon & Parkhurst, 2023).

SD is chosen as the primary modeling approach in this article due to its unique ability to capture the structural foundations and long-term behavior of complex systems such as those found in the circular economy. Unlike other modeling paradigms that focus primarily on events or individual entities, SD emphasizes systemic structure, causal relationships, and feedback loops that drive system evolution over time. This makes it particularly suitable for identifying leverage points, simulating policy interventions, and understanding the accumulation and depletion of resources across production and recovery cycles. Moreover, its capacity to integrate both qualitative insights and quantitative data allows for flexible and transparent model construction, which is essential in domains where uncertainty, delays, and indirect effects are common (Bala et al, 2017). By using SD as the foundation, the model ensures a coherent, system-wide representation upon which more detailed methods like ABM and DES can be layered when needed.

The foundation for building a DT of a physical system is the creation of its simulation model (Biller et. al., 2023). This chapter presents a comparison of SD with some other simulation modeling methods, to fill in the gaps that are missing from SD. The applicability and choice of methods for simulation modeling in general, and in the circular economy, depend on fundamental properties and expected behaviors of the physical system, such as predominantly static versus dynamic behavior; deterministic versus stochastic behavior; linearity versus nonlinearity; discrete versus continuous state changes; and the presence or absence of feedback. It should also be considered whether the objective of modeling is to understand and analyze the physical system, optimize individual aspects, predict behavior, or construct a DT.

ABM is a simulation method used to model the interactions of autonomous agents which can represent individuals, organizations, or entities, within an environment (Khazaii, 2016). It is based on bottom-up modeling, where agents follow defined rules and make decisions based on local information, often leading to complex, emergent phenomena (Khazaii, 2016).. It captures the evolving state of the system over time, which is useful for studying how complex behaviors emerge from simple agent interactions, making it useful for modeling systems with non-linear dynamics. It is flexible for a wide variety of different domains, scenarios, from human behaviors social dynamics, and decision-making processes in areas like economics, sociology, and politics, through ecological modeling of interactions between species, resource usage, and

environmental changes, to simulating complex circular economy and supply chain behaviors and interactions of agents. In all these applications, ABM focuses on individual agents, which allows for granular analysis of specific behaviors, interactions and outcomes (Khazaii, 2016).

There are problems for which ABM is not a good solution for simulation modelling:

- problems that are highly deterministic or where interactions are simple and predictable
- detailed information about individual agents or their behaviors is unavailable
- optimizing processes at the system level, especially when interactions are not as important as high-level parameters
- problems requiring high-volume data processing or real-time decision-making when modeling large numbers of agents or complex environments
- change in agent behavior, which would require to calibrate the models accurately, especially when there is limited data on agent behavior.
- overfitting and to specific agents leading to reduced generalizability.

ABM is useful in many areas, and research touches on healthcare systems, smart cities, manufacturing, and the like. Research (Ding et al., 2018) presented a concrete example of SD implementation in the Vensim tool and filled in the gaps that SD cannot fill with ABM. In the area of complex systems, Cassidy et al. conducted a systematic review of the application of SD and ABM models (Cassidy et al., 2019). Their research shows that SD can help in understanding complex interaction. Agent-based models provide insight into the behavior of individual entities within the system, which can be complementary to the SD approach. SD and ABM can be successfully used complementary in simulation modeling. SD is used to study macro-level system behavior such as the movement of resources or quantities in the system over time. ABM models micro-level system behavior through the behavior of individual agents and their interactions. Examples are people, their decisions and their interactions (Cassidy et al., 2019). The combination of these methods allows for a detailed understanding of micro-dynamics and its impact on macro-dynamics. ABM is a bottom-up approach, simulating changes in the state of individual agents in the system, which can be aggregated in SD. The combination of SD and AB can reduce computational time, provide a strategic overview characteristic of SD, while at the same time capturing relevant elements of individual heterogeneity and stochasticity of agents and interactions in processes (Guerrero et al., 2016).

Petri Nets are a modeling method with a solid mathematical foundation and graphic representation used to represent and analyze discrete event systems, particularly for systems with concurrent, asynchronous, or parallel processes (Dai et al., 2020). Digital twins driving model based on Petri net in industrial pipeline. They consist of places representing conditions or resources, transitions representing events or activities, tokens that reside in places and move between places via transitions, representing the state of the system (Dai et al., 2020). Petri nets are widely used for modeling and simulating dynamic systems, with a strong focus on the flow of information or resources and how they evolve over time. Petri nets are particularly effective at modeling systems with multiple concurrent processes and synchronization constraints. They are used for modeling various types of systems, such as time or resource-dependent processes in manufacturing, resource management and scheduling, assembly and other workflow systems with multiple interacting components. In the field of communications, they are used for modeling communication protocols, distributed systems, and network flows, where concurrency and synchronization are important, including concurrent and parallel processing (Dai et al., 2020). Areas and problems where Petri nets modeling is not a good choice are:

- modeling continuous or analog systems, such as fluid dynamics or electrical circuits, where differential equations are used,
- very complex or large systems where Petri nets are very large, computationally expensive and difficult to analyze or simulate, leading to state-space explosion problems,
- data-intensive systems that require detailed and complex data modeling, because Petri nets focus on control flow and resource allocation, but they do not directly model detailed data structures or system-level behaviors,
- systems where decision-making is complex and strategic, as they focus more on process flow and less on high-level cognitive behavior.

Petri Nets and **SD** can complement each other effectively in simulation modeling (Duggan, 2006; Viswanadham & Srinivasa Raghavan, 2000). Petri Nets are excellent for capturing the detailed, event-driven behaviors of specific processes (Viswanadham, & Srinivasa Raghavan, 2000), such as material flows, recycling processes, or product disassembly, where concurrency, resource allocation, and synchronization are important. SD, on the other hand, excels at modeling feedback loops, stock-flow relationships, and long-term behavior, such as the impact of recycling rates on resource stocks or the effects of consumer demand on product life cycles. By combining Petri Nets with SD, detailed process flows and interactions which are the result of events that occur and changes in the state at certain points in time can be captured alongside the broader, system-level dynamics. All of the above is applicable to the circular economy because it enables both operational insights and strategic, policy-driven decision-making (Duggan, 2006).

DES modeling method is used to represent systems where changes occur at discrete points in time due to specific events. In DES, the system's state is updated only when an event occurs, and events trigger changes in the system (Ragazzini et al. 2024). DES is particularly useful for modeling and simulations of the behavior of systems that are driven by discrete events, providing a detailed and realistic simulation of how systems operate over time. It can be used to model a wide range of systems, from simple queuing systems to complex manufacturing or healthcare systems and transportation networks (Ragazzini et al. 2024). DES can help in identifying bottlenecks, delays and inefficiencies in system performance and resource utilization. in a way that allows for experimentation with different system configurations, resource allocations, and process flows to test various scenarios and outcomes (Ragazzini et al. 2024). DES has limitations in modeling and simulations:

- useful only for systems where discrete events happen at distinct points in time and not for continuous or analog systems where time is a continuous variable
- can be time-consuming and resource-demanding, particularly for complex systems when accurate modeling of many events and interactions is required,
- In large or complex systems, the number of possible states and events explodes exponentially, making it difficult to manage, analyze, or interpret the results.
- may not be effective in environments requiring real-time decision-making, such as some emergency systems, where fast and dynamic adaptation to changes is crucial.

DES and SD can be complementary in simulation modeling, especially in the context of the circular economy, by combining their strengths (Morgan, 2011). While SD is excellent for modeling large-scale, long-term, and continuous feedback loops in systems like resource flows, waste recycling, and stock accumulation, DES can be used to model specific, time-dependent events like the processing of materials, product repairs, or resource transactions that occur at discrete points in time (Reed et al., 2021). When combined, SD can capture the overall dynamics and feedback mechanisms within a circular economy, while DES can model the detailed, event-driven activities (such as product disassembly or resource recovery) that drive those dynamics. Together, they provide a more comprehensive view, combining the holistic, feedback-rich system modeling of SD with the event-based, process-level detail of DES, making them effective in analyzing systems that require both high-level strategic insights and detailed operational efficiency (Jovanovski et al., 2012). Combining system dynamics and discrete event simulations-Overview of hybrid simulation models. *Journal of Applied Engineering Science*, *10*(3). .

Several modeling methods can complement SD in circular economy contexts. **DTs** support real-time model validation and system monitoring, though they require substantial data and resources. **ABM** enhances SD by simulating individual behaviors and interactions, offering detailed system insights, but is complex to define and implement. **Petri Nets** are effective for modeling discrete events and system synchronization, though limited to event-driven systems. **DES** excels at modeling time-specific processes like material flows and repairs, but is less suited for continuous dynamics. Each method addresses specific system characteristics, enabling a more comprehensive simulation when combined with SD.
The method selection matrix is presented in Table 1.

Table 1. Method selection matrix for combination with SD

Method	System-level understanding and analysis	Implementation complexity	Application in circular economy	Real-time analysis	Big data integration	Process management and optimization
DT	+ (Liu et al, 2021)	–	+ (Rocca, et al., 2005)	+ (Choi et al., 2023)	+ (Stojanovic et al, 2021)	+ (Liu et al, 2021)
ABM	+ (Guerrero et al., 2016)	+ (Ding et al., 2018)	+ (Lange et al., 2021)	–	–	+ (Ding et al., 2018)
Petri Nets	+ (Viswanadham, & Srinivasa Raghavan, 2000)	–	+ (Viswanadham, & Srinivasa Raghavan, 2000)	–	–	+ (Viswanadham, & Srinivasa Raghavan, 2000)
DES	+ (Ganguli & Adhikari, 2020)	+ (Ganguli & Adhikari, 2020)	+ (Charnley et al., 2019)	–	–	+ (Ganguli & Adhikari, 2020)

3 Building a DT based on a SD and complementary methods

DTs enable the creation of virtual replicas of physical systems that are continuously updated with real-world data. The combination of DTs and SD enables the modeling and simulation of phycsical system in near real-time. They help to identify and analyze changes in the system in real time, which can improve system prediction and management (Ganguli & Adhikari, 2020). According to Stojanovic et al. (2021) the modern approach to DTs considers them as software entities that follow the entire life cycle of their physical counterparts, including the engineering, operations, and disposal phases. A DT becomes a replica of reality by implementing the real world as closely as possible and calibrating it with data. It was originally proposed

as a concept to support decision-making in the product design and manufacturing, but is used as an analysis tool throughout the entire life cycle. (Choi et al., 2023).

Regarding the application of SD as a methodological foundation of DTns, Gejo-García et al. applied the concept of DT to manufacturing systems to assess the long-term effects of different manufacturing strategies. Their research shows how SD can simulate and optimize manufacturing processes by considering the dynamic interactions between different components of the system (Gejo-García et al., 2022). A similar approach is used by Choi et al. in simulating naval ship operations using DTs, which shows how SD can be used to optimize operational strategies and improve ship performance (Choi et al., 2023). Both studies explore the importance of SD in creating DTs for the simulation and analysis of complex systems. In the context of the importance of DTs, it is also discussed in Ganguli & Adhikari article (2020) regarding the development of the framework of the virtual model of the physical system in which the DT plays a role, but it is still necessary to create and validate the simulation model, which is from the field of SD.

SD enables the simulation of the wide world with high fidelity, and the DT is not a term referring to a specific technology, but an idea that simulates the fusion of many existing state-of-the-art technologies. Researches on combining DTs based on SD with other methods, such as Gejo-García et al., (2022) for manufacturing systems, make it possible to draw conclusions about why combining these methods has advantages. DTs based on SD simulations provide a platform for CPS in which the DT well reflects the state of the physical system in real time. By enabling simulations of states and changes in states in the near future, the DT can serve:

1. Process management and real-time optimization, and indirectly improving flexibility and resilience in unexpected conditions Goodwin et al., 2024).
2. Decision support based on data and predictions, and indirectly proactive identification of disruptions within physical systems, optimizing resource utilization, increasing efficiency, etc. (Vetrivel et al. 2024).
3. Continuous Optimization and Lifecycle Management by leveraging real-time data and simulations, and indirectly improving planning capabilities, process stability, and resource optimization (Zhou, 2024).
4. More comprehensive approach to model building that includes different levels of models, different relevant areas, flows, and decision options within the domains of single entity, value chains or ecosystems (Biller et. al., 2023).

The methodological framework for the development of DTs necessarily unites the theoretical structure, approaches and methods of two areas: Simulation modeling methodology and Model Technology - DTs are executable models that embody the features, behaviors, and properties of physical systems. To construct such executable, efficient, reusable, and interoperable models, an appropriate modeling technology is essential (Liu et al., 2019); Systems Engineering methodology, as used for the development of SCADA (Supervisory Control and Data Acquisition) systems and software for automatic process control.

4 Practical aspects of developing DT based on SD

This chapter focuses on the practical aspects of developing DTs based on SD and complementary modeling methods. It outlines the most relevant modeling tools and discusses the structural elements and application considerations essential for creating effective SD-based DTs in circular economy contexts.

4.1 Tools for modeling and simulation

Various software tools support the implementation of SD and its complementary methods in the development of DTs. Vensim, Stella, and AnyLogic are commonly used for SD modeling, based on concepts such as stocks, flows, feedback loops, and time delays (Sapiri et al., 2017). AnyLogic, in particular, enables hybrid modeling, combining SD, ABM, and DT functionality, including real-time data validation (Farhan et al, 2020; Chakraborti et al., 2020). Tools such as NetLogo and Repast specialize in ABM, supporting simulation of heterogeneous agents with adaptive behaviors and local interactions, which is especially useful for studying micro-level dynamics and emergent phenomena (Robertson, 2005;Antelmi et al., 2022). Petri nets are implemented through tools like CPN Tools and PIPE, providing precise modeling of discrete transitions, parallel processes, and resource flows within systems characterized by concurrency and synchronization (Dworzański & Lomazova, 2013; Llado, 2022). In the context of DES, tools such as Simul8, Arena, or even AnyLogic support modeling of event-driven systems where state changes occur at specific points in time. DES is particularly effective for analyzing queueing systems, resource scheduling, and operational workflows, offering time-based precision and performance optimization capabilities (Ganguli & Adhikari, 2020).

Among the available tools, AnyLogic stands out as the most versatile and comprehensive platform for developing DTs based on SD and complementary methods. Its support for hybrid modeling, which integrates SD, ABM, and DES within a single environment, allows researchers and practitioners to simulate systems at multiple abstraction levels — from

aggregate feedback-driven dynamics to individual agent behavior and discrete event flows (Farhan et al, 2020; Chakraborti et al., 2020). Moreover, AnyLogic supports real-time data integration, enabling dynamic model calibration and validation, which is essential for implementing operational DTs.

4.2 Conceptual and system architecture for digital twin development

Fig. 1 illustrates the logical relationship between SD and other modeling methods within a DT development context.

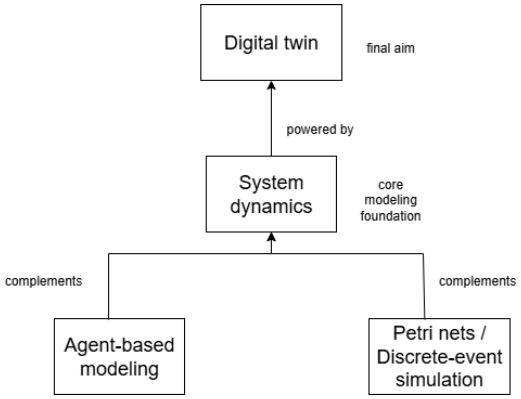


Figure 1. Conceptual modeling structure of a DT based on SD and complementary methods

SD serves as the core modeling foundation for the development of DTs, providing insight into long-term system behavior through feedback loops, stock-flow relationships, and continuous dynamics. SD has limitations when it comes to representing discrete events or individual-level behavior. To address these gaps, complementary methods are integrated: ABM enables simulation of autonomous agents and their interactions, capturing emergent and heterogeneous behaviors at the micro level; while Petri Nets and DES allow modeling of discrete, time-specific events, resource flows, and synchronization in operational processes. Petri Nets and DES are grouped together due to their shared focus on event-driven systems and process-level precision. Together, these methods complement SD by adding granularity, stochasticity, and temporal accuracy, enabling the development of DTs that are both systemic and detailed, strategic and operational, predictive and responsive.

After establishing the conceptual modeling logic in Fig. 1, which highlights the integration of SD with complementary methods such as ABM and DES, Fig. 2 presents a layered architecture that shows how such a model can be implemented within a CPS. It connects the physical environment to the modeling logic through communication infrastructure and real-time data integration, ultimately supporting applications such as monitoring, control, and decision-making.

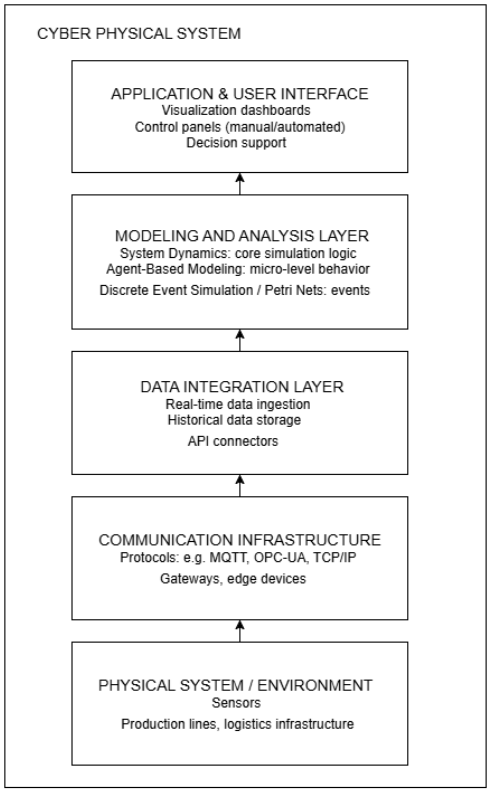


Figure 2. Layered architecture of a CPS for DT development (modified from Januario et al., 2019)

Architecture (Fig. 2) illustrates how DTs based on SD, ABM, and DES are operationalized. It spans from physical infrastructure and sensors, through communication and data integration layers, up to the simulation and application interface layers. This structure can enable real-time simulation, monitoring, and decision support in complex systems.

4 Conclusion and future research

This research analyzes SD and the need to complement it with other simulation modeling methods, such as ABM, and Petri Nets, especially in building DTs. Although SD has many advantages, such as the ability to model interactions within complex systems and analyze feedback loops, its shortcomings include limited ability to predict unexpected behaviors and external influences. By combining SD with other methods, these shortcomings can be compensated for and the accuracy and reliability of modeling can be improved, as demonstrated in this paper.

The research proposes the application of a methodological approach in building simulation models and DTs. Layered architecture is introduced to integrate various modeling methods with data infrastructure and application layers, enabling simulation models to evolve into real-time, decision-supporting DTs. This approach is particularly suited to

circular economy systems, where feedback, flow dynamics, and emergent behavior all play central roles.

Connecting different modeling paradigms like system dynamics, agent-based modeling, and discrete-event simulation enhances the representational fidelity of circular systems. This integration allows simultaneous representation of long-term trends, operational process flows, and individual behavior dynamics. In circular economy applications, such as product lifecycle extension, waste reduction, and material recovery, this multiscale perspective is essential for designing more resilient and resource-efficient systems.

Future research will focus on operationalizing the proposed conceptual framework within the textile industry. Using the architectural foundation presented in this work, a DT will be developed for a production-recycling flow with the objective of reducing generated waste. The model will incorporate real-time data streams, dynamic stock-flow structures, and event-based control, enabling scenario analysis and optimization of material reuse and resource efficiency strategies in line with circular economy goals.

References

Akhtar, P., Ghouri, A. M., Saha, M., Khan, M. R., Shamim, S., & Nallaluthan, K. (2022). Industrial digitization, the use of real-time information, and operational agility: Digital and information perspectives for supply chain resilience. IEEE Transactions on Engineering Management. doi: http://dx.doi.org/10.1109/TEM.2022.3182479

Antelmi, A., Cordasco, G., D'Ambrosio, G., De Vinco, D., & Spagnuolo, C. (2022). Experimenting with agent-based model simulation tools. *Applied Sciences, 13*(1), 13. doi: https://doi.org/10.3390/app13010013

Bala, B. K., Arshad, F. M., & Noh, K. M. (2017). *SD*. In *Springer Texts in Business and Economics*. Springer Singapore. doi: https://doi.org/10.1007/978-981-10-2045-2

Biller, S., Venditti, P., Yi, J., Jiang, X., & Biller, B. (2023). Simulation-driven Digital Twins: the DNA of Resilient Supply Chains. In *2023 Winter Simulation Conference (WSC)* (pp. 1463-1477). IEEE

Cassidy, R., Singh, N. S., Schiratti, P. R., Semwanga, A., Binyaruka, P., Sachingongu, N., Blanchet, K. (2019). Mathematical modelling for health systems research: a systematic review of SD and agent-based models. *BMC health services research*, *19*(1), 845. doi: https://doi.org/10.1186/s12913-019-4627-7

Chakraborti, A., Heininen, A., Koskinen, K. T., & Lämsä, V. (2020). Digital twin: Multi-dimensional model reduction method for performance optimization of the virtual entity. *Procedia CIRP, 93*, 240–245.

Charnley, F., Tiwari, D., Hutabarat, W., Moreno, M., Okorie, O., & Tiwari, A. (2019). Simulation to enable a data-driven circular economy. *Sustainability*, 11(12), 3379. doi: https://doi.org/10.3390/su11123379

Choi, J., Moon, S., & Min, S. (2023). Digital twin simulation modeling process with SD: An application to naval ship operation. *International Journal of Robust and Nonlinear Control, 33*(16), 10136–10150. doi: https://doi.org/10.1002/rnc.6893

Dai, Y., Shi, Y., Zhang, Z., Tao, R., & Fang, F. (2020). Digital twins driving model based on Petri net in industrial pipeline. In *2020 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)* (pp. 283-291). IEEE.

Ding, Z., Gong, W., Li, S., & Wu, Z. (2018). SD versus agent-based modeling: A review of complexity simulation in construction waste management. *Sustainability, 10*(7), 2484. doi: https://doi.org/10.3390/su10072484

Duggan, J. (2006). A comparison of Petri net and SD approaches for modelling dynamic feedback systems. *Proceedings of the 24th International Conference of the SD Society*.

Dworzański, L. W., & Lomazova, I. A. (2013). CPN tools-assisted simulation and verification of nested Petri nets. *Automation and Remote Control, 47*(7), 393–402. doi: https://doi.org/10.3103/S0146411613070201

Farhan, M., Gohre, B., & Junprung, E. (2020). Reinforcement learning in AnyLogic simulation models: A guiding example using Pathmind. In *2020 Winter Simulation Conference (WSC)* (pp. 3212–3223). IEEE. doi: https://doi.org/10.1109/WSC48552.2020.9383916

Ganguli, R., & Adhikari, S. (2020). The digital twin of discrete dynamic systems: Initial approaches and future challenges. *Applied Mathematical Modelling, 77*, 1110–1128.

Garcia, J. M. (2020). *Theory and practical exercises of SD*. Amazon Fulfillment.

Gejo-García, J., Reschke, J., Gallego-García, S., & García-García, M. (2022). Development of a SD simulation for assessing manufacturing systems based on the digital twin concept. *Applied Sciences, 12*(4), 2095. doi: https://doi.org/10.3390/app12042095

Goodwin, T., Xu, J., Celik, N., & Chen, C. H. (2024). Real-time digital twin-based optimization with

predictive simulation learning. *Journal of Simulation*, *18*(1), 47-64.

Guerrero, C. N., Schwarz, P., & Slinger, J. H. (2016). A recent overview of the integration of SD and agent-based modelling and simulation. In *SD Conference* (pp. 1–13).

Januario, F., Cardoso, A., & Gil, P. (2019). A distributed multi-agent framework for resilience enhancement in cyber-physical systems. *IEEE Access*, *7*, 1–1. doi: https://doi.org/10.1109/ACCESS.2019.2903629

Jovanovski, B., Minovski, R., Voessner, S., & Lichtenegger, G. (2012). Combining system dynamics and discrete event simulations-Overview of hybrid simulation models. *Journal of Applied Engineering Science*, *10*(3).

Khazaii, J. (2016). Agent-based modeling. In *Advanced Decision Making for HVAC Engineers: Creating Energy Efficient Smart Buildings* (pp. 137-144). Cham: Springer International Publishing.

Lange, K. P. H., Korevaar, G., Oskam, I. F., Nikolic, I., & Herder, P. M. (2021). Agent-based modelling and simulation for circular business model experimentation. *Resources, Conservation & Recycling Advances, 12*, 200055. doi: https://doi.org/10.1016/j.rcradv.2021.200055

Liu, M., Fang, S., Dong, H., & Xu, C. (2021). Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems, 58*, 346–361. doi: https://doi.org/10.1016/j.jmsy.2020.06.017

Llado, C. M. (2022). PIPE 2.7 overview: A Petri net tool for performance modeling and evaluation. *SIGMETRICS Performance Evaluation Review, 49*(4), 76–80. doi: https://doi.org/10.1145/3543146.3543163

Malbon, E., & Parkhurst, J. (2023). SD modelling and the use of evidence to inform policymaking. *Policy Studies, 44*(4), 454–472. doi: https://doi.org/10.1080/01442872.2022.2080814

Morgan, J., Howick, S., & Belton, V. (2011, December). Designs for the complementary use of system dynamics and discrete-event simulation. In *Proceedings of the 2011 Winter Simulation Conference (WSC)* (pp. 2710-2722). IEEE.

Ragazzini, L., McGinnis, L. F., Negri, E., & Macchi, M. (2024). Modeling Operational Control in Discrete-Event Logistics Systems and Their Digital Twins. In *2024 Winter Simulation Conference (WSC)* (pp. 1-12). IEEE.

Reed, S., Löfstrand, M., & Andrews, J. (2021). Modelling cycle for simulation digital twins. *Manufacturing Letters*, *28*, 54-58.

Robertson, D. A. (2005). Book and resource reviews: Agent-based modeling toolkits NetLogo, RePast, and Swarm. *Academy of Management Learning & Education, 4*(4), 524–527. https://doi.org/10.5465/amle.2005.19086798

Rocca, R., Rosa, P., Sassanelli, C., Fumagalli, L., & Terzi, S. (2020). Integrating virtual reality and digital twin in circular economy practices: A laboratory application case. *Sustainability, 12*(6), 2286. doi: https://doi.org/10.3390/su12062286

Sabri, E. H., & Shaikh, S. N. (2010). Lean and agile value chain management: a guide to the next level of improvement. J. Ross Publishing.

Sapiri, H., Zulkepli Hew, J., Ahmad, N., Zainal Abidin, N., & Hawari, N. N. (2017). *Introduction to system dynamic modelling and Vensim software*. UUM Press. doi: https://doi.org/10.32890/9789672064084

Sterman, J. (2018). SD at sixty: The path forward. *SD Review*, 34(1–2), 5–47. doi: https://doi.org/10.1002/sdr.1601

Stojanovic, L., Usländer, T., Volz, F., Weißenbacher, C., Müller, J., Jacoby, M., & Bischoff, T. (2021). Methodology and tools for digital twin management—The FA3ST approach. *IoT*, *2*(4), 717-740.

Chen, X., Memon, H. A., Wang, Y., Marriam, I., & Tebyetekerwa, M. (2021). Circular economy and sustainability of the clothing and textile industry. *Materials Circular Economy*, *3*(1), 12. doi: 10.1007/s42824-021-00026-2

Vetrivel, S. C., Sowmiya, K. C., & Sabareeshwari, V. (2024). Digital twins: Revolutionizing business in the age of AI. In *Harnessing AI and digital twin technologies in businesses* (pp. 111-131). IGI Global.

Viswanadham, N., & Srinivasa Raghavan, N. R. (2000). Performance analysis and design of supply chains: A Petri net approach. *Journal of the Operational Research Society, 51*(10), 1158–1169. doi: https://doi.org/10.1057/palgrave.jors.2600063

Zhou, E. (2024). Data-driven Simulation Optimization in the Age of Digital Twins. In *Proceedings of the 38th ACM SIGSIM Conference on Principles of Advanced Discrete Simulation* (pp. 2-2).

Data and Knowledge Bases

Ana Petrovic, Danijela Jaksic, Kristian Stancin
The Role of Ontologies and Semantic Technologies in Developing Educational Chatbots

Mikel Ferrer-Oliva, José Amelio Medina-Merodio, José Javier Martínez-Herraiz, Carlos Cillerue-lo-Rodríguez
Relational Modelling of Cyberattacks: A Taxonomy for the Analysis of Threat Interactions in Digital Environments

Sven Vujević, Maja Cerjan, Kornelije Rabuzin
Data Quality in Distributed Information Systems: Conceptual Challenges and Empirical Findings

Martina Asenbrener Katic, Sanja Candrlic, Marina Rauker Koch
A Case Study on the Two-Sided Version of the Node of Knowledge Conceptual Framework

Nađa Klještanović, Sonja Ristić, Darko Stefanović
Parallel Utilization of SQL and NoSQL Databases –A Systematic Literature Review

Dajana Marnika, Maja Gligora Marković, Božidar Kovačić
Unlocking Data Quality for Responsible Artificial Intelligence

Ana Pasarić, Patrizia Poščić, Danijela Jakšić
Building Trustworthy Knowledge: Data Quality and Validation Protocols for Educational Knowledge Graphs

The Role of Ontologies and Semantic Technologies in Developing Educational Chatbots

Ana Petrovic, Danijela Jaksic, Kristian Stancin
University of Rijeka, Faculty of Informatics and Digital Technologies
Radmile Matejcic 2
Rijeka, Croatia
{apetrovic, danijela.jaksic , kristian.stancin}@inf.uniri.hr

Abstract. This paper explores the role of ontologies and semantic technologies in the development of advanced educational chatbots and virtual assistants. It focuses on two research questions: "*What role do ontologies play in making chatbots easier to use, more accurate, and more effective for personalized learning?",* and "*What are the application domains and case studies that demonstrate the use of ontologies in educational contexts?".* The study synthesizes findings from a literature review conducted using Google Scholar and Scopus. Findings highlight the application of ontology-driven chatbots in areas such as university admissions, health and nutrition education, programming instruction, cultural heritage learning, and general e-learning environments. The results underscore the value of ontologies in enhancing chatbot performance and personalizing student support.

Keywords. ontology, chatbot, education, semantic technology

1 Introduction

The integration of ontologies and semantic technologies in the development of educational chatbots and virtual assistants represents a significant advancement in the field of educational technology. These innovations lay solid groundwork for developing smart learning systems that can engage with students, enhance their learning, and effectively respond to educational inquiries. Ontologies are structured frameworks that define the relationships between concepts within a specific domain, enabling machines to understand and interpret the meaning of terms in context. Semantic techniques involve the use of these ontologies, along with natural language processing and reasoning algorithms, to derive meaningful insights from data and support intelligent interactions. Ontologies, which outline the connections between concepts in a particular field, empower chatbots to navigate through complex educational material and guide students through organized learning resources (Lecce et al., 2010). For example, systems based on semantics have been created to streamline the arrangement of learning materials, simplifying resource management for instructors and allowing students to interactively explore content via chatbots (Lecce et al., 2010).

Additionally, the use of natural language processing (NLP) and artificial intelligence (AI) in educational chatbots allows personalized learning experiences. These systems can simulate human interaction, providing students with immediate feedback and assistance, which is crucial for motivation and cognitive development (MA et al., 2024). The potential of chatbots to serve as virtual teaching assistants is further highlighted by their ability to engage with students using evaluative language and engagement strategies, although challenges such as ideological bias and lack of contextual awareness remain (Van Poucke, 2024).

The development of educational chatbots also benefits from the integration of semantic web frameworks, which facilitate the creation of smart assistants capable of processing and modeling information dynamically. Such frameworks enable the incorporation of voice recognition and ontology-powered inference, making chatbots more effective in educational context (Sermet & Demir, 2020). As educational systems continue to embrace digital transformation, the role of ontologies and semantic technologies in enhancing the capabilities of chatbots and virtual assistants will likely expand, offering new opportunities for innovation in teaching and learning.

The paper is structured as follows: Section 2 describes the methodology, Section 3 gives literature review and answers to set research questions, Section 4 highlights main implications and gives a brief discussion, and Section 5 gives some conclusions.

2 Methodology

2.1. Research Approach

This study follows a qualitative literature review approach to examine the intersection of ontologies and chatbot development. The research process was guided by a systematic search of two academic databases, Google Scholar and Scopus, to identify relevant and recent publications focused on ontology based chatbots. These two databases were chosen for complementary reasons. Google Scholar provides broad coverage across disciplines and includes grey literature, which ensured that recent and potentially influential works outside traditional journals could be considered. Scopus, on the other hand, is a curated and highly structured database that indexes peer-reviewed journals and conference proceedings, offering more reliable metadata, filtering options, and citation analysis. Using both sources increased the comprehensiveness and balance of the literature search. The following research questions were formulated:

- **RQ1:** *What role do ontologies play in making chatbots easier to use, more accurate, and more effective for personalized learning?*
- **RQ2:** *What are the application domains and case studies that demonstrate the use of ontologies in educational contexts?*

These questions aim to capture both the functional benefits of ontologies in chatbot design and the diversity of their practical applications within educational settings.

2.2. Google Scholar

The initial search was conducted on Google Scholar using the query "Ontology chatbot" restricted to article titles and limited to the publication years from 2021 to 2025. The Google Scholar search was intentionally kept simple due to the platform's limited support for complex queries and filtering. This allowed us to capture a broad initial pool of results, even if further manual screening was required.

This search gave a total of 24 results. A first-level screening based on the titles and abstracts was performed to determine relevance and accessibility. The filtering process resulted in: 1 article excluded due to being in a non-English language, 15 articles excluded due to lack of access or availability, and 8 articles deemed potentially relevant for full-text reading.

Following a comprehensive reading of these 8 articles, 7 were selected as relevant and included in the literature review.

2.3. Scopus

Compared to Google Scholar, Scopus allows for advanced query construction with precise filters for publication year, subject area, document type, and language. This motivated the use of a more refined and structured query in Scopus to ensure that only peer-reviewed, computer science–oriented, and accessible works were included. This strategy reduced noise in the results and increased the methodological rigor of the review. A more refined and structured query was applied for the Scopus database, defined as follows:

(TITLE-ABS-KEY (chatbots) AND TITLE-ABS-KEY (ontology)) AND PUBYEAR > 2020 AND PUBYEAR < 2026 AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (LANGUAGE , "English"))

This query aimed to retrieve peer-reviewed conference papers and journal articles published between 2021 and 2025, in the subject area of computer science, written in English, and with open access availability. The search returned 28 results.

Title and abstract screening led to the following categorization: 9 articles selected for full-text review and 19 excluded (including 4 duplicates).

Out of the 9 fully reviewed texts, 7 articles were ultimately included in the analysis.

2.4. Selection Outcome

From both databases, a combined total of 14 publications were selected for the literature review: 7 from Google Scholar, and 7 from Scopus. These sources were chosen based on relevance to the topic, accessibility, and their contribution to understanding the role of ontologies in chatbot systems.

While this study employed Google Scholar and Scopus to identify relevant literature, the approach has several limitations. First, relying only on two databases may have excluded potentially relevant works indexed elsewhere or published in venues not covered by these platforms. Second, the search queries, although systematically applied, were constrained by the functionalities of each database and may not have captured all pertinent studies. Third, emerging tools such as AI-powered assistants can increasingly support literature exploration by offering more advanced semantic retrieval and cross-paper reasoning. However, these tools were not employed in this study in order to ensure transparency and reproducibility of results.

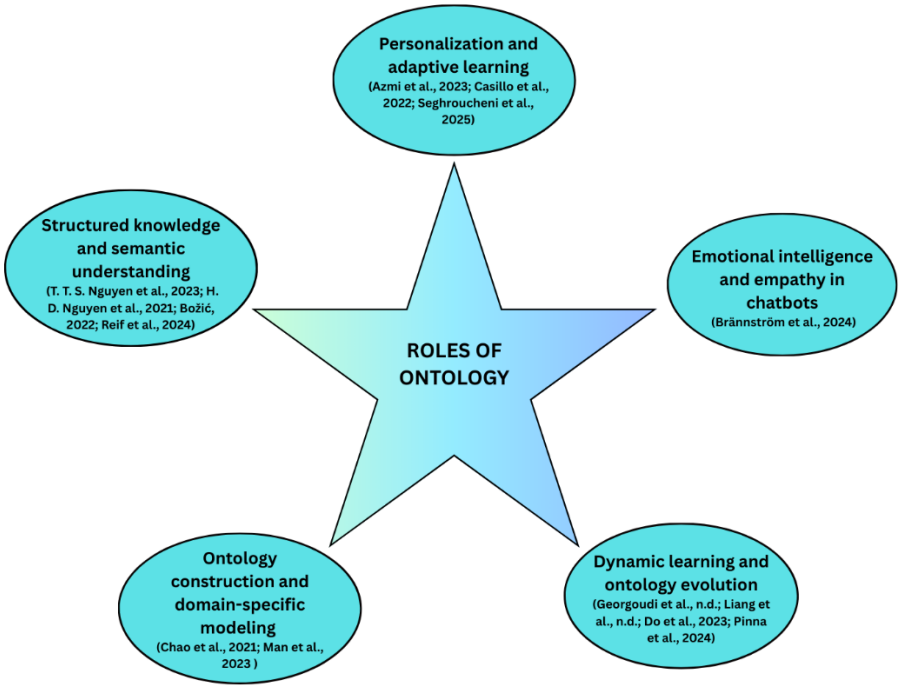


Figure 1. Roles of Ontology

3 Literature review

3.1. RQ1: What role do ontologies play in making chatbots easier to use, more accurate and more effective for personalized learning?

Based on the literature review, five key roles of ontology were identified that enhance the effectiveness and functionality of chatbots (as illustrated in Fig. 1): Structured knowledge and semantic understanding; Personalization and adaptive learning; Emotional intelligence and empathy in chatbots; Dynamic learning and ontology evolution; Ontology construction and domain-specific modelling.

3.1.1. Structured knowledge and semantic understanding

Ontologies play a key role in helping chatbots understand user queries by providing a structured way to represent domain knowledge. This improves the accuracy, consistency, and relevance of chatbot responses, especially in educational settings.

A university admissions chatbot was developed using a domain-specific ontology that enabled it to generate detailed responses and handle a wide range of student questions. The system can reason over structured data to provide long, meaningful answers. It's also designed to be expandable, allowing new knowledge to be added over time (T. T. S. Nguyen et al., 2023). In the context of programming education, a chatbot was built using the Rela-model ontology, which represents concepts, relationships, and rules relevant to the curriculum. This structure allowed the chatbot to classify questions accurately and support personalized guidance for students learning to code (H. D. Nguyen et al., 2021). Ontologies also play a role in improving the reliability and scalability of chatbots. By using structured semantic rules, developers can perform metamorphic testing to ensure consistent behaviour and easier system maintenance (Božić, 2022). In more complex environments, ontologies are combined with large language models to make knowledge more accessible. One approach enables users to ask questions in natural language, which the system then translates into SPARQL queries for retrieving accurate information, improving usability for non-expert users (Reif et al., 2024).

These examples show how structured knowledge, enabled by ontologies, supports better understanding, higher accuracy, and more scalable chatbot solutions in education.

3.1.2. Personalization and adaptive learning

Ontologies are widely used to personalize chatbot interactions by adapting content to individual users' needs, preferences, or learning contexts. This makes educational chatbots more engaging, responsive, and effective in supporting user goals.

A chatbot designed to recommend healthy food options for teenagers uses an ontology to deliver personalized suggestions based on individual health conditions and dietary needs. The system applies Semantic Web Rule Language (SWRL) rules to reason over user input, such as calorie goals and medical restrictions, ensuring accurate and tailored advice (Azmi et al., 2023). A chatbot called HeriBot

personalizes educational experiences for museum visitors. By using a purpose-built ontology representing the Archaeological Urban Park of Naples, the system adjusts its tone, language, and depth of information based on user motivations and profiles (Casillo et al., 2022). Ontologies also support adaptive learning by helping chatbots interpret and respond to learners' evolving needs. In e-learning platforms, they enable the conversion of tacit knowledge into structured, usable formats. When combined with NLP, it allows the chatbot to provide responses aligned with the user's intent and the knowledge base. This approach enhances both personalization and content delivery (Seghroucheni et al., 2025).

3.1.3. Emotional intelligence and empathy in chatbots

Ontologies can also play a key role in enabling chatbots to understand and respond to users' emotions, making interactions more humanized and supportive, especially in sensitive domains like health and well-being.

A proposed empathy ontology offers a structured way for chatbots to model emotional intelligence by organizing empathy-related concepts into a clear, hierarchical framework. This allows the chatbot to interpret emotional cues more accurately and respond in appropriate way (Brännström et al., 2024).

3.1.4. Dynamic learning and ontology evolution

In dynamic learning environments, ontologies are used not only to organize existing knowledge but also to expand and evolve as new information becomes available. This adaptability is crucial for chatbots that operate in changing domains and interact with users in unpredictable ways.

One approach introduces a dynamic ontology extension system that updates the chatbot's knowledge base in response to new user input and shifts in domain content. New entity types and relationships are added automatically, improving the chatbot's ability to provide accurate and up-to-date responses (Georgoudi et al., n.d.). Another study categorizes dynamic learning into tasks such as discovering new user intents and slot values during real-time interactions. This technique, known as ontology expansion, allows chatbots to handle novel inputs and maintain flexibility in unpredictable conversational contexts (Liang et al., n.d.). A hybrid model combines ontological reasoning with an autoregressive language model (XLNet) to fill in gaps where structured knowledge alone is not enough. This allows the chatbot to maintain semantic precision while generating fluent, natural responses (Do et al., 2023). To support long-term learning, a modular framework has been proposed that enables domain-specific chatbots to continuously incorporate new data and improve over time. The use of ontologies in this system allows scalable knowledge integration and sustained relevance across different educational contexts (Pinna et al., 2024).

3.1.5. Ontology construction and domain-specific modeling

In some applications, the focus is not just on using ontologies but on constructing them to fit specific domains. These domain-specific models help chatbots understand specialized topics, organize complex information, and deliver relevant responses.

One study applies text mining techniques such as keyword frequency analysis, clustering, and Latent Dirichlet Allocation (LDA) to build an ontology from patent data (Chao et al., 2021). Another example comes from the healthcare education domain, where a chatbot uses a modular ontology system divided into general and user-specific components. The chatbot personalizes information delivery using a layered ontology design and ensures regulatory compliance, including data privacy and GDPR adherence (Man et al., 2023).

3.2. RQ2: What are the application domains and case studies that demonstrate the use of ontologies in educational contexts?

This review identifies five key domains where ontologies are applied to enhance chatbot functionality in educational contexts: University admissions; Health and nutrition education**;** Programming and technical education**;** Cultural heritage and experiential learning**;** General educational chatbot systems.

In the domain of **university admissions**, chatbots support students by answering questions about application procedures, entry requirements, and academic offerings. Ontologies are used to structure institutional knowledge, enabling the chatbot to reason and generate accurate, context-aware responses (T. T. S. Nguyen et al., 2023), (Do et al., 2023).

Within **health and nutrition education**, chatbots are designed to provide personalized guidance based on user-specific data. One paper (Azmi et al., 2023) delivers nutrition advice tailored to adolescents' health profiles, while other supports applicants with relevant health program information and ensures privacy and compliance with data regulations (Man et al., 2023).

The **programming and technical education** domain features chatbots that guide learners through course content and problem-solving tasks (H. D. Nguyen et al., 2021). In more informal and experience-based learning contexts, **cultural heritage and experiential learning** benefit from ontology driven chatbots that adapt educational content to the user. One study presents HeriBot, which personalizes museum visits by adjusting language, tone, and content depth based on visitor profiles and motivations (Casillo et al., 2022).

Finally, several papers explore **general educational chatbot systems** that span multiple learning domains or are designed for long-term adaptability. For example, one study focuses on

systems that evolve their knowledge bases dynamically based on user interactions (Georgoudi et al., n.d.). Similarly, there is a study that provides a scalable and reusable architecture that supports chatbot development across varied educational settings (Pinna et al., 2024).

These diverse case studies illustrate how ontology-based chatbots are being successfully applied across formal, informal, and hybrid learning environments.

3.3. Findings overview

Table 1. presents, for each analysed paper, the role of ontology and its application domain.

Table 1. Paper analysis

Paper Title	Role of Ontology	Application Domain
An Ontology-Based Question Answering System for University Admissions Advising *(T. T. S. Nguyen et al., 2023)*	Structured Knowledge and Semantic Understanding	University Admissions
A Collaboration of an Ontology and an Autoregressive Model to Build an Efficient Chatbot Model *(Do et al., 2023)*	Dynamic Learning and Ontology Evolution	University Admissions
Recommendation System in the Form of an Ontology-Based Chatbot for Healthy Food Recommendations for Teenagers *(Azmi et al., 2023)*	Personalization and Adaptive Learning	Health and Nutrition Education
An Ontology-Based Chatbot to Enhance Experiential Learning in a Cultural Heritage Scenario *(Casillo et al., 2022)*	Personalization and Adaptive Learning	Cultural Heritage
Ontology-Based Integration of Knowledge Base for Building an Intelligent Searching Chatbot *(H. D. Nguyen et al., 2021)*	Structured Knowledge and Semantic Understanding	Programming and Technical Education
A Formal Understanding of Computational Empathy in Interactive Agents *(Brännström et al., 2024)*	Emotional Intelligence and Empathy in Chatbots	Health and Nutrition Education
Systematic Review and Framework for AI-Driven Tacit Knowledge Conversion Methods and Machine Learning Algorithms for Ontology-Based Chatbots in E-Learning Platforms *(Seghroucheni et al., 2025)*	Personalization and Adaptive Learning	General Educational Chatbots and Systems
Chatbot-Based Ontology Interaction Using Large Language Models and Domain-Specific Standards *(Reif et al., 2024)*	Structured Knowledge and Semantic Understanding	General Educational Chatbots and Systems
DYNAMO: Dynamic Ontology Extension for Augmenting Chatbot Intelligence through BabelNet *(Georgoudi et al., n.d.)*	Dynamic Learning and Ontology Evolution	General Educational Chatbots and Systems
A Survey of Ontology Expansion for Conversational Understanding *(Liang et al., n.d.)*	Dynamic Learning and Ontology Evolution	General Educational Chatbots and Systems
The Innovative Use of Intelligent Chatbot for Sustainable Health Education Admission Process: Learnt Lessons and Good Practices *(Man et al., 2023)*	Ontology Construction and Domain-Specific Modeling	Health and Nutrition Education
Technology Mining for Intelligent Chatbot Development *(Chao et al., 2021)*	Ontology Construction and Domain-Specific Modeling	General Educational Chatbots and Systems
Ontology-based Metamorphic Testing for Chatbots *(Božić, 2022)*	Structured Knowledge and Semantic Understanding	General Educational Chatbots and Systems
A Modular Framework for Domain-Specific	Dynamic Learning and	General Educational

Paper Title	Role of Ontology	Application Domain
Conversational Systems Powered by Never-Ending Learning *(Pinna et al., 2024)*	Ontology Evolution	Chatbots and Systems

The graphs below present an overview of the number of studies that discussed each specific role of ontologies (Fig. 2), as well as the number of publications categorized by their application domains (Fig. 3).

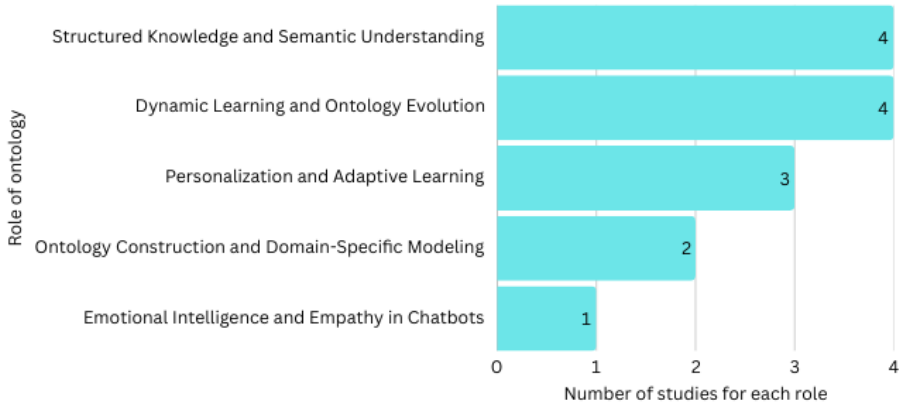


Figure 2. Role of ontology for each paper

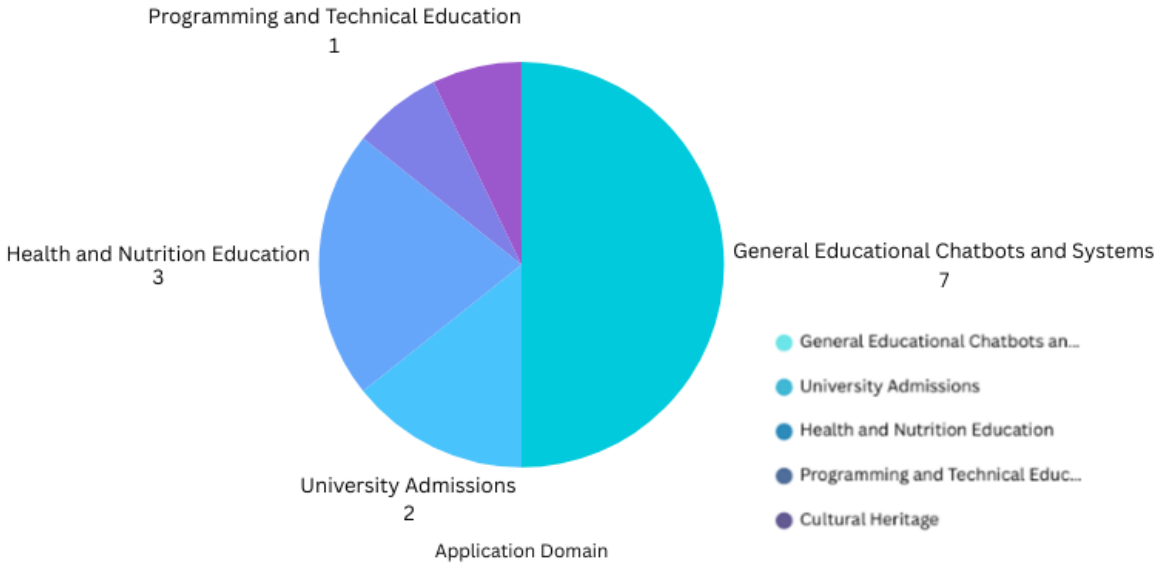


Figure 3. Frequency of application domains

4 Discussion and implications

Considering the results of the first research question, the review shows that ontologies significantly enhance the usability, accuracy, and overall effectiveness of chatbots. Ontologies enable a structured representation of knowledge, allowing chatbots to interpret user inputs more precisely. The ability to structure complex information ensures that these systems can deliver personalized experiences. In the field of education, structured knowledge and semantic understanding are key to creating successful learning experiences. This trend is particularly evident in the **4 papers** focused on **dynamic learning and ontology evolution**, **4 papers** highlighting **structured knowledge and semantic understanding**, and **3 papers** dedicated to **personalization and adaptive learning**. These findings suggest a growing need for dynamic ontology models that can continuously evolve to reflect emerging knowledge and user needs. Additionally, incorporating structured knowledge and adaptive learning could be essential to enhancing learning experience.

Ontology-based chatbots are being applied across diverse domains, with a significant focus on **general educational systems** (7 papers), **health and nutrition education** (3 papers) and **university admissions** (2 papers). These systems provide personalized learning experiences and enhance user engagement by adapting to individual needs. Future research can focus on cross-domain applications, where ontologies from various sectors are integrated to enhance chatbot intelligence.

5 Conclusion and future work

Ontologies play a foundational role in enhancing the design, functionality, and impact of chatbots in educational contexts. As demonstrated across multiple studies, ontologies provide structured, semantic frameworks that enable chatbots to understand user queries more accurately. The reviewed research highlights that ontologies contribute not only to technical precision but also to pedagogical value.

The main contributions of this paper are insights into the role that ontologies play in making chatbots easier to use, more accurate and more effective for personalized learning, as well as categorization of application domains that demonstrate the use of ontologies in educational contexts.

In domains such as university admissions, programming education, and cultural heritage learning, domain-specific ontologies allow chatbots to deliver contextually relevant information. Furthermore, adaptive and personalized systems enhance learner engagement and support differentiated instruction, especially when combined with natural language processing and rule-based inference.

Future research could explore the use of specific ontologies to support the development of chatbots designed to assist students in higher education institutions within e-learning platforms. Integrating some of the analysed ontologies with existing Learning Management Systems (LMS) could enable automated progress tracking, learning recommendations, and real-time answers to frequently asked questions. At the same time, research should not be limited to ontologies alone. Emerging technologies such as large language models, knowledge graphs, and recommender systems also offer promising opportunities to enhance the intelligence and adaptability of virtual assistants in education.

Acknowledgements

The research has been co-funded by “uniri-iskusni-drustv-23-209 (3203)” under the project “Development of an e-learning course design model based on recommendations for managing student satisfaction in higher education”

References

Azmi, N., Richasdy, D., & Hasmawati. (2023). Recommendation System in the Form of an Ontology-Based Chatbot for Healthy Food Recommendations for Teenagers. Jurnal Penelitian Pendidikan IPA, 9(7), 5085–5091. https://doi.org/10.29303/jppipa.v9i7.4401

Božić, J. (2022). Ontology-based metamorphic testing for chatbots. Software Quality Journal, 30(1), 227–251. https://doi.org/10.1007/s11219-020-09544-9

Brännström, A., Wester, J., & Nieves, J. C. (2024). A formal understanding of computational empathy in interactive agents. Cognitive Systems Research, 85. https://doi.org/10.1016/j.cogsys.2023.101203

Casillo, M., De Santo, M., Mosca, R., & Santaniello, D. (2022). An Ontology-Based Chatbot to Enhance Experiential Learning in a Cultural Heritage Scenario. Frontiers in Artificial Intelligence, 5. https://doi.org/10.3389/frai.2022.808281

Chao, M. H., Trappey, A. J. C., Wu, C. T., & Su, Y. A. (2021). Technology mining for intelligent chatbot development. Advances in Transdisciplinary Engineering, 16, 123–132. https://doi.org/10.3233/ATDE210090

Do, P. M. T., Nguyen, N. T. A., Ho, D. H. T., & Nguyen, T. T. S. (2023). A collaboration of an ontology and an autoregressive model to build an efficient chatbot model. International Journal of Intelligent Information and Database Systems, 1(1). https://doi.org/10.1504/ijiids.2023.10059348

Georgoudi, A., Meditskos, G., Mavropoulos, T., Vrochidis, S., & Kompatsiaris, I. (n.d.).

DYNAMO: Dynamic Ontology Extension for Augmenting Chatbot Intelligence through BabelNet.

Lecce, V., Calabrese, M., Soldo, D., & Giove, A. (2010). Semantic Management Systems for the Material Support of E-learning Platforms. Journal of E-Learning and Knowledge Society, 6, 61–70. https://doi.org/10.20368/1971-8829/445

Liang, J., Wu, Y., Fang, Y., Fei, H., & Liao, L. (n.d.). A Survey of Ontology Expansion for Conversational Understanding. https://github.com/liangjinggui/

MA, W., MA, W., HU, Y., & BI, X. (2024). The who, why, and how of ai-based chatbots for learning and teaching in higher education: A systematic review. Education and Information Technologies. https://doi.org/10.1007/s10639-024-13128-6

Man, S. C., Matei, O., Faragau, T., Andreica, L., & Daraba, D. (2023). The Innovative Use of Intelligent Chatbot for Sustainable Health Education Admission Process: Learnt Lessons and Good Practices. Applied Sciences (Switzerland), 13(4). https://doi.org/10.3390/app13042415

Nguyen, H. D., Tran, T. V., Pham, X. T., Huynh, A. T., & Do, N. V. (2021). Ontology-based integration of knowledge base for building an intelligent searching chatbot. Sensors and Materials, 33(9), 3101–3123. https://doi.org/10.18494/SAM.2021.3264

Nguyen, T. T. S., Ho, D. H. T., & Nguyen, N. T. A. (2023). An Ontology-Based Question Answering System for University Admissions Advising. Intelligent Automation and Soft Computing, 36(1), 601–616. https://doi.org/10.32604/iasc.2023.032080

Pinna, F. C. de A., Hayashi, V. T., Néto, J. C., Marquesone, R. de F. P., Duarte, M. C., Okada, R. S., & Ruggiero, W. V. (2024). A Modular Framework for Domain-Specific Conversational Systems Powered by Never-Ending Learning. Applied Sciences (Switzerland), 14(4). https://doi.org/10.3390/app14041585

Reif, J., Jeleniewski, T., Gill, M. S., Gehlhoff, F., & Fay, A. (2024). Chatbot-Based Ontology Interaction Using Large Language Models and Domain-Specific Standards. https://doi.org/10.1109/ETFA61755.2024.10711065

Seghroucheni, O. Z., Lazaar, M., & Al Achhab, M. (2025). Systematic Review and Framework for AI-Driven Tacit Knowledge Conversion Methods and Machine Learning Algorithms for Ontology-Based Chatbots in E-Learning Platforms. International Journal of Interactive Mobile Technologies , 19(1), 126–139. https://doi.org/10.3991/ijim.v19i01.51051

Sermet, Y., & Demir, I. (2020). A Semantic Web Framework for Automated Smart Assistants: COVID-19 Case Study. ArXiv, abs/2007.00747. https://consensus.app/papers/a-semantic-web-framework-for-automated-smart-assistants-sermet-demir/11ae2ec1f56652c3923fa0b8d30ff0d8/

Van Poucke, M. (2024). ChatGPT, the perfect virtual teaching assistant? Ideological bias in learner-chatbot interactions. Computers and Composition. https://doi.org/10.1016/j.compcom.2024.102871

Relational Modelling of Cyberattacks: A Taxonomy for the Analysis of Threat Interactions in Digital Environments

Mikel Ferrer-Oliva, José-Amelio Medina-Merodio, José-Javier Martínez-Herraiz, Carlos Cilleruelo-Rodríguez

Departamento de Ciencias de la Computación, Universidad de Alcalá, España

Alcalá de Henares, España

{mikel.ferrer, josea.medina, josej.martinez, carlos.cilleruelo}@uah.es

Abstract. *Cyberattacks have evolved into coordinated operations that combine multiple techniques to disrupt digital infrastructures. Existing classification methods often overlook the interdependence between attack vectors, limiting early detection and strategic response. This study introduces a relational taxonomy composed of eight attack groups and twenty connections, enabling a non-hierarchical understanding of how threats interact and escalate.*
The model captures dynamic relationships across multiple attack phases and supports the integration of emerging techniques. Its open structure enhances adaptability and analytical depth, offering practical value for cybersecurity operations within corporate and critical environments. This contribution addresses key limitations in static threat classification frameworks.

Keywords. Cybersecurity, taxonomy, threats, risk modelling, cyber defence.

1 Introduction

Digital transformation has significantly redefined the landscape of technological infrastructures, exposing organizations and institutions to a broader spectrum of advanced cyber threats. These threats have evolved from isolated actions into complex operations where diverse techniques are orchestrated to maximize impact and compromise resilience across interconnected systems.

Although current cybersecurity frameworks provide valuable mechanisms for classifying and exchanging threat information, such as MITRE ATT&CK (MITRE, 2025), STIX2 (TC, 2025a), TAXII 2.1 (TC, 2025b), and MISP (CIRCL, 2025), their descriptive nature tends to focus on individual tactics. These approaches frequently fail to convey how techniques interact within attack sequences, thereby limiting their predictive capacity and response effectiveness.

Documented analyses of real incidents demonstrate that attacks typically unfold as a chain of events, where an initial intrusion enables successive steps, including privilege escalation, long-term access, and lateral expansion across digital environments (ENISA, 2021). This progression highlights the necessity of a relational understanding of how attack vectors reinforce one another throughout the adversarial lifecycle.

For example, social engineering continues to serve as a catalyst for compromising credentials and delivering malware, granting unauthorized access to high-value environments (Bhardwaj & Sapra, 2020; Hellemann, 2023). In parallel, the exploitation of software vulnerabilities remains a prevalent mechanism for maintaining persistence and infiltrating sensitive systems (Clavijo Mesa et al., 2024; Connolly et al., 2023). However, the prevailing classification models lack the structural capacity to reflect these strategic interactions. This paper introduces a relational taxonomy that groups cyber threats according to shared tactical and operational characteristics, while mapping the connections that enable their interaction. Unlike rigid categorizations, the model is designed as a non-hierarchical system capable of representing complex relationships among diverse offensive techniques. Its purpose is to support more accurate threat analysis, informed risk assessment, and adaptive response strategies in dynamically evolving cybersecurity environments.

The manuscript is structured as follows: Section 2 presents a review of existing taxonomies and their limitations. Section 3 details the proposed model, including its categories and relational links. Section 4 discusses practical implications and real-world applications. Section 5 concludes with a summary and future lines of research.

2 Theoretical Framework

Cyber threats evolve constantly, prompting the creation of taxonomies to structure and classify

malicious behaviours. Ontology-driven efforts like CASE formalise events but show limited adoption, while platforms like MISP facilitate sharing yet do not model the strategic interplay among vectors (CIRCL, 2025; C. Community, 2025). Most systems remain static: even MITRE ATT&CK—despite its granularity—cannot depict the dynamic sequences observed in real incidents (MITRE, 2025). In practice, attacks unfold through chained techniques that escalate access or maintain persistence, underscoring the need for taxonomies that capture interdependencies and progression (ENISA, 2021). Empirical patterns—watering hole leveraging trusted sites (Alrwais et al., 2016), coordinated ransomware + DDoS in industrial/maritime contexts (Clavijo Mesa et al., 2024; Ivanov et al., 2021), and the surge of ransomware amid remote work and credential leaks—reinforce this need (Beaman et al., 2021; Rauf et al., 2023; Salim et al., 2019).

2.1 Limitations of Static Taxonomies

Frameworks like MITRE ATT&CK (MITRE, 2025) catalogue techniques in rigid categories. Although informative, they overlook how tactics combine across stages. This limits their capacity to model threat progression.

Attackers adapt their strategies based on the environment, chaining actions that static models fail to capture (Connolly et al., 2023). Watering hole attacks, for instance, involve multiple phases of deception and exploitation (Alrwais et al., 2016). In industrial and maritime settings, coordinated ransomware and DDoS incidents underline the need for models that reflect overlapping behaviours (Clavijo Mesa et al., 2024; Ivanov et al., 2021). The surge in ransomware due to remote work and credential leaks reinforces this urgency (Beaman et al., 2021).

Without accounting for technique interrelations, traditional taxonomies struggle to support predictive analysis or adaptive defence (Rauf et al., 2023; Salim et al., 2019).

2.2 Intrusion Chains and Implications for Risk Assessment and Incident Response

Cyberattacks typically evolve as ordered sequences, where each step enables the next (Hutchins et al., 2011; Javeed et al., 2020). Frequent entry vectors—phishing and vishing—lead to credential theft or malware deployment (Álvarez et al., 2024; Bhardwaj & Sapra, 2020; CISA & FBI, 2021). Campaigns such as Lucifer combine cryptojacking, DDoS, and exploitation in a single flow, while IoT botnets assemble large-scale infrastructures (Gelgi et al., 2024; Networks, 2025; Niño, 2023; Wu et al., 2021). Static frameworks often obscure these dependencies, reducing their usefulness for anticipatory defence. Standards like STIX 2.1 and TAXII 2.1 enable structured sharing but, by design, do not impose relational semantics; MISP aids classification and exchange yet lacks native constructs for stage transitions (CIRCL, 2025; TC, 2025a, 2025b). As emerging vectors—e.g., supply chain attacks—demand models that capture coordination (ENISA, 2021), prior proposals such as the Diamond Model and VERIS acknowledge links but remain limited in operational use (Caltagirone et al., 2013; V. Community, 2025; Sedano Pinzón, 2024). A relational taxonomy aligns classification with how real attacks unfold, making explicit the enabling role of each stage and strengthening both risk assessment and incident response (Clavijo Mesa et al., 2024; Hutchins et al., 2011; Javeed et al., 2020).

3 Proposed Cyberattack Taxonomy

The dynamic nature of cyber threats requires models capable of representing the interconnection between offensive techniques, rather than viewing them as isolated events. In adversarial contexts, attacks frequently manifest as interrelated actions, where each tactic contributes to achieving broader strategic goals. Recognizing this complexity is essential for improving the accuracy of threat analysis and the responsiveness of cybersecurity frameworks.

This section introduces a relational taxonomy of cyberattacks designed to map the functional role and interdependencies of each technique within an evolving threat environment. The model comprises eight distinct attack groups, each representing a set of behaviours or tactics commonly observed in real incidents. These groups are not organized in hierarchical order, but rather positioned within a network of relationships that reflects how they interact and reinforce one another.

The proposed structure captures both direct and indirect connections between attack groups. This approach enables analysts to identify common escalation paths, detect potential facilitators of advanced threats, and understand how certain techniques may contribute to system-wide compromise. By incorporating twenty well-defined relationships, the taxonomy supports a non-linear representation of attack progression, providing a more realistic basis for scenario-based risk modelling. Rather than introducing isolated categories disconnected from established knowledge, this model recontextualizes existing techniques within a relational framework that emphasises functional interdependence. By focusing on how techniques interact rather than how they are segmented, the approach facilitates the seamless incorporation of new attack modalities while preserving structural integrity. Its adaptability allows the taxonomy to remain operationally consistent as threat landscapes evolve, offering a scalable tool for understanding the convergence of diverse cyber tactics. In the following subsections, each attack group is described in detail

(Section 3.1), followed by an explanation of the strategic relationships that connect them (Section 3.2), and finally, a visual representation of these connections through a non-hierarchical diagram (Section 3.3).

3.1 Classification of Attacks Groups

This model facilitates the identification of recurring patterns in threat evolution, supporting the anticipation of escalation strategies commonly observed in complex cyber incidents. Table 1 presents the eight attack groups proposed in the taxonomy, each defined by the nature of the techniques involved and their role in documented cases. The classification does not follow a sequential or hierarchical structure but reflects distinct operational behaviours. Social engineering (such as phishing and vishing) exploits human factors to obtain credentials or trigger malicious actions (Bhardwaj & Sapra, 2020; Hellemann, 2023). These tactics often precede malware-based attacks involving ransomware, trojans or botnets used to disrupt systems or enable lateral movement (Ivanov et al., 2021).

Many incidents also involve exploiting software vulnerabilities in applications or firmware to gain privileged access (Networks, 2025; Wu et al., 2021). These are frequently combined with identity and authentication attacks based on credential stuffing or brute-force methods (Hellemann, 2023).

Other vectors include attacks on network infrastructure, which affect routers or internal devices to degrade services or bypass segmentation (Gelgi et al., 2024), and protocol-based attacks that intercept or manipulate data flows (Javeed et al., 2020).

In operational environments, threats against IT/OT infrastructure may compromise industrial processes (Clavijo Mesa et al., 2024). Advanced persistent threats and cyberespionage campaigns combine several tactics to maintain long-term access to strategic systems (Connolly et al., 2023; Rauf et al., 2023).

Table 1. Attack Groups in the Proposed Taxonomy

Strike Group	Description
Social Engineering (SE)	Use of psychological manipulation to gain access to credentials or resources. Phishing attacks have been identified as one of the leading causes of security breaches in companies (Bhardwaj & Sapra, 2020).
Malware-based attacks (MBA)	Use of malware to compromise systems, facilitating attacks such as ransomware, Trojans, and botnets. The analysis of new malware variants has proven its adaptability to evade detections (Ivanov et al., 2021).
Network Infrastructure Attacks (NIA)	Attacks targeting network devices, servers, and interconnected systems to disrupt services or facilitate unauthorized access. Recent research has looked at the impact of IoT botnets on massive DDoS attacks (Gelgi et al., 2024).
Exploiting Software Vulnerabilities (ESV)	Exploiting flaws in software to escalate privileges, install malware, or compromise critical systems. The persistence of unpatched vulnerabilities has been a key factor in multiple attack campaigns (Wu et al., 2021).
Attacks on Protocols and Communications (APC)	Compromise of communication protocols through techniques such as Man in the Middle attacks and DNS hijacking, facilitating the interception of sensitive data. Studies have shown the vulnerability of industrial protocols to targeted attacks (Javeed et al., 2020).
Identity and Authentication Attacks (IAA)	Compromise of credentials and authentication systems through attacks such as credential stuffing, brute force, and dictionary attacks. Exploiting credentials in corporate environments remains one of the main initial access tactics (Hellemann, 2023).
Attacks on critical IT/OT infrastructure (CIIA)	Attacks targeting industrial networks and SCADA systems for sabotage and espionage. Device tampering in critical environments has been documented as a significant risk to operational safety (Clavijo Mesa et al., 2024).
APTs and Cyberespionage (APT)	Persistent infiltration and espionage operations that combine multiple attack vectors to maintain longtermly access to strategic government and enterprise networks. Cyberespionage continues to evolve with more sophisticated persistence tactics (Connolly & Wall, 2019).

3.2 Relationships Between Attack Groups

A central strength of the proposed taxonomy lies in its capacity to model how attack groups interact strategically throughout the lifecycle of cyber incidents. Rather than depicting static or isolated behaviours, the taxonomy captures how techniques from different domains reinforce each other, forming interconnected sequences that reflect the operational logic observed in real-world threats.

Table 2 outlines twenty directional relationships supported by case-based evidence and intelligence reports. These links describe how the use of one type

of technique often enables, amplifies or conditions the success of others within multi-stage campaigns.

Initial compromise frequently involves social engineering tactics such as phishing or vishing (Álvarez et al., 2024; Bhardwaj & Sapra, 2020), which often lead to the extraction of access credentials (Hellemann, 2023; Ivanov et al., 2021). With these credentials, attackers can bypass authentication mechanisms, escalate privileges or deploy malware components tailored to specific targets (Clavijo Mesa et al., 2024; Gelgi et al., 2024). Malware, in turn, operates as a pivot between multiple attack vectors, supporting system manipulation, network disruption or covert persistence.

Exploitation of software vulnerabilities plays a complementary role, allowing unauthorized modifications in services and applications to establish control over critical assets (Networks, 2025; Wu et al., 2021). This technique is commonly associated with identity-based intrusions, particularly when authentication protocols are weak or reused credentials are exposed (Connolly et al., 2023; Rauf et al., 2023).

Attacks targeting communication protocols and network infrastructure add another layer of complexity, enabling packet interception, redirection or disruption of legitimate traffic flows (Javeed et al., 2020; Salim et al., 2019). When these same techniques are directed toward operational environments, they can severely impact critical IT/OT infrastructure, compromising industrial processes and degrading service continuity (Clavijo Mesa et al., 2024).

These diverse vectors converge in advanced persistent threats and cyberespionage operations, where multiple techniques are orchestrated to maintain long-term access and extract strategic information (Connolly et al., 2023; Rauf et al., 2023)).

By mapping these interdependencies, the taxonomy offers a relational view of attack progression that supports predictive modelling, improves early detection and enhances threat intelligence capabilities.

Table 2. Relationships between Attack Groups in Taxonomy

Initial Attack	Facilitated Attack	Description
Identity and Authentication Attacks	Social Engineering	Credentials access aids manipulation, increasing deception effectiveness for successful attacks (Bhardwaj & Sapra, 2020; Hellemann, 2023).
Identity and Authentication Attacks	Malware based attacks	Compromised systems via stolen credentials allow running unauthorized software, bypassing security to spread malware (Beaman et al., 2021; Ivanov et al., 2021).
Identity and Authentication Attacks	Exploiting Software Vulnerabilities	Improper auth/privileged access helps attackers find flaws, increasing compromise risk (Rauf et al., 2023; Wu et al., 2021).
Identity and Authentication Attacks	Attacks on Protocols and Communications	Exposed credentials allow channel access for data interception/alteration without needing vulnerability exploits (Javeed et al., 2020; Wu et al., 2021).
Identity and Authentication Attacks	Network Infrastructure Attacks	Improper admin access aids network manipulation for persistence or undetected info flow modification (Gelgi et al., 2024; Salim et al., 2019).
Social Engineering	Malware based attacks	Persuasive deception effectively induces users to run unsafe programs, enabling harmful actions (Álvarez et al., 2024; Bhardwaj & Sapra, 2020).
Social Engineering	Exploiting Software Vulnerabilities	Personalized persuasion leads victims to insecure platforms, risking unauthorized code execution (Alrwais et al., 2016; Rauf et al., 2023).
Malware based attacks	Attacks on Protocols and Communications	Some malware uses system communication for data manipulation/exfiltration without needing credentials (Javeed et al., 2020; Wu et al., 2021).
Malware based attacks	Network Infrastructure Attacks	Compromised devices can degrade networks and enable unauthorized internal access without human interaction (Clavijo Mesa et al., 2024; Gelgi et al., 2024).
Malware based attacks	Attacks on critical IT/OT infrastructure	Harmful software exploited to affect essential control systems, interrupting operations strategically (Clavijo Mesa et al., 2024; Niño, 2023).
Malware based attacks	Exploiting Software Vulnerabilities	Automated access helps find flaws in compromised environments, increasing persistence ability (Rauf et al., 2023; Wu et al., 2021).

Exploiting Software Vulnerabilities	Attacks on Protocols and Communications	Altering apps/services enables communication manipulation for data access without user intervention (Javeed et al., 2020; Wu et al., 2021).
Exploiting Software Vulnerabilities	Network Infrastructure Attacks	Software protection flaws allow compromising key network devices, affecting stability and enabling unauthorized access (Salim et al., 2019; Wu et al., 2021).
Exploiting Software Vulnerabilities	APTs and Cyberespionage	Persistent attacks use software flaws to infiltrate sensitive areas, allowing prolonged undetected access (Connolly & Wall, 2019; Rauf et al., 2023).
Attacks on Protocols and Communications	APTs and Cyberespionage	Manipulating device communication aids info exfiltration without needing credentials, creating high risks (Rauf et al., 2023; Wu et al., 2021).
Attacks on Protocols and Communications	Attacks on critical IT/OT infrastructure	Manipulating industrial communication systems allows remote process control, impacting operational continuity (Clavijo Mesa et al., 2024; Wu et al., 2021).
Attacks on Protocols and Communications	Network Infrastructure Attacks	Compromising communication protocols enables persistent attackers to establish network footholds for undetected data extraction (Connolly & Wall, 2019; Rauf et al., 2023).
Network Infrastructure Attacks	APTs and Cyberespionage	Compromised networks facilitate access to industrial environments, with significant strategic/operational impact (Clavijo Mesa et al., 2024; Niño, 2023).
Network Infrastructure Attacks	Attacks on critical IT/OT infrastructure	Prolonged infiltrations via network compromise enable espionage/process alteration, threatening stability/confidentiality (Connolly & Wall, 2019; Niño, 2023).
APTs and Cyberespionage	Attacks on critical IT/OT infrastructure	Prolonged APT access enables process manipulation and key info theft in globally important infrastructures (Clavijo Mesa et al., 2024; Rauf et al., 2023).

3.3 Taxonomy Framework

Fig. 1 presents a non hierarchical relational diagram. Each attack group is a node with evidence based directed links that indicate tactical enablement.

The view reveals multiple paths from phishing or credential compromise to malware and vulnerability exploitation or unauthorized access and also cyclic patterns typical of APTs. The scheme supports threat correlation and path simulation and predictive risk modelling.

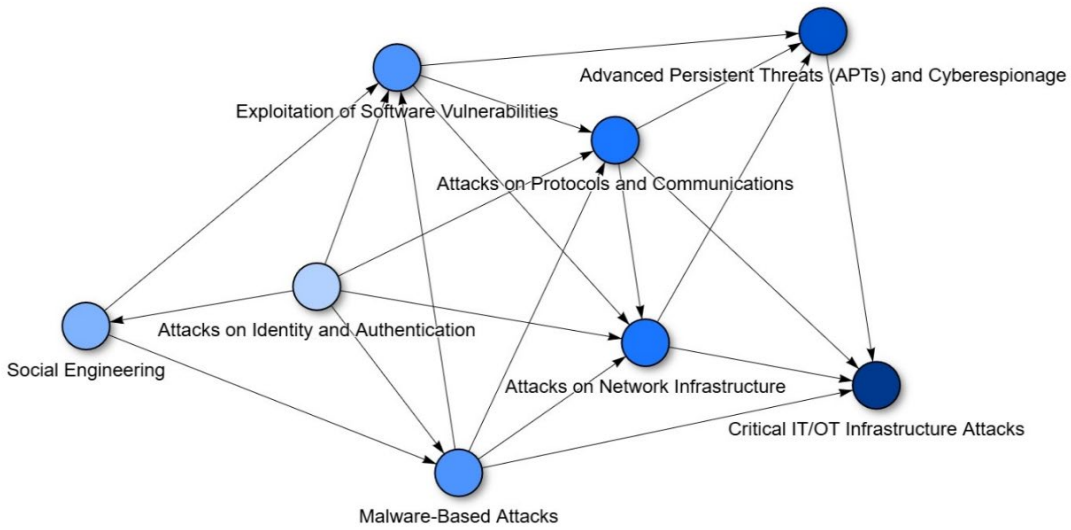


Figure 1. Framework of the Proposed Relational Taxonomy

4 Methodology

The taxonomy is built as a functional partition of eight groups defined by the tactical role they play within real intrusion chains, independent of platform, sector, or malware family. The decision to fix eight groups simultaneously pursues broad functional coverage and clear semantic separation so that the model avoids overlaps and preserves traceability to TTP catalogues and cyber-threat intelligence exchange frameworks (ENISA, 2021; Hutchins et al., 2011; MITRE, 2025).

The twenty directed relationships arise from an iterative refinement that retains only tactical-enablement links repeatedly observed in technical literature and incident reports. Each edge expresses a group's capacity to facilitate or condition the activation of another with causal meaning (not mere chronological succession). The structure supports the incorporation of emerging techniques as instances or, where appropriate, as new edges without reopening the base ontology. The result is directly integrable with STIX 2.1/TAXII 2.1, can be annotated in MISP, and is compatible with VERIS/CASE for operational exchange and correlation (CIRCL, 2025; C. Community, 2025; V. Community, 2025; TC, 2025a, 2025b).

5 Discussion

The increasing sophistication of cyberattacks demands models capable of reflecting the strategic interplay between techniques. Traditional taxonomies, while useful for cataloguing behaviours, often fail to represent how tactics interact during incident progression (Hutchins et al., 2011; MITRE, 2025). Many intrusions follow coordinated chains, where early actions enable more damaging phases (Alrwais et al., 2016; Connolly & Wall, 2019). Social engineering methods such as phishing are frequently used to gain credentials or introduce malicious payloads, which often lead to deeper compromise of critical systems (Bhardwaj & Sapra, 2020; Sedano Pinzón, 2024). Similarly, exploiting software vulnerabilities facilitates privilege escalation and persistence in sensitive environments (Clavijo Mesa et al., 2024; Wu et al., 2021).

The relational taxonomy proposed here addresses these limitations by mapping how one technique facilitates others within multi-stage attacks. It includes eight attack groups and twenty relationships derived from documented cases, enabling flexible representation of complex scenarios.

This adaptability is key when analysing emerging threats such as supply chain compromises or attacks in cloud-based infrastructures. By linking new behaviours to existing groups, the model avoids constant reclassification while maintaining consistency (ENISA, 2021; Hutchins et al., 2011).

Concrete examples reinforce its relevance. Malware like Lucifer combines cryptojacking, denial of service and exploitation in a single campaign (Networks, 2025; Salim et al., 2019). IoT botnets, exploiting protocol and firmware weaknesses, demonstrate how limited-entry vectors can lead to large-scale compromise (Gelgi et al., 2024). Protocol-based threats also support industrial espionage through data interception (Clavijo Mesa et al., 2024; Javeed et al., 2020).

Human factors remain central. Training and awareness have been shown to reduce social engineering effectiveness (Hellemann, 2023; Nassir et al., 2025). Existing frameworks like CASE or MISP offer structural approaches but lack the flexibility to capture threat progression in operational contexts (CIRCL, 2025; C. Community, 2025).

5.1 Case-led validation and scalability

The validity of the approach is evident in Stuxnet, where an APT-type operation established persistence, industrial-environment awareness, and evasion to enable a direct impact on OT infrastructure. The intrusion combined zero-day vulnerabilities, stolen certificates, and stealth techniques to infiltrate SCADA and alter Siemens PLCs, covertly modifying centrifuge process parameters; in the taxonomy, this path is modelled as APTs and Cyberespionage → Attacks on Critical IT/OT Infrastructure, confirming the directionality and causal meaning of the proposed relationship (Shakarian, 2012). The robustness of the network is reinforced by other campaigns that cover distinct edges of the graph: the WannaCry case illustrates Malware-Based Attacks → Attacks on Protocols and Communications by propagating the worm through an SMBv1 exploit (Smart, 2018); additionally, Mirai evidences Malware-Based Attacks → Network Infrastructure Attacks by compromising IoT devices and launching DDoS against network services (Cloudflare, 2022); and Sea Turtle shows Attacks on Protocols and Communications → Network Infrastructure Attacks/APTs via DNS hijacking that sustains covert and persistent access (CISCO, 2019). Maintaining a fixed conceptual base of eight nodes (groups) and twenty edges (relationships) facilitates scalability: operational growth comes from instances and relationship weights as events and sectors are incorporated, preserving structural coherence without relying on specific operational minutiae.

5.2 Integration with Standards and Knowledge Graphs

Each group is modelled as an abstract entity in STIX 2.1 (attack-pattern, malware, intrusion-set, or infrastructure, depending on role) and each of the twenty relationships as a directed enablement link

between entities. Distribution via TAXII 2.1 allows updatable collections where new evidence adds edges or annotations without modifying the base ontology (TC, 2025a, 2025b). In MISP, the taxonomy is annotated via a relational taxonomy/galaxy and coexists with MITRE ATT&CK to maintain TTP ↔ group traceability (CIRCL, 2025; MITRE, 2025). In knowledge graphs, the groups are functional nodes and the edges encode directed enablement with evidence/confidence attributes; this enables queries for probable paths, pivot-node detection, and flow analysis over incidents annotated with VERIS/CASE (C. Community, 2025; V. Community, 2025).

6 Conclusions

This work shows cyberattacks are interrelated sequences enabling escalation, not isolated actions. Addressing prior taxonomy limits, we proposed a relational model (8 groups, 20 relationships) capturing the dynamic complexity of real incidents.

The model contributes to understanding attack organization and progression by explaining technique interdependencies, helping identify patterns across various environments, including critical infrastructures. Organizationally, the model aids early detection and response. Analysing technique interaction helps anticipate attack evolution and mitigate impact, enabling efficient, coordinated responses and resource allocation, fostering preventative cultures.

Socially, the model highlights the human factor in social engineering and credential mismanagement. It underscores investigating human/technical vulnerability interactions and promotes awareness /training programs to foster good practices and reduce susceptibility to manipulation like phishing.

Limitations exist: the theoretical basis continues to be tested in operational threat environments. Constant threat evolution necessitates regular updates for ongoing adaptability. Future research includes (i) modeling the eight identified groups to empirically analyze characteristic attack behaviors, (ii) establishing explicit STIX 2.1 representations by formally defining entities, relationships, and inference rules, and (iii) validating the taxonomy's scalability in high-volume, real-time environments (e.g., SOC operations), where its performance has not yet been fully demonstrated.

Acknowledgments

This work has been developed within the "Recovery, Transformation and Resilience Plan", project C084/23 Ada Byron INCIBE-UAH, funded by the European Union (Next Generation).

References

Alrwais, S., Yuan, K., Alowaisheq, E., Liao, X., Oprea, A., Wang, X., & Li, Z. (2016). *Catching predators at watering holes: finding and understanding strategically compromised websites*. https://doi.org/10.1145/2991079.2991112

Álvarez, A. L., Cruz, J. A., Cruz, S. B., Gallardo, J. d. C., López, I. M., & García, R. E. (2024). El phishing como amenaza en la ciberseguridad corporativa de grandes empresas. *Investigaciones Latinoamericanas en Ingeniería y Arquitectura* (1), 26-33. https://doi.org/10.51378/ilia.vi1.8496

Beaman, C., Barkworth, A., Akande, T. D., Hakak, S., & Khan, M. K. (2021). Ransomware: Recent advances, analysis, challenges and future research directions. *Comput Secur, 111*, 102490. https://doi.org/10.1016/j.cose.2021.102490

Bhardwaj, A., & Sapra, V. (2020). Why is phishing still successful? *Computer Fraud & Security, 2020*, 15-19. https://doi.org/10.1016/S1361-3723(20)30098-1

Caltagirone, S., Pendergast, A., & Betz, C. (2013). *The Diamond Model of Intrusion Analysis* (ADA586960).

CIRCL. (2025, 2025-03-19). *MISP taxonomies and classification as machine tags.* CIRCL.LU. Retrieved 19-mar-2025 from https://www.misp-project.org/

CISA, & FBI. (2021). *TrickBot Malware* (AA21-076A). https://www.cisa.gov

CISCO. (2019). *Threats of the Year: A look back at the tactics and tools of 2019* (Cisco Cybersecurity Series 2019. Threat Report, Issue. http://www.cisco.com/go/securityreports

Clavijo Mesa, M. V., Patino-Rodriguez, C. E., & Guevara Carazas, F. J. (2024). Cybersecurity at Sea: A Literature Review of Cyber-Attack Impacts and Defenses in Maritime Supply Chains. *Information*, 15(11). https://doi.org/10.3390/info15110710

Cloudflare. (2022). *DNS and the Threat of DDoS*.

Community, C. (2025, 2025-03-19). *CASE: Cyber-investigation Analysis Standard Expression*. CASE Community. Retrieved 19-mar-2025 from https://caseontology.org/

Community, V. (2025, 2025-03-19). *VERIS: Vocabulary for Event Recording and Incident Sharing*. VERIS Community. https://verisframework.org/

Connolly, K., Klempay, A., McCann, M., & Brenner, P. (2023). Dark Web Marketplaces: Data for Collaborative Threat Intelligence. *ACM Digital*

Threats: Research and Practice, 4(4). https://doi.org/10.1145/3615666

Connolly, L., & Wall, D. S. (2019). The rise of crypto-ransomware in a changing cybercrime landscape: Taxonomising countermeasures. *Computers & Security, 87*. https://doi.org/10.1016/j.cose.2019.101568

ENISA. (2021). *ENISA Threat Landscape for Supply Chain Attacks*. https://www.enisa.europa.eu/sites/default/files/publications/ENISA%20Threat%20Landscape%20for%20Supply%20Chain%20Attacks.pdf

Gelgi, M., Guan, Y., Arunachala, S., Samba Siva Rao, M., & Dragoni, N. (2024). Systematic Literature Review of IoT Botnet DDOS Attacks and Evaluation of Detection Techniques. *Sensors (Basel),* 24(11). https://doi.org/10.3390/s24113571

Hellemann, N. (2023). *Human Risk Review 2023* [Report]. S. A. GmbH. https://www.sosafe-awareness.com/

Hutchins, E. M., Cloppert, M. J., & Amin, R. M. (2011). Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains. *Leading Issues in Information Warfare & Security Research,* 1(1), 80-106. https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf

Ivanov, M. A., Kliuchnikova, B. V., Chugunkov, I. V., & Plaksina, A. M. (2021). *Phishing Attacks and Protection Against Them 2021* IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus),

Javeed, D., MohammedBadamasi, U., Ndubuisi, C. O., Soomro, F., & Asif, M. (2020). Man in the Middle Attacks: Analysis, Motivation and Prevention. *International Journal of Computer Networks and Communications Security*, 8(7), 52-58. https://doi.org/10.13140/RG.2.2.22752.81928

MITRE. (2025). ATT&CK: Adversarial Tactics, Techniques, and Common Knowledge. In: MITRE Corporation.

Nassir, N. F. M., Rauf, U. F. A., Zainol, Z., & Ghani, K. A. (2025). Revealing the multi-perspective factors behind insider threats in cybersecurity. *Journal of Media and Information Warfare*, 17(2), 65-82.

Networks, P. A. (2025). *Lucifer: New Cryptojacking and DDoS Hybrid Malware Exploiting High and Critical Vulnerabilities to Infect Windows Devices*. https://unit42.paloaltonetworks.com

Niño, F. Y. Á. (2023). Ransomware, una amenaza latente en Latinoamérica. *InterSedes,* 24(49). https://doi.org/10.15517/isucr.v24i49

Rauf, U., Mohsen, F., & Wei, Z. (2023). A Taxonomic Classification of Insider Threats: Existing Techniques, Future Directions & Recommendations. *Journal of Cyber Security and Mobility.* https://doi.org/10.13052/jcsm2245-1439.1225

Salim, M. M., Rathore, S., & Park, J. H. (2019). Distributed denial of service attacks and itsdefenses in IoT: a survey. *The Journal of Supercomputing*, 76(7), 5320-5363. https://doi.org/10.1007/s11227-019-02945-z

Sedano Pinzón, J. J. (2024). El contexto actual e histórico de la ingeniería social. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades,* 5(5). https://doi.org/10.56712/latam.v5i5.2691

Shakarian, P. (2012). Stuxnet: Revolución de Ciberguerra en los Asuntos Militares. *Air and Space Power Journal*.

Smart, W. (2018). *Lessons learned review of the WannaCry Ransomware Cyber Attack*.

TC, O. C. T. I. (2025a). STIX Version 2.1. Committee Specification 02. In: OASIS.

TC, O. C. T. I. (2025b). TAXII Version 2.1. Committee Specification 01. In: OASIS.

Wu, Q., Zhang, S., Zheng, B., You, C., & Zhang, R. (2021). Intelligent Reflecting Surface-Aided Wireless Communications: A Tutorial. IEEE *Transactions on Communications,* 69(5), 3313-3351. https://doi.org/10.1109/tcomm.2021.3051897

Data Quality in Distributed Information Systems: Conceptual Challenges and Empirical Findings

Sven Vujević
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
svujevic21@foi.hr

Maja Cerjan, Kornelije Rabuzin
University of Zagreb Faculty of Organization and Informatics
Laboratory for Data Technologies
Pavlinska 2, 42000 Varaždin, Croatia
{macerjan,krabuzin}@foi.unizg.hr

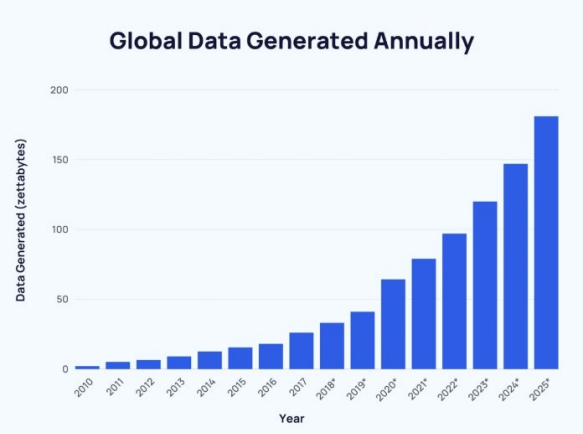
Abstract. *As is widely known by now, the quality of data is of great importance. However, conundrums arise if the data cleansing process isn't done appropriately from the very source. Especially in the digital world, driven by data in which its amounts are growing exponentially (Duarte, F., 2025). This paper emphasizes the importance of the quality of data in distributed information systems where the situation is even more perplexing and harder to resolve providing an error occurs.*
Therewithal, you will find the theoretical guidelines of things to do and ones to avoid, complying with ISO/IEC 25012 standard, either you are trying to get the grasp of your own data or you're working for an external business customer. To shed light on this growing problem there will also be concrete, firsthand examples from industry showing some of the most common fallacies and therefore challenges to overcome to ensure the quality of data for reliable decision making in the future.
This reading provides best practices when it comes to handling the data optimally both in theory and practice wherefore it is a great foundation for both field experts and field experts to be.

Keywords. data quality, distributed systems, data consistency, information systems, data cleansing, ISO/IEC 25012

1 Introduction

Regardless of the sector the company operates in, data-driven decisions have become the norm of business excellence and are inevitable if your goal is to stay competitive in the market. "Consequently, along with people, data can actually be considered as one of the most important assets for organizations" (Gualo, F. et al., 2021). It is trivial that data and its quality play an immense and crucial part in the process of decision making (Wang, R.Y., 1998), (Woodall, P. et al., 2012). According to IBM Data quality measures how well a dataset meets criteria for accuracy, completeness, validity, consistency, uniqueness, timeliness and fitness for purpose, and it is critical to all data governance initiatives within an organization. Furthermore, according to research by consulting firm Gartner, poor data quality costs companies an average of $12.9 million per year, while in the financial sector this amount can be significantly higher, and this is just one problematic example of this underlying problem. This only magnifies with the shire volume of ever new and exponentially bigger amounts of data being recorded every year (Duarte, F., 2025). "According to the latest estimates, 402.74 million terabytes of data are created each day" (Duarte, F., 2025).

Figure 1. Global data generated annually (Duarte, F., 2025)



This paradigm with the data as a cornerstone of the organizational activity and its success implies that the companies must fully commit to ensure the highest standards for the best outcomes, and the practice bolsters it (Gualo, F. et al., 2021). For example, when American Express began to foster more the importance of data quality with the accent on customer data accuracy, fraud detection, and behavioral analytics it had a huge positive impact which can be divided into three main parts: (1) delivering greater security for customers with 60% reduction in fraud by utilizing big data; (2) faster and smarter decisions backed with real-time data intelligence (3) long-term investment in innovation thus building a data-driven advantage (Llano, M., 2022). Another example of getting the best out of it is the story of Netflix. "We have embedded this approach into our culture from when the company was founded and call it Consumer (Data) Science. Broadly speaking, the main goal of Consumer Science

is to effectively innovate for users by using data to drive product decisions." (Amatriain, X., & Basilico, J., 2015). All this comes as motivation to invest your time and resources in studying this topic and implementing your findings in practice.

The sole process of ensuring high standards when it comes to data quality is composite and delicate work. There exist a lot of concepts and frameworks on this topic, some of which include the ISO/IEC 25012 standard that defines 15 dimensions that are nested across three different categories. By ISO/IEC 25012:2008 they are as follows: (1) Inherent; (2) Inherent and System-Dependent and (3) System-Dependent (Guerra-García, C. et al., 2023), (Miller, R. et al., 2024). Another approach is the DAQUAVORD methodology and its "data quality by design" that embeds data quality requirements into the early stages of information system development. DAQUAVORD is a relatively new concept (introduced in early 2023 by Guerra-García, C. et al.), nonetheless, based on the ISO/IEC 25012 standard.

In the following section one can find deeper overview of other relevant work done on this topic.

2 Related Work

2.1 Industrial Certification (ISO 25012)

Data quality has become a core asset for organizations, driving the need for systematic evaluation and certification (Gualo, F. et al., 2021). The ISO/IEC 25012 standard emerged as the gold standard and became central to data quality certification, offering a structured framework for evaluating and managing data quality within information systems (Gualo, F. et al., 2021). Initially, organizations recognized the importance of data quality for operational, tactical, and strategic activities (Gualo, F. et al., 2021). Poor data quality led to significant losses and hindered business initiatives, emphasizing the need for mechanisms to ensure data reliability (Gualo, F. et al., 2021). Just one of many such tremendous examples are Redman reported losses of about $3,1 billion for American companies due to poor and inadequate data quality, and the ever-growing complexity in system architecture which accounts for their high price (Redman, T. C., 2016), (Ballou, D. P., & Tayi, G. K., 1989).

Rather than treating all quality issues the same, international standards as ISO/IEC 25012 suggest evaluating data from multiple angles. Some aspects of quality stem from the data itself, such as whether it's correct or complete, while others depend on how systems manage, store, or present that data. By recognizing this distinction, organizations can take a more strategic approach by improving the data at its source while also ensuring that technological systems maintain their integrity throughout their lifecycle.

Apart from ISO/IEC 25012, other important standards include ISO/IEC 25024 and ISO/IEC 25040 standards, which are both part of the broader SQuaRE (Software Product Quality Requirements and Evaluation) series developed by ISO and IEC. They are helping to ensure quality in software and data systems in a broader, more holistic approach. ISO/IEC 25024:2015 defines a set of data quality measures for evaluating the quality of data used in systems and thus supporting the ISO/IEC 25012. Its purpose is to provide practical metrics and methods for measuring the actual quality of datasets, whether for internal auditing, improvement, certification, or regulatory purposes. Additionally, it defines quantitative metrics for each characteristic. It proves to be useful for data profiling and quality audits among others. ISO/IEC 25040:2024 defines a generic, structured process for evaluating software product quality, from planning to result reporting, based on requirements and context. It fulfils its purpose as repeatable evaluation framework that can be applied consistently across software products to determine whether they meet quality goals. It often works alongside ISO/IEC 25010:2023, which defines the quality model for software (e.g., reliability, usability, performance). Benefits of data quality certification include long-term organizational sustainability, better internal knowledge of data, and more efficient data quality management (Gualo, F. et al., 2021).

2.2 Alternative Certifications and Improvements on ISO 25012

However, challenges of data quality remain, including a lack of standardized terminology and the need for adaptable frameworks across different domains (Miller, R. et al., 2024). Recent work suggests enhancing the ISO 25012 standard with additional dimensions like governance, usefulness, quantity, and semantics to improve its applicability and relevance (Miller, R. et al., 2024). (1) Governance covers internal roles, authority, and accountability structures guiding data handling; (2) usefulness measures the extent to which data meets user/application needs, focusing on adaptability and reusability; (3) quantity considers whether there's a sufficient volume of data for a complete and accurate representation and (4) semantics captures the ability of data to convey correct and consistent meaning, crucial in contexts like ontologies or knowledge graphs (Miller, R. et al., 2024). This evolution aims to create a universal framework that supports effective data management and decision-making across various sectors (Miller, R. et al., 2024).

Another effort to improve upon ISO/IEC 25012 standard comes from Guerra-García et al. (2023) in their paper "ISO/IEC 25012-based methodology for managing data quality requirements in the development of information systems: Towards Data Quality by Design". Authors proposed structured methodology that significantly extends and operationalizes the ISO/IEC 25012:2008 standard. The solution included: (1) Integration of Data Quality into

Software Development Life Cycle (SDLC) - which encompassed the DAQUAVORD methodology, which embeds data quality requirements directly into the requirements engineering phase of system development (Guerra-García, C. et al., 2023). This shift supports the principle of "Data Quality by Design", ensuring that data quality concerns are proactively addressed, not retrofitted, which authors strongly encouraged (Guerra-García, C. et al., 2023). (2) Conceptual Extension is another such addition to current earlier discussed standard; it provides transformation from DQ dimensions to DQRs (Data Quality Requirements) and DQSRs (Data Quality Software Requirements) (Guerra-García, C. et al., 2023). Other key components are (3) Viewpoint-Oriented Requirements Definition (VORD) Integration; (4) Step-by-Step Methodology and Role Assignments (DAQUAVORD framework introduces a six-phase process) and (5) Realistic Use Case Implementation (the methodology is demonstrated using a university student registration system case study), (Guerra-García, C. et al., 2023).

2.3 Data Quality in Distributed Information Systems

Data quality in distributed data networks involves unique challenges compared to traditional, centralized systems due to the variety of data sources and the complexity of integrating data across multiple institutions (Kahn, M. G. et al. 2015). This implies additional steps in the process of data cleansing and establishing a sufficient level of the quality of data. Therefore, authors have been proposing different approaches to handling the challenge.

Some examples of the mentioned differences between the centralized and distributed systems include differences in a few key areas.

Table 1. Differences over data ownership and autonomy (Kahn, M. G. et al. 2015)

Centralized Systems:	**Distributed systems:**
Single point of control over data management	Multiple autonomous data providers/sites
Easier to enforce uniform data quality polices	Quality assessment varies per site and may lack transparency

Another significant difference between these two system architectures are the tools used to build, maintain and manipulate them. For example, for ETL process (Extract, Transform, Load) in centralized systems, the most common tools used are Informatica and Talend while for distributed systems some of the most used tools include Apache Spark, Kafka and Delta Lake among others (Margara, A. et al. 2022).

Table 2. Differences over cleansing and reconciliation complexity (Margara, A. et al. 2022)

Centralized Data Cleansing:	**Distributed Data Cleansing:**
Cleansing rules apply uniformly	Each node may apply its own logic or delay synchronization
Easier duplicate and inconsistency detection	Harder duplicate and inconsistency detection

Table 3. Quality assessment reporting (Kahn, M. G. et al. 2015)

Traditional systems:	**Distributed systems:**
Single-point metrics (accuracy, completeness)	Requires multi-site, temporal and source-specific indicators
Errors detected globally	Errors may be site-specific, temporal

3 Theoretical Framework: Dimensions of Data Quality

3.1 Data quality dimensions by ISO 25012 standard

There are multiple theoretical frameworks describing the dimensions of data quality in IT. However, some of those frameworks proved to be more credible than others over time. One such example is certainly the ISO/IEC 25012 standard.

ISO/IEC 25012:2008 standard describes data across 15 different dimensions which could be broadly arranged in three categories (Guerra-García, C. et al., 2023), (Miller, R. et al., 2024):

I. **Inherent**
II. **Inherent and System Dependent**
III. **System Dependent**

Inherent Data Quality Dimensions are the same regardless of the usage of the data (Guerra-García, C. et al., 2023), (Miller, R. et al., 2024). They assess the quality independent of the system in which the data resides, and the key idea is, if moved to another system, these qualities would remain the same, there are 5 of them and they are as follows: (1) accuracy, (2) completeness, (3) consistency, (4) credibility and (5) currentness.

Inherent and System-Dependent Data Quality Dimensions, which means they are hybrid of both inherent to a certain degree as well as partially dependent on the system they are used in (Guerra-García, C. et al., 2023), (Miller, R. et al., 2024). Their quality is determined by how well the data and the system work in cohesion: (6) accessibility, (7) compliance, (8) confidentiality, (9) efficiency, (10) precision, (11) traceability and (12) understandability.

System-Dependent Data Quality Dimensions, which means that they are different based on the nature and setting they are being used in (Guerra-García, C. et al., 2023), (Miller, R. et al., 2024). Given the different technical environments, they would always be the

same as before, these qualities change if the system changes even if the data stays the same: (13) availability, (14) portability and (15) recoverability.

Wang & Strong (1996) define quality dimensions as "a set of data quality attributes that represent a single aspect or construct of data quality". Data quality dimensions provide a means to quantify and manage data quality. When defining data quality measures, one should try to focus on dimensions that are meaningful and relevant to the business, with maximum return on investment. On the other hand, measuring all the different dimensions of data quality provides a complete picture. This is the motivation to go in more detail for the five core dimensions of data quality listed in the next subchapter.

3.2 Five core dimensions of data quality

The analysis of many studies showed that **completeness, accuracy, timeliness, consistency and relevance** are the first five dimensions of data quality that are most often mentioned in studies, so they will be described in more detail just below (Wang, R. Y. & Strong, D. M., 1996).

3.2.1 Completeness

Completeness is the most basic dimension in the family of data quality dimensions. Completeness is a measure of the presence or absence of data. In a relational database, "present data" means non-empty values in a data field in a table; "absent data" means null or empty values in a data field in a table. Sometimes values such as "unknown" or "not applicable" are also used to represent missing data (Mahanti, R., 2018). Typical questions that need to be asked are (Mahanti, R., 2018):

1. Is all the necessary information available?
2. Are critical values missing from the data records?
3. Are all data sets recorded?
4. Are all mandatory data recorded?

In some cases, missing data or information is irrelevant. However, in cases where missing data or information is critical to a particular business process or task being performed, completeness becomes a concern (Mahanti, R., 2018). Missing optional data is fine for data completeness. For example, a customer's middle name is optional, so the record can be considered complete even if the middle name is not available.

The percentage of missing/present data can be measured not only at the data item/element level but also at the record, dataset or database level. Missing records can have a huge impact, e.g. lost opportunities, additional costs, customer dissatisfaction, etc.

3.2.2 Accuracy

Accuracy generally means that the recorded value is consistent with a real-world fact or value. Accuracy refers to the absence of error and is considered by consumers to be the most important characteristic of data quality (Fisher, C. W., & Kingma, B. R., 2001). According to Wang & Strong (1996.) accuracy refers to the degree to which the data accurately represent the actual value or condition they describe. This includes the match between the data and the real world, that is, how similar the data is to the real entities they represent (Wang, R. Y. & Strong, D. M., 1996).

3.2.3 Timeliness

Timeliness refers to how up to date the data is for the task being performed. Data timeliness can be expressed as a function of: (1) how current the data is for the task for which it is being used, and (2) whether the data is available in time for use (Wang, R. Y. & Strong, D. M., 1996). According to Ballou and Pazer (1985), timeliness refers to whether information is available when it is needed. The value of timeliness decreases with age or as information changes due to new discoveries (Ballou, D. P., & Pazer, H. L., 1985). Data that does not comply with timeliness is often the result of processing delays or insufficiently frequent updates (Umar, A. et al. 1999). Outdated or delayed data can lead to inaccurate analyses or missed opportunities. For example, in the financial sector, timeliness is crucial when trading stocks. If an investor makes decisions based on stock price data that are not up to date, there is a risk of buying or selling at unfavorable conditions, which can result in financial losses.

3.2.4 Consistency

Consistency, by ISO 8000-8:2015 is a property of data that describes the degree to which the data is free of contradictions and consistent with other data in a specific context of use. Data consistency assesses the absence of obvious contradictions and discrepancies in the data set and the extent to which the data conforms to defined business rules, formats, and domains (Umar, A. et al. 1999). Consistency refers to the absence of contradictions in the data. For example, data that a customer is 20 years old but also has historical transactions recorded 25 years ago represents an inconsistency (Ballou, D. P., & Pazer, H. L., 1985).

Batini and Scannapieco (2006) expand on this concept: Consistency refers to the preservation and satisfaction of semantic rules defined over a set of data. Consistency can be defined in terms of models, constraints, and rules specific to application domains (Batini, C. & Scannapieco, M., 2006). Redman (2001) emphasizes the importance of consistency for business: Inconsistent data can lead to erroneous conclusions, duplicate efforts, and suboptimal decisions. Data consistency is a fundamental prerequisite for the integrity of an information system and the reliability of business processes (Redman, T. C., 2001).

3.2.5 Relevance

Relevance is a dimension of data quality that refers to the usefulness of the data for a specific task. It describes how well the information meets the needs of the user in a specific context of use (Kahn, B. K. et al., 2002). The ISO/IEC 25012:2008 standard defines relevance as the degree to which data has attributes that are appropriate and provide added value for a specific task in a specific context of use.

Relevance refers to the degree to which data is appropriate and useful for the specific task or decision at hand. It ensures that only the data that directly contributes to the objectives of a business process or analytical activity is collected, maintained, and used. High-quality data must not only be accurate and complete but also tailored to meet the needs of its intended users. Irrelevant data increases noise, storage costs, and processing time, and may even lead to flawed decision-making if it distracts from key indicators.

These five dimensions work synergistically: a deficiency in any one of them significantly reduces the overall quality of data. For example, complete but inaccurate data will lead to erroneous conclusions, accurate but outdated data may be useless for current decisions, timely but inconsistent data creates distrust, while consistent but irrelevant data wastes resources without creating value.

These dimensions also form the basis for most data quality assessment methodologies and standards such as ISO 8000 and DAMA-DMBOK, further confirming their central importance in data quality management.

4 Empirical Analysis of Data Quality Challenges

The foundation for the empirical analysis of the data quality challenges discussed in this academic paper is firsthand experience in the industry. Through this chapter there will be real observed examples of data quality issues, their causes and consequences and the way they were both discovered and handled in each specific scenario. We will try to present the wide array of different and unique challenges relating to the theoretical concepts described in the previous sections to connect the importance of understanding the theory and to be able to apply it accordingly.

The project was about data analysis for the purposes of fraud detection modelling for a big company within the SAS (Statistical Analysis System) environment. The analysis was carried out through SQL queries over DWH (data warehouse) (Oracle database) where two layers were covered: (1) The Dimensional Layer – in which the data model used for internal reporting was already built and (2) The Stage Layer – in which the original tables were mapped 1 to 1 due to the need for additional data that is not in the dimensional model.

4.1 Example 1: Inconsistent labelling (referencing) of the same entity (document) (violated dimension: consistency)

In one source system, a document is uniquely defined by a combination of four columns, while in the second system a unique ID is used, without the availability of all four columns that make up the key in the first system. This difference prevents simple linking of records between systems. It was detected by analysis through SQL queries on tables from both systems. Due to the impossibility of linking, it was decided to use data from only one system, which has more complete and significant information for the project.

Apart from the violation of the **consistency** (differing representations of the same entity across systems), this problem also directly affected other important dimensions. This is most usually the case since rarely the specific problem violates only one dimension because of their close interconnection. Other violations from the five core dimensions include violating: **completeness** (the second system lacks the four columns used in the composite key in the first system) and **relevance** (if the second system lacks detailed or contextual, then some of it may not be sufficiently relevant to the analytical needs). There are also multiple other dimensions violated from the ISO/IEC 25012:2008 standard as well.

4.2 Example 2: Different practice of recording data on entities (violated dimension: relevance)

We have 2 entities: (1) individuals who have acquired a certain right and (2) individuals who are using that right (users). Some data, although for the user of the right, are entered under the ID of the person who has acquired the right, and some data are entered under the ID of the person who is the user, so for each data item/table, it is necessary to separately check with the system users under which ID the data is entered. It was detected through the analysis of data obtained by SQL queries and those displayed in the application where there was a discrepancy, and through communication with the system users, an inconsistency was observed. Through communication with the system users, it was defined for which data item/table which ID is viewed.

In this example, apart from the violation of the **relevance** dimension (the data is not reliably associated with the correct person, reducing its usefulness in this context), **consistency** (the same type of data is recorded under different entity IDs across tables or cases) and **accuracy** (data may be technically correct in content (e.g., a date or status), but linked to the wrong person) were also violated as well as multiple dimensions from ISO/IEC 25012:2008 data quality standard.

4.3 Example 3: Differences between application view and DWH model – data inconsistency (violated dimension: accuracy)

The entitlement start date differs between the application and the DWH model, as DWH does not include an additional table of changes. The date calculation logic is based on a combination of multiple sources that the DWH model does not fully cover. The differences were observed through the analysis of the obtained data and the presentation in the application part. Due to the limitations of the DWH dimensional model (which is not under our control), it was decided to abandon the use of the DWH model for that part and use the stage part, which is a copy of the original table from the operating systems, and a clear rule for determining the mentioned date was defined with the end users.

Violated dimensions include **accuracy** (the DWH contains an incorrect or misleading start date), **completeness** (the DWH is missing necessary data (the change-tracking table)) and **consistency** (the same business concept appears with different values in the application and the DWH) as well as multiple dimensions from ISO/IEC 25012:2008 data quality standard.

4.4 Example 4: Missing data within the DWH system (violated dimension: completeness)

Continuing the previous problem where some dates are calculated from multiple entered dates, in addition to the insufficiently detailed published rule, there is also the problem that not all dates that should be considered exist within the DWH system. One such example was for the service start date where information on the service resignation date was missing. Without this information, it can be mistakenly considered that the service has started. The differences were observed through the analysis of the obtained data and the display in the application part. The analysis with the system users determined that the scope of the missing data is relatively small and does not significantly affect the results, so it was decided to continue the implementation without this data, with a documented assumption.

Violated dimensions include **completeness** (not all necessary data is present in the DWH (e.g., deletion records), which are critical for correct interpretation and processing), **accuracy** (due to the absence of key data, derived values (such as "start of service") may be factually incorrect, even though technically valid within the limited data scope) and **consistency** (discrepancy between data available and presented in the application and the DWH with different output due to missing records) as well as multiple dimensions from ISO/IEC 25012:2008 data quality standard.

4.5 Example 5: Data unreliability (violated dimension: accuracy)

The analysis determined that the data indicating a certain user characteristic was not aligned with the state in the application. For some people it was correct, but for others it was not. The error occurred due to a historical problem during data migration, and it is not possible to precisely determine which data and time periods were affected. Due to the unknown scope and the impossibility of correction within the deadlines defined by the project scope, the data was excluded from further analysis and modelling. The decision was documented and confirmed with the system users. The data was detected mainly through the analysis of data obtained through SQL queries and comparison with the state in the application and resolved by validation with the system users and documentation of business rules and exceptions.

Violated dimensions include **accuracy** (the data does not correctly reflect the real-world situation, especially for a subset of users), **completeness** (due to the unknown scope of the issue and missing context (e.g., historical change logs), the dataset is incomplete) and **consistency** (the same attribute has conflicting values between the DWH and the source application, showing inconsistent behaviour across systems) as well as multiple dimensions from ISO/IEC 25012:2008 data quality standard. It is, also important to note that in this example another violated, and very important attribute is credibility. It has been compromised because users lose trust in the data due to known inconsistencies and the lack of a way to trace or fix them.

4.6 Additional examples of data quality challenges in distributed information systems

The full list of all the possible challenges divided into a standardised guidebook would consist of countless examples.

However, here are some additional examples of data quality challenges: (1) Inconsistent User Identifiers Across Systems (primary violation: **consistency**); (2) Inconsistent User Attribute Data (e.g., Address) (primary violation: **accuracy**); (3) Unsynchronized Timing of Data Updates (primary violation: **timeliness**); (4) Inconsistent User Status Across Systems (primary violation: **consistency**) and many more.

To conclude this chapter, the mistakes in the data are omnipresent and unavoidable. One of the most important questions is how to deal with the newly recognized problem. From just a few real-life cases described above it is obvious that there are many ways of handling these kinds of situations and sometimes it is difficult to choose the best one. Nonetheless, the best way to be prepared for such non-program situations is the deep knowledge about the data quality frameworks

and standards to be able to accurately decide what amount of data loss is acceptable etc.

5 Discussion

Identified problems in data quality point to several causes, both technical and organizational. Technical causes include: **(1) Outdated IT infrastructure and disconnection between application modules** – certain data is in one application, others in another and are not connected, and even after bringing all the data into the DWH system, connection sometimes isn't possible. **(2) Historical migration errors** – incomplete control over filling processes and insufficient documentation of the transfer and control process itself makes it impossible to determine the correctness of the data.

Whilst organizational causes include: **(1) Lack of clear responsibility** – it is not clearly defined who is responsible for the accuracy and maintenance of certain data within the system, so there is often "passing the ball" instead of solving the problem. **(2) Poor coordination between IT and business users** – this leads to the situation that specific rules are not clearly defined and documented, and a lot of time is wasted only on "discovering" rules for an individual data. **(3) External system maintenance** – various external companies maintain individual parts of the system, which makes it more difficult to determine the root of the problem, as communication with each of them is required, and leads to slow corrections.

Data quality greatly affects the reliability of reporting systems. Incomplete data slows down the decision-making process due to additional checks and increases operational costs. Incomplete or inaccurate data in a fraud detection model can lead to false positives or false negatives, which reduces the efficiency of the system and undermines trust.

To improve data quality in distributed systems, organizations should establish clear ownership of data across departments, coupled with transparent governance policies that define responsibilities and escalation paths.

6 Future Work

There are countless ways to add on this ever-relevant problem in theoretical as well as pragmatical approach. Possible future discussions over the topic could encompass the quantitative case studies showing how frequent are specific problems in data quality across different distributed information systems. This approach would involve usage of metrics and statistical analysis to unbiasedly access individual challenges.

Another useful approach lies in the be development of the software for automatic inconsistency detection. This software implies creating tools that can identify data conflicts, violations of rules etc. without manual intervention. This is particularly important in environments where data is collected, stored, and processed across multiple platforms or databases because in distributed environments it is that much harder to revert and nullify the errors made in the previous phases of, for example, the ETL process.

Other useful improvements are standardization across systems in the means of creating and implementing interoperable standards and schemas to ensure consistent data formats and data definitions across different systems and organizations. Decentralized data governance models with the emphasis on how can blockchain or other distributed ledgers technologies we used to support trust, accountability, and auditability in data quality management.

7 Conclusion

Challenges with data quality to meet the appropriate standard are not anything new or previously undiscussed, they are present and will be present as long as the data itself is. However, in the data-driven world we know today, with the ever-growing need for the new data the problem is magnified exponentially if not addressed correctly (Duarte, F., 2025). That is why it is extremely important to implement the right measures and industry best practices from the very beginning. The standards such as ISO/IEC 25012 and others play a significant role in mitigating those challenges but, the standards, without the theoretical background for comprehending them fully, fall short in the real-life project tangles.

As seen countless times throughout the past not being on top of your data and managing its quality optimally, often, results in huge money, reputation and other losses (Redman, T. C., 2016). That is why any company with an aspiration to become or stay in a position of competitive player on the market should make data quality governance its priority. In other words, data quality has become the cornerstone of a profitable business in a digital and very much globalised world of 21st century. It makes or breaks it on its way to success.

References

Amatriain, X., & Basilico, J. (2015). Recommender systems in industry: A Netflix case study. In F. Ricci, L. Rokach, & B. Shapira (Eds.), Recommender systems handbook (pp. 385–419). Springer. https://doi.org/10.1007/978-1-4899-7637-6_11

Ballou, D. P., & Pazer, H. L. (1985). Modeling data and process quality in multi-input, multi-output

information systems. Management Science, 31(2), 150–162. https://doi.org/10.1287/mnsc.31.2.150

Ballou, D. P., & Tayi, G. K. (1989). Methodology for allocating resources for data quality enhancement. Communications of the ACM, 32(3), 320–329. https://doi.org/10.1145/62065.62071

Batini, C., & Scannapieco, M. (2006). Data quality: Concepts, methodologies and techniques. Springer. https://doi.org/10.1007/3-540-33173-5

Duarte, F. (2025, April 24). Amount of data created daily (2025). Exploding Topics. https://explodingtopics.com/blog/data-generated-per-day

Fisher, C. W., & Kingma, B. R. (2001). Criticality of data quality as exemplified in two disasters. Information & Management, 39(2), 109–116. https://doi.org/10.1016/S0378-7206(01)00083-0

Gualo, F., Rodriguez, M., Verdugo, J., Caballero, I., & Piattini, M. (2021). Data quality certification using ISO/IEC 25012: Industrial experiences. Journal of Systems and Software, 176, 110938. https://doi.org/10.1016/j.jss.2021.110938

Guerra-García, C., Nikiforova, A., Jiménez, S., Perez-Gonzalez, H. G., Ramírez-Torres, M., & Ontañon-García, L. (2023). ISO/IEC 25012-based methodology for managing data quality requirements in the development of information systems: Towards data quality by design. Data & Knowledge Engineering, 145, 102152. https://doi.org/10.1016/j.datak.2023.102152

International Organization for Standardization. (2008). ISO/IEC 25012:2008 — Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model [Standard]. ISO.

International Organization for Standardization. (2015). ISO/IEC 25024:2015 — Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality [Standard]. ISO.

International Organization for Standardization. (2015). ISO 8000-8:2015 — Data quality — Part 8: Information and data quality — Concepts and measuring [Standard]. ISO.

International Organization for Standardization & International Electrotechnical Commission. (2023). ISO/IEC 25010:2023 (E) — Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Product quality model (2nd ed.) [Standard]. ISO/IEC.

International Organization for Standardization & International Electrotechnical Commission. (2024). ISO/IEC 25040:2024 — Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality evaluation framework (2nd ed.) [Standard]. ISO/IEC.

Kahn, B. K., Strong, D. M., & Wang, R. Y. (2002). Information quality benchmarks: Product and service performance. Communications of the ACM, 45(4), 184–192. https://doi.org/10.1145/505248.506007

Kahn, M. G., Raebel, M. A., Glanz, J. M., Riedlinger, K., Steiner, J. F., & Hamilton-Lopez, M. (2015). Transparent reporting of data quality in distributed data networks. eGEMS, 3(1), Article 1052. https://doi.org/10.13063/2327-9214.1052

Llano, M. (2022, October 2). American Express: Using big data to prevent fraud. Harvard Business School Digital Initiative. https://d3.harvard.edu/platform-digit/submission/american-express-using-big-data-to-prevent-fraud/

Mahanti, R. (2018). Data governance implementation: Critical success factors. Software Quality Professional, 20(4), 4–16.

Margara, A., Cugola, G., Felicioni, N., & Cilloni, S. (2022). A model and survey of distributed data-intensive systems. arXiv. https://arxiv.org/abs/2203.10836

Miller, R., Whelan, H., Chrubasik, M., Whittaker, D., Duncan, P., & Gregório, J. (2024). A framework for current and new data quality dimensions: An overview. Data, 9(12), 151. https://doi.org/10.3390/data9120151

Redman, T. C. (2001). Data quality: The field guide. Digital Press.

Redman, T. C. (2016, September 22). Bad data costs the U.S. $3 trillion per year. Harvard Business Review. https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year

Umar, A., Karabatis, G., Ness, L., Horowitz, B., & Elmagarmid, A. (1999). Enterprise data quality: A pragmatic approach. Information Systems Frontiers, 1(3), 279–301. https://doi.org/10.1023/A:1010006529488

Wang, R. Y. (1998). A product perspective on total data quality management. Communications of the ACM, 41(2), 58–65. https://doi.org/10.1145/269012.269022

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, 12(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

Woodall, P., Parlikad, A. K., & Lebrun, L. (2012). Approaches to information quality management: State of the practice of UK asset-intensive organisations. In J. Mathew, L. Ma, A. K. Parlikad, & M. A. Tiwari (Eds.), Asset condition, information systems and decision models (pp. 1–18). Springer. https://doi.org/10.1007/978-1-4471-4392-5_1

A Case Study on the Two-Sided Version of the Node of Knowledge Conceptual Framework

Martina Asenbrener Katic, Sanja Candrlic
University of Rijeka
Faculty of Informatics and Digital Technologies
Radmile Matejcic 2, 51000 Rijeka, Croatia
{masenbrener, sanjac}@inf.uniri.hr

Marina Rauker Koch
University of Applied Sciences of Rijeka
Vukovarska 58, 51000 Rijeka, Croatia
mrauker@veleri.hr

Abstract. *Today, data plays a central role in supporting business decision-making, planning and operations. Most business information systems rely on relational databases to store and manage structured data. Much of the valuable information contained in reports, documents or natural language statements cannot be directly integrated into relational databases due to their rigid structure and predefined schemas. This limitation requires new approaches that can bridge the gap between unstructured text and structured data storage. One promising solution lies in the conceptual framework of Node of Knowledge (NOK), which provides a structured method for translating natural language knowledge into formats suitable for storage and retrieval in relational databases.*

This paper presents a system for converting knowledge expressed in natural language into a relational database format. The transformation process is based on the NOK method, in particular its graphical representation known as Diagram Node of knowledge. For this purpose, a metamodel was developed that maps concepts from NOK diagrams to the concepts used in entity-relationship diagrams and thus enables a direct translation of DNOK structures into relational database entries.

A case study is used to show how knowledge from simple natural language sentences can be entered into a relational database.

Keywords. Knowledge representation, Node of Knowledge, NOK, DNOK, relational database

1 Introduction and related work

In business information systems, data is stored in relational databases. To date, various technologies have been developed to improve working with databases. There are many additional systems and tools of business intelligence, e.g. data warehousing tools (Subotic et al., 2014), (Ahmadi, 2023), which can improve the conclusion about data in the information system. However, there is a problem with storing information embedded in unstructured or semi-structured texts (Pavlic, Jakupovic, et al., 2013). Only the data that is structured as prescribed by a database schema created in the design process according to a methodology of information system development is entered into databases. This means that only data that fulfils certain criteria and for which a number of restrictions apply (type of values, amounts, complex terms and similar) can be entered into a database. On the other hand, there is a great need for systems that enable the storage of textual knowledge (knowledge expressed in text form) in databases.

However, if we want to store the new knowledge in the form of natural language sentences, the question arises as to how natural language sentences can be entered into the relational database. The problem is how to enter knowledge from texts in books and natural language sentences into a relational database that allows us to perform semantic queries. If the sentence is entered into the text field, similar to entering text into editors, then it is not possible to perform the query over sentences, as is the case with the other data in the database. One of the methods that makes this possible is the Node of Knowledge (NOK) method, which is part of the Node of Knowledge conceptual framework

Large language models (LLMs) have recently dominated the field of natural language processing and question answering systems. Unlike knowledge bases, which involve large amounts of data, LLMs show a strong ability to generalise across a wide range of textual, tabular, general and even mathematical tasks to answer questions with few examples in context (Li et al., 2023). Currently, the best known and most widely used LLM is the GPT developed by OpenAI (ChatGPT, 2025). Large language models look for answers to queries that require a large amount of knowledge, which is why the accuracy of the answers obtained is questioned (Lewis et al., 2020), (Romano et al., 2024), (Athaluri et al., 2023). Large language models use a large amount of unstructured knowledge to search for answers to questions, while the NOK conceptual framework uses structured knowledge. This makes the NOK method suitable for solving problems that require structured knowledge and a precisely selected set of knowledge in which answers to questions are sought.

The conceptual framework Node of Knowledge consists of a Node of Knowledge method and several formalisms (formalism for graphical representation – DNOK – diagram node of knowledge, formalized textual record – FNOK – formalized node of knowledge, formalized questions – qFNOK – question formalized node of knowledge).

The NOK method represents knowledge graphically in the form of diagrams of Nodes of Knowledge (Asenbrener Katic et al., 2015), (Rauker Koch et al., 2014), (Tomljanovic et al., 2014). The DNOK formalism (as well as other graphical methods of knowledge representation) shows knowledge in the form of a graph consisting of nodes and links. The nodes represent concepts and the links between the nodes are used to represent the relationships between the concepts. When comparing the NOK method with other graphical methods (Basic Conceptual Graphs (BG), Multi-layered extended semantic networks (MULTINET), Hierarchical Semantic Form (HSF), Resource Description Framework (RDF), it can be concluded that the NOK method is simpler than other methods (it has a smaller number of elements), more expressive (it allows the representation of knowledge at different levels of abstraction) and easier to read (in the NOK method it is possible to read the knowledge starting from any node, but using roles of relations) (Jakupovic et al., 2013). These characteristics of the NOK method enable easy deployment and thus broad applicability and are one of the main reasons why the NOK method was chosen.

The conceptual framework of NOK (Jakupovic et al., 2014) facilitates the transformation of knowledge from natural language texts into a Question Answering (QA) system by organising it into structured records suitable for integration into a relational database.

Studies (Rauker Koch et al., 2014), (Rauker Koch et al., 2015b), (Rauker Koch et al., 2022) have shown that the NOK conceptual framework enables the encoding of texts in multiple languages (Candrlic et al., 2020), into a formalised format suitable for input into a relational database. To achieve this transformation of natural language sentences into a structured format, it is necessary to analyse word types, define formalisation rules and develop a language metamodel. Previous research has looked at different word classes — nouns (Asenbrener Katic et al., 2021), verbs (Asenbrener Katic et al., 2018), adjectives (Pavlic et al., 2017) and adverbs and prepositions (Asenbrener Katic et al., 2022) and confirmed that the construction of a language metamodel and the encoding of sentences in a formalized record is indeed feasible.

Previous research has used the so-called one-way NOK method, in which the relationship between nodes is analysed in one direction only. In this paper, we present an extended variant of graphical knowledge representation that uses two-sided (or two-way) links, where each link carries two semantic roles. While DNOK also contains one-sided links — where a link has a single role — two-sided links offer greater semantic richness and expressive power. Due to the characteristics of the two-sided NOK method, we believe that a question answering (QA) system based on it would provide better results.

This paper presents the two-sided NOK method, including the DNOK diagrams for the given examples and the corresponding data sets stored in a relational database.

After the introduction and presentation of related work in Section 1, the motivation for the research is presented in Section 2. Section 3 describes the research methodology. Sections 4 and 5 describe the NOK meta-model and the results of the case study. Section 6 provides conclusions and a plan for future work.

2 Research Motivation

The aim of the authors' research is to store knowledge derived from textual sources in a structured knowledge database that enables answers to questions to be generated on the basis of the stored information. The main motivation is to develop a QA system based on a relational database that can process and integrate knowledge from any text, regardless of the language in which it is written.

Previous research has used the so-called one-sided NOK method, in which the relationship between nodes is only analysed in one direction.

The question arises: How can two-role (two-sided) links be transformed into a relational database?

3 Methodology

The research methodology is based on the development of a metamodel for the two-sided NOK method and its representation as a relational database. Ten sentences were selected and transformed into DNOK diagrams. Based on the DNOK, tables were inserted into the database.

The Node of Knowledge method is based on researching and analysing natural human language. It analyses sentences, words and their meaning as well as the order in which words are linked together to form more complex thoughts. It is used to create knowledge models, i.e. to organise a knowledge network via ICT. NOK enables the storage of knowledge in a different way than language and writing, namely with the help of the human mind. The aim of the method is to enable the structuring of a knowledge network contained in a text form in any natural language (Pavlic, Mestrovic, et al., 2013). The basic concepts of the NOK method used in this are presented in Table 1.

Table 1. Basic concepts in the DNOK method

Concept	Symbol
Regular node	Iva
Process node	Drive
Link	Role 1 Role 2

Let us present the concepts of the NOK method by using a simple example. DNOK for sentence "A good Vedran drives a green car." is shown in Fig. 1.

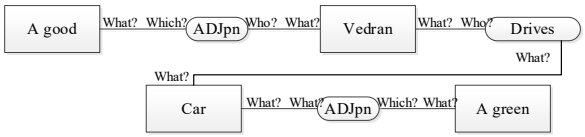


Figure 1. DNOK for sentence "A good Vedran drives a green car."

DNOK for this sentence consists of four regular nodes: "Vedran", "A good", "Car" and "A green" and the process node "Drives". Process nodes "ADJpn" are used on DNOK to connect two regular nodes whose link in a sentence is implied but is not explicitly stated. Nodes "Vedran" and "Car" are connected to the process node "Drives". In addition, the node "Vedran" is connected to the node "A good" with an adjective process node "ADJpn", while the node "Car" is connected to the node "A green" with another adjective process node "ADJpn". The nodes "A good" and "A green" in more detail describe the nodes "Vedran" and "Car" respectively, suggested by the process node "ADJpn". Two questions that define the role are listed on each link between nodes, and they are very important in storing and querying knowledge.

For example, the node "Vedran" and the process node "Drives" are connected by a link with two roles, "What?" and "Who?". If we ask "What Vedran?", The answer is "Drives". If we ask, "Who drives?" the answer is "Vedran". Node "Vedran" is also connected to an adjective process node "ADJpn" with a link with two roles, "What?" And "Who?". If we ask "What Vedran", the answer is "ADJpn". If we ask "Who ADJpn" the answer is "Vedran". The same process node "ADJpn" is on the other side connected by a link with two roles, "What?" and "Which?" with regular node "A good".

If we ask, "What a good?", the answer is "ADJpn". If we ask, "Which ADJpn?" the answer is "A good". In this case, the adjective process node, as previously mentioned, is used to connect two regular nodes, and from this we can conclude that node ADJpn, the adjective process node is used for a more detailed explanation of one of the regular nodes, i.e. node "Vedran" is determined in more detail by node "A good". The mechanism of the NOK method functioning is described in detail in previous papers on NOK method (Asenbrener Katic et al., 2015).

The next chapter presents a brief description of how simple sentences can be transformed into the DNOK model.

4 NOK metamodel

For the requirements of text transformation with the NOK method, it was necessary to define a model of the output database, the so-called NOK text database, into which the input data records are to be transformed and entered after application of the NOK method. For this purpose, a metamodel of the NOK method is defined, which is shown in Fig. 2. The NOK method enables recording natural language sentences in graphical form. By analysing the rules of the NOK method, the model of the NOK method (metamodel) can be represented using the Entity-Relationship method.

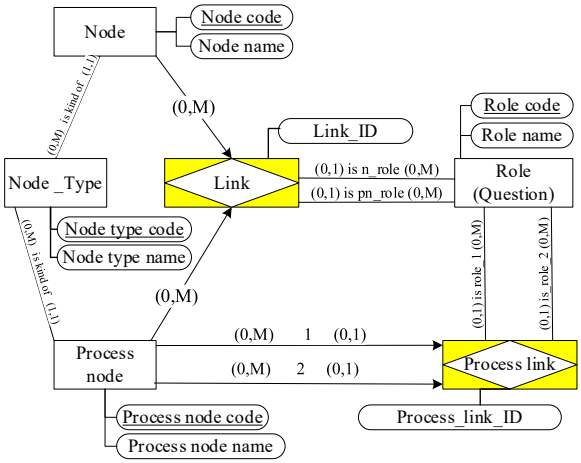


Figure 2. NOK Metamodel

Let us describe the proposed model in more detail. Entity type *Node_Type* contains a list of all possible types of nodes. A single occurrence of entity type *Node_Type* can have none or more occurrences of entity type *Node* and *Process node*. A single occurrence of the entity types *Node* and *Process node* belong to one and only one entity *Node_Type*. Entity types *Node* and *Process node* contain a list of all nodes in the NOK. This version of NOK method consists of two types of nodes, ordinary nodes (called simply nodes) and process nodes.

In NOK, there are links between nodes. A single occurrence of entity type *Node* can be connected to none or many occurrences of entity type *Process node*. A single occurrence of entity type *Process node* can be connected to none or many occurrences of entity type *Node*. Due to this, an aggregated entity type *Link* is introduced for the relationship between nodes, consisting of node and process node.

Each link has a relationship role to the node. The role (question) provides a clearer description of knowledge. These are usually interrogative pronouns and all

interrogative words that can be set. Each link has two roles (questions).

Entity type *Role (Question)* is connected with two relationships to the aggregated entity type *Link*. A single occurrence of entity type *Role* can be a role in none or more occurrences of aggregated entity type *Link*. A single occurrence of the aggregated entity type *Link* has two ends (for example, Node (Role_1) is connected to one, and Process node (Role_2) is connected to the other. The number of roles is (0,1), which means that a role may not have a name from the codebook.

In the case of the need to connect two entity types *Process node*, we come to the aggregated entity type *Process link*. Just like with the aggregated entity type *Link*, the aggregated entity type *Process link* is connected to the entity type *Role* with two relationships. A single occurrence of entity type *Role* can have none or more occurrences of aggregated entity type *Process link.* A single occurrence of aggregated entity type *Process link* belongs to none or only one entity *Role*.

Metamodel of the NOK method enables the transformation and recording of natural language sentences into a relational database using the NOK method.

Once we have entered the knowledge into the relational database, we have the possibility to store and query this knowledge using SQL queries. The metamodel provides enrichment of sentences with questions.

The relational database model (database schema) based on the metamodel NOK (from Fig. 2) is:

NODE_TYPE (Node type code, Node type name)

NODE (Node code, Node name, *Node type code*)

PROCESS NODE (PN code, PN name, *Node type code*)

ROLE (QUESTION) (Role code, Role name)

LINK (Node code, PN Code, Link_ID, *N_Role Code*, *PN_Role Code*)

PROCESS LINK (PN_1 Code, PN_2 Code, Process_link_ID, *Role Code_1, Role Code_2*)

5 Research Results and Discussion (Case Study)

This section will demonstrate how the input natural language sentences are entered into a relational database. According to rules for interpreting entity-relationship diagrams into a relational schema (Pavlić, 2011), based on the metamodel of the NOK method from Fig. 2, a relational database schema of the NOK method for storing text is defined. Algorithm for transformation of DNOK (with one-way link) is given in paper (Pavlic, Mestrovic, et al., 2013). Below you can see the transformation of two-way relationship of DNOK.

Examples of selected sentences in natural language that are transformed into DNOK diagrams are:

1. Peter reads a journal.
2. Mark drives a red car.
3. Mary's car is blue.
4. John drives a motor car.
5. A good Vedran drives a green car.
6. Jelena sits and reads a comic.
7. Tom swims.
8. The student reads a wonderful book.
9. Girls are on the beach.
10. Tom has two cars.

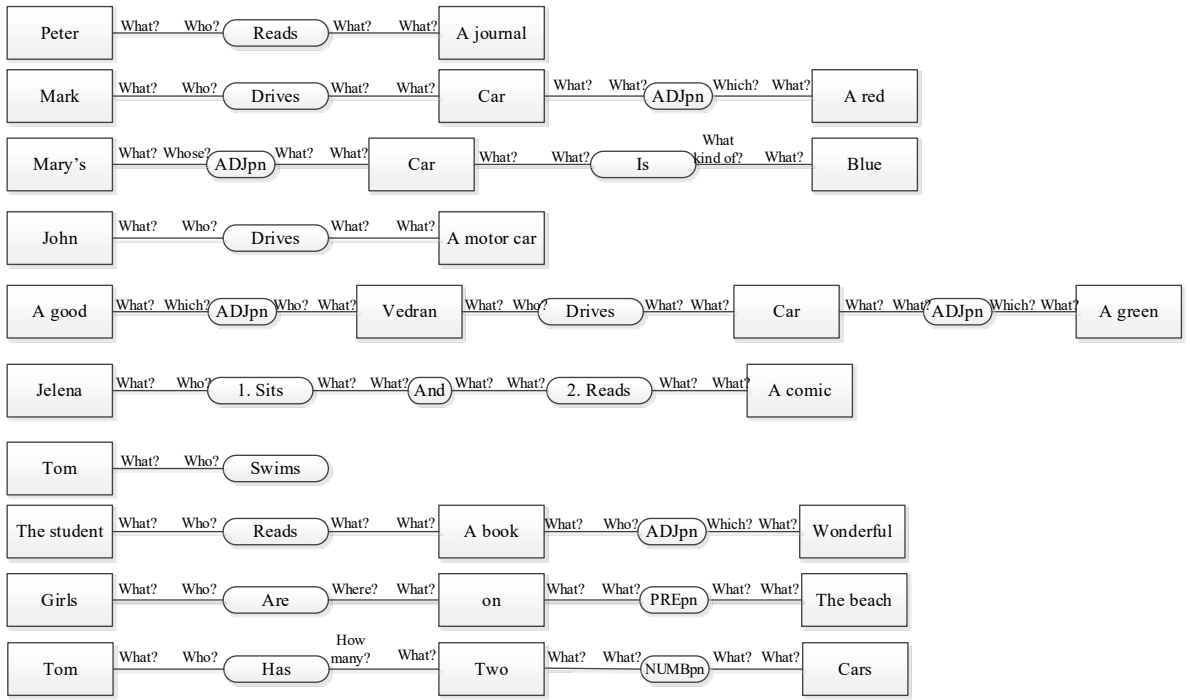


Figure 3. DNOKs for natural language sentences

In DNOK, articles "a", "an" and "the" are joined to words which they preceed (while in the process of querying in SQL queries, those articles are ignored). Each row represents a single sentence.

The corresponding DNOKs for sentences in natural language are shown in Fig. 3.

The selected examples show different types of words and their modeling using the NOK method: nouns, verbs, adjectives, prepositions, conjunctions and numbers. Nouns and Verbs are two basic types of nodes. Adjectives, numbers and prepositions belong to nouns and are connected with special process nodes. Two types of adjectives are shown (definite and indefinite adjective) and their model in DNOK. Conjunctions are shown as process nodes.

Process nodes "ADJpn", "PREpn" and "NUMBpn", i.e. adjectival process node, prepositional process node and numeric process node respectively, in DNOK have no special semantics as words, but represent the class. They are inserted in order to connect two regular nodes, and also serve as a link between nodes that are closely related and enable the preservation of semantics in sentences. In the sentence: Jelena sits and reads a comic, "Sits" and "Reads" are the process nodes, with roles "what?" and "who?", as shown in Fig. 3. The conjunction "And" from this sentence is drawn on DNOK as a process node that connects the two process nodes.

Using the DNOKs from Fig. 3, we can fill in the tables based on the relational database schema (Tables 2-7).

Table 2. NODE_TYPE

Node type code	Node type name
O	Regular node
P	Process node
K	Context node
ADJ	Adjectival process node
PRE	Prepositional process node
NUM	Numeric process node

Table 3. NODE_TYPE

Node code	Node name	Node type code
a	Peter	O
c	A journal	O
d	Mark	O
f	A red	O
g	Car	O
h	Mary's	O
j	Car	O
k	Blue	O
l	John	O
n	A motor car	O
o	A good	O
p	Vedran	O
s	A green	O
t	Car	O
u	Jelena	O
x	A comic	O
aa	Tom	O
ac	The student	O
ae	A book	O
ag	Wonderful	O
ah	Girls	O
aj	On	O
ak	The beach	O
al	Tom	O
an	Two	O
ao	Cars	O

Table 4. PROCESS NODE

PN code	PN name	Node type code
b	Reads	P
e	Drives	P
e_1	ADJpn	ADJ
i	ADJpn	ADJ
i_1	Is	P
m	Drives	P
r	Drives	P
r_1	ADJpn	ADJ
r_2	ADJpn	ADJ
v	Sits	P
z	And	P
y	Reads	P
ab	Swims	P
ad	Reads	P
ad_1	ADJpn	ADJ
ai	Are	P
ai_1	PREpn	PRE
am	Has	P
am_1	NUMBpn	NUM

Table 5. ROLE (QUESTION)

Role code	Role name
I	Who
II.	What
III	Whom
IV	What kind of
V	Where?
VI	How many?
VII	Which?
VIII	Whose?

The first five columns in Table 6 represent the content of relational database. The last two columns *Node Name* and *PN name (Process node name)*, separated by an empty column, were introduced to facilitate monitoring of the transformation of the sentences' content record into the relational database. The same applies to the table 7. Process link.

Although a small number of sentences are involved, the data in Tables 2-7 show how knowledge from simple sentences in natural language can be entered into a relational database that allows knowledge to be queried and stored.

For example, if you ask "Who is driving?" based on the tables, we can get answers: "Marko", "Ivan" and "Vedran". If the question is expanded with another

node, we get more precise answers. For example, "Who drives a motor car?" The answer is "Ivan". To the question "Who is driving the car?" we will get answers, "Marko" and "Vedran", while the question "Who is driving a red car?" results in the answer: "Marko".

In addition, integration with the dictionary is required, allowing the inclusion of various semantic relations. In the case of the example shown, for example, one could ask: "Who drives automobile?". The information about automobile is not added directly to the database. However, according to the data from the monolingual dictionary, there is a connection between the word "car" and its synonym "automobile". Therefore, the answer would be: "Marko" and "Vedran".

The accuracy and concreteness of the answers (one or more) depends on how detailed the question is formulated and how many nodes are used. Sometimes the answer is easy to find in the link and process link tables with one pass through the tables, while sometimes the answer requires several passes through the table, i.e. the complexity of the SQL query can vary.

6 Conclusion and future work

The article describes a system that enables the transformation of textually expressed knowledge into a relational database. The translation of natural language sentences into the relational database is based on the NOK method, i.e. the graphical formalism DNOK. In the paper, a metamodel was defined that enables the

Table 6. LINK

Link_ID	Node code	PN Code	N_Role Code	PN_Role Code		*Node name*	*PN Name*
1.	a	b	II (What)	I. (Who)		*Peter*	*Reads*
2.	c	b	II (What)	II (What)		*A Journal*	*Reads*
3.	d	e	II (What)	I. (Who)		*Mark*	*Drives*
4.	g	e	II (What)	II (What)		*Car*	*Drives*
5.	g	e_1	II (What)	II (What)		*Car*	*ADJpn*
6.	f	e_1	II (What)	VII (Which)		*A red*	*ADJpn*
7.	h	i	II (What)	VIII. (Whose)		*Mary's*	*ADJpn*
8.	k	i	II (What)	II (What)		*Car*	*ADJpn*
9.	k	i_1	II (What)	II (What)		*Car*	*Is*
10.	j	i_1	II (What)	IV (What kind of)		*Blue*	*Is*
11.	l	m	II (What)	I. (Who)		*John*	*Drives*
12.	n	m	II (What)	II (What)		*A motor car*	*Drives*
13.	p	r_1	II (What)	I (Who)		*Vedran*	*ADJpn*
14.	o	r_1	II (What)	VII (Which)		*A good*	*ADJpn*
15.	p	r	II (What)	I. (Who)		*Vedran*	*Drives*
16.	t	r	II (What)	II (What)		*Car*	*Drives*
17.	t	r_2	II (What)	II (What)		*Car*	*ADJpn*
18.	s	r_2	II (What)	VII (Which)		*A green*	*ADJpn*
19	u	v	II (What)	I. (Who)		*Jelena*	*Sits*
20.	x	y	II (What)	II (What)		*A comic*	*Reads*
21.	aa	ab	II (What)	I. (Who)		*Tom*	*Swims*
22.	ac	ad	II (What)	I. (Who)		*The student*	*Reads*
23.	ae	ad	II (What)	II (What)		*A book*	*Reads*
24.	ae	ad_1	II (What)	I. (Who)		*A book*	*ADJpn*
25.	ag	ad_1	II (What)	VII (Which)		*Wonderful*	*ADJpn*
26.	ah	ai	II (What)	I. (Who)		*Girls*	*Are*
27.	aj	ai	II (What)	V (Where)		*On*	*Are*
28.	aj	ai_1	II (What)	II (What)		*On*	*PREpn*
29.	ak	ai_1	II (What)	II (What)		*The beach*	*PREpn*
30.	al	am	II (What)	I. (Who)		*Tom*	*Has*
31.	an	am	II (What)	VI (How many?)		*Two*	*Has*
32.	an	am_1	II (What)	II (What)		*Two*	*NUMBpn*
33.	ao	am_1	II (What)	II (What)		*Cars*	*NUMBpn*

Table 7. PROCESS LINK

Process_link_ID	PN_1 Code	PN_2 Code	Role Code_1	Role Code_2		*PN_1 Name*	*PN_2 Name*
1	v	z	II (What)	II (What)		*Sits*	*And*
2	z	y	II (What)	II (What)		*And*	*Reads*

translation of concepts from NOK diagrams into concepts of entity-relationship diagrams. This has also enabled the direct transformation of data from DNOK into a relational database.

Finally, as a case study, we presented the input of knowledge into a relational database and explained it using the example of simple natural language sentences. In this paper, the NOK method is applied to simple sentences, in later stages we will apply it to complex sentences.

Querying an integrated information system with a new module (NOK system) or defining queries attached to both relational databases (database of the existing information system and database of the new NOK system) makes it possible to obtain answers to the complex question posed, which should be used to scan both text and data. For the query itself, various algorithms for entering data into the database must be defined, which requires great effort and further research.

The plan for further research includes analyzing more dictionary attributes and developing the dictionary model. Without the dictionary model, which uniquely names words, separates homonyms and semantically links synonyms, it is not possible to write algorithms required to develop the system for storing sentences in the relational databases. Our future goal is to develop a module for an information system based on the metamodel, which will enable the input of natural language sentences in two-way NOK records and their retrieval.

Acknowledgments

This work has been fully supported by the University of Rijeka under project uniri-iskusni-drustv-23-216.

References

Ahmadi, S. (2023). Optimizing Data Warehousing Performance through Machine Learning Algorithms in the Cloud. *International Journal of Science and Research*, *12*(12), 1859–1867. https://doi.org/10.21275/SR231224074241

Asenbrener Katic, M., Candrlic, S., & Koch, M. R. (2022). Adverb and Preposition Representation in Croatian and English using the Node of Knowledge Method. *Central European Conference on Information and Intelligent Systems*, 41–48.

Asenbrener Katic, M., Candrlic, S., & Pavlic, M. (2018). Modeling of verbs using the node of knowledge conceptual framework. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings*, 1022–1027. https://doi.org/10.23919/MIPRO.2018.8400187

Asenbrener Katic, M., Candrlic, S., & Pavlic, M. (2021). Nouns in the Conceptual Framework "Node of Knowledge." *Tehnicki Vjesnik - Technical Gazette*, *28*(6), 2088–2093. https://doi.org/10.17559/TV-20200731095329

Asenbrener Katic, M., Pavlic, M., & Candrlic, S. (2015). The representation of database content and structure using the NOK method. *Procedia Engineering*, *100*(January). https://doi.org/10.1016/j.proeng.2015.01.469

Athaluri, S. A., Manthena, S. V., Kesapragada, V. S. R. K. M., Yarlagadda, V., Dave, T., & Duddumpudi, R. T. S. (2023). Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus*, *15*(4). https://doi.org/10.7759/CUREUS.37432,

Candrlic, S., Asenbrener Katic, M., & Jakupovic, A. (2020). Preliminary multi-lingual evaluation of a question answering system based on the node of knowledge method. In *Lecture Notes in Networks and Systems* (Vol. 69). https://doi.org/10.1007/978-3-030-12388-8_69

ChatGPT. (2025). *ChatGPT*. https://chatgpt.com/

Jakupovic, A., Pavlic, M., & Dovedan Han, Z. (2014). Formalisation method for the text expressed knowledge. *Expert Systems with Applications*, *41*(11), 5308–5322. https://doi.org/10.1016/j.eswa.2014.03.006

Jakupovic, A., Pavlic, M., Mestrovic, A., & Jovanović, V. (2013). Comparison of the Nodes of Knowledge method with other graphical methods for knowledge representation. *Proceedings of the 36th International Convention /CIS/, MIPRO 2013*, 1276–1280.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W. T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, *2020-December*. https://arxiv.org/pdf/2005.11401

Li, T., Ma, X., Zhuang, A., Gu, Y., Su, Y., & Chen, W. (2023). Few-shot In-context Learning for Knowledge Base Question Answering. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, *1*, 6966–6980. https://github.com/ltl3A87/KB-BINDER

Pavlic, M., Dovedan Han, Z., Jakupovic, A., Asenbrener Katic, M., & Candrlic, S. (2017). Adjective representation with the method Nodes of Knowledge. *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1221–1226. https://doi.org/10.23919/MIPRO.2017.7973610

Pavlic, M., Jakupovic, A., & Meštrović, A. (2013). Nodes of knowledge method for knowledge representation. *Informatologia*, *46*(3), 206–214.

Pavlic, M., Mestrovic, A., & Jakupovic, A. (2013). Graph-based formalisms for knowledge representation. *Proceedings of the 17th World Multi- Conference on Systemic Cybernetics and Informatics (WMSCI 2013)*, 200–204.

Rauker Koch, M., Čandrlić, S., & Ašenbrener Katić, M. (2022). Automation of the conversion of natural language to formalized node of knowledge record. *Zbornik Veleučilišta u Rijeci*, *10*(1), 57–71. https://doi.org/10.31784/zvr.10.1.4

Rauker Koch, M., Pavlic, M., & Asenbrener Katic, M. (2015). Homonyms and Synonyms in NOK Method. *Procedia Engineering*, *100*, 1055–1061. https://doi.org/10.1016/j.proeng.2015.01.466

Rauker Koch, M., Pavlic, M., & Jakupovic, A. (2014). Application of the NOK method in sentence modelling. *Proceedings of the 37th Internation Convention MIPRO 2014*, 1426–1431.

Romano, S., Angius, R., Kerby, N., Bouchaud, P., Amidei, J., & Kaltenbrunner, A. (2024). A Dataset to Assess Microsoft Copilot Answers in the Context of Swiss, Bavarian and Hessian Elections. *Proceedings of the International AAAI Conference on Web and Social Media*, *18*, 2040–2050. https://doi.org/10.1609/ICWSM.V18I1.31446

Subotic, D., Jovanovic, V., & Poscic, P. (2014). Data Warehouse Schema Evolution: State of the Art. *Central European Conference on Information and Intelligent Systems CECIIS*.

Tomljanovic, J., Pavlic, M., & Asenbrener Katic, M. (2014). Intelligent question — Answering systems: Review of research. *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1228–1233. https://doi.org/10.1109/MIPRO.2014.6859755

Parallel Utilization of SQL and NoSQL Databases – A Systematic Literature Review

Nađa Klještanović, Sonja Ristić, Darko Stefanović
University of Novi Sad
Faculty of Technical Sciences
Department of Industrial Engineering and Engineering Management
Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia
{nadja.kljestanovic, sdristic, darko.stefanovic}@uns.ac.rs

Abstract. *Complex information systems often face challenges when choosing an appropriate database for storing business logic data. SQL and NoSQL systems represent standard approaches with clearly defined areas of application. However, systems that lie at the intersection of these approaches encounter difficulties, as it is often necessary to take advantage of both models. The modern trend of polyglot persistence enables combination of multiple data models within a single system. This paper conducts a systematic literature review of the concept of polyglot persistence, as well as multi-model databases, providing a clear overview of the advantages, limitations and use cases of these approaches in various business and technical environments.*

Keywords. SQL, NoSQL, ACID, BASE, polyglot persistence, multi-model databases

1 Introduction

Many applications nowadays are data-intensive. Data ecosystem evolves in several dimensions like volume, variety, complexity and velocity. Traditional relational database management systems (RDBMS) with their Structured Query Language (SQL) enable efficient data organization, access and manipulation. SQL allows users to retrieve, read and modify data in the database in a simple way, providing intricate query patterns (Majhadi & Machkour, 2021). In that way relational databases (often referred as SQL databases) are suitable for analytics and report systems where data integrity is paramount.

Although RDBMSs have many advantages they have several limitations. They are primarily aimed for processing structured data. Working with semi-structured and unstructured data requires more complex customization and that can slow down system efficiency. Besides, RDBMSs are designed for vertical scaling, which involves upgrading existing server resources to increase database capacity. In addition, working with these systems requires a defined database schema, which can be a challenge in dynamic environments with rapidly changing data structures. The adherence to ACID (atomicity, consistency, isolation, durability) properties in transaction management slows down performance when dealing with large amounts of data. These strict properties require complex synchronization mechanisms. Finally, providing high system availability is a challenge because traditional relational systems often require complex data replication and failure recovery strategies (Ohlsson & Persson, 2019).

An umbrella term *NoSQL databases* encompasses a variety of data systems that do not adhere to relational data model and have different approaches to data storage and retrieval. Although the first models of this type of databases were created in the 1960s, their wider utilization and popularization began only after 2007, when the company Amazon presented its NoSQL database DynamoDB. Compared to relational databases, NoSQL databases offer a lot of advantages. They provide high performance in terms of data processing speed and capacity during storing large amounts of data. These advantages often come with a certain compromise in relation to ACID properties. NoSQL databases very often adhere BASE (Basically Available, Soft State, Eventual Consistency) instead of ACID properties. (Ohlsson & Persson, 2019).

Although NoSQL databases offer many advantages, like scalability, flexibility and high performance, they also have some challenges (Nurhadi et al., 2024). Adhering BASE properties can lead to temporary data inconsistency and problems with integrity of stored data. Also, NoSQL databases are not designed for complex queries and data transactions like SQL databases. That can make difficult analysing data or working with relations between entities. In addition, NoSQL databases are often schema-less, unlike SQL databases. Because of that, determining an appropriate data structure requires careful system design and understanding. The deficiency of standardization

among different types of NoSQL databases can make transfer across data management systems or integration into existing data systems difficult.

The practice of using both NoSQL and SQL databases within a single application is increasingly common, likewise using one system providing characteristics of both database types. These approaches are applied especially in complex architectures that require a combination of flexibility and scalability provided by NoSQL and the structure and integrity of data ensured by means of SQL databases. First approach, known as polyglot persistence, involves the use of different types of databases. Second approach, known as the multi-model database approach, implies the existence of just one database with SQL and NoSQL properties. Both approaches enable organizations to optimize system performance according to specific requirements, thereby ensuring efficient handling of different types of data, scalability, flexibility and integrity (Glake et al., 2022).

The goal of this research is to conduct a systematic literature review (SLR) to analyse and compare different types of NoSQL databases in combination with the relational database, addressing the lack of comprehensive studies on how these combinations are applied in the same projects. The result is a systematic overview of the advantages and disadvantages of various SQL/NoSQL combinations with special emphasis on their use in several use cases within the reviewed papers.

Following Introduction, this paper proceeds as follows. The second chapter includes an overview of the research planning, the method of its implementation and the analysis of the selected papers using descriptive statistical analysis. The third chapter presents the discussion about the research with answers to the research questions, while the fourth and final chapter are based on presenting opinions on the topic and potential directions for future work.

2 Systematic Literature Review

According to Barbara Kitcheman (2004), SLR consists of three key phases which are planning, conducting the review and reporting the review.

As part of this study, scientific papers dealing with applications that simultaneously utilize SQL and NoSQL databases were analysed. Special emphasis was placed on the overview of database systems, as well as their applications across different domains.

2.1 Planning the review

At the beginning of the first phase of the research, a search for existing papers on the same or similar topic was conducted. Several SRLs were found, but they are different from this paper in several aspects. First, they focus just on the analysis of the concept of polyglot persistence or multi-model databases. Second, they analyze only one specific combination of SQL and NoSQL databases. Finally, they do not provide a comprehensive overview of the various concepts and possible ways of their utilization.

The next step in planning an SLR is defining its objective. The aim of this research is to compare use cases involving different combinations of SQL and NoSQL databases. A comparative analysis presents commonly used combinations of these database systems. Furthermore, the focus is placed on the specific purpose of each presented combination.

2.1.1 Research Questions

The following research questions were formulated from the previously defined objective:

RQ1: What are the characteristics, advantages and limitations of SQL and NoSQL databases?

RQ2: What are polyglot persistence and multi-model approaches and how are they applied in contemporary information systems?

- **RQ2.1**: What are the key benefits of polyglot persistence approach?
- **RQ2.2**: What are the advantages of multi-model approach?

RQ3: What combinations of SQL and NoSQL databases are commonly used?

- **RQ3.1**: In which scenarios is a specific combination of SQL and NoSQL databases utilized?

2.2 Conducting the review

Conducting the review refers to the implementation of research, which includes the identification of relevant studies, their selection and quality assessment, as well as data extraction and summarization of primary research results (Kitchenham, 2004). The conducted research steps are shown in Fig. 1.

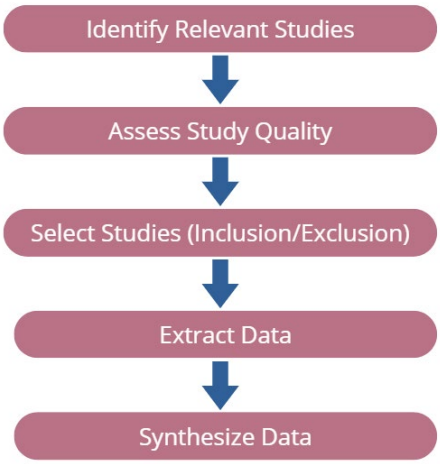


Figure 1. Conducted research steps

2.2.1 Identification of Research

Index databases used to search for relevant sources on the selected topic are Google Scholar and Scopus.

It was necessary to define the search string with different variations of key terms, while conditions related to the year of publication, type of work and language were set to maintain the relevance of the results. The search string for Scopus database is:

("SQL" OR "relational database" OR "relational databases") AND ("NoSQL" OR "non-relational database" OR "non-relational databases") AND ("polyglot persistence" OR "hybrid database" OR "hybrid architecture" OR "multi-model") AND PUBYEAR > 2016 AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar")) AND LIMIT-TO (LANGUAGE , "English").

Since Google Scholar uses a different search syntax compared to Scopus, the following text provides a search string relevant for this index database.
Additionally, it is necessary to apply filters for the publication year and document type after entering the string:

("SQL" OR "relational database" OR "relational databases") AND ("NoSQL" OR "non-relational database" OR "non-relational databases") AND "polyglot persistence" OR "hybrid database" OR "hybrid architecture" OR "multi-model").

2.2.2 Study Selection

Inclusion and exclusion criteria were defined based on the specific needs of the research. Criteria that were not applied during the search itself were implemented during the selection of works obtained through the search. A tabular overview of the inclusion and exclusion criteria is provided in Table 1 and Table 2, respectively.

Table 1. Inclusion criteria

ID	Description
I1	The paper must be written in English.
I2	The paper must be published in 2017 or later.
I3	Only journal articles or conference papers are accepted.
I4	The works must be accessible.
I5	The paper must describe use cases of parallel use of SQL and NoSQL databases in the same application.

Table 2. Exclusion criteria

ID	Description
E1	Duplicates must be deleted.
E2	If there are several works by the same author on the same topic, only one is taken.
E3	Papers dealing only with comparing the use of SQL or NoSQL databases in specific use cases are excluded.
E4	Papers discussing migration between SQL and NoSQL database systems are excluded.

During the further analysis of the selected works, the focus was placed on specific technologies and use cases involving the combination of NoSQL and SQL databases. Additionally, the descriptive statistical analysis considers all previously established criteria to provide a clearer statistical representation of the selected works.

2.2.3 Study Quality Assessment

Quality of study can be assessed like inclusion and exclusion criteria, but with a greater level of detail. This assessment provides that the studies are reliable for research. In order to ensure scientific quality, it is crucial to ask the following questions during evaluation process:

QA1: Were appropriate data collection methods used?
QA2: Are the research questions clearly formulated and relevant to the research area?
QA3: Are the results of the study reproducible under the same experimental conditions?
QA4: Are the used methodologies presented clearly?
QA5: Were the limitations of the study and their potential impact on the results considered?

The quality assessment was conducted on the studies that satisfied the defined inclusion and exclusion criteria. Only studies that fully met all quality assessment criteria were selected for the next phase of the research.

2.2.4 Data Extraction

The data extraction phase was conducted manually by the authors to provide a consistent and systematic analysis of studies, enabling answers to research questions and achieving quality criteria (Kitchenham, 2004). It involves a several groups of extracted information. The first group includes basic details about each study: ID, title, authors, year and place of publication. The second group of extracted data categorize studies according to the type of publication. The third data group classifies studies based on keywords, abstracts and titles. Finally, data related to the technologies used in the studies are extracted. This approach provides an objective analysis of the common utilization of these technologies.

On this basis, the grouping of data is done, which forms the foundation for the presenting data in the next phase, data synthesis.

2.2.5 Data Synthesis

The initial search of both selected index databases resulted in a total of 599 papers. Then the papers were selected based on defined inclusion and exclusion criteria, as well as based on the abstract, content, and the numbers of selected papers are shown in Table 3.

Table 3. Number of selected papers by search phases

	Scopus	Google Schoolar
Initial search	403	196
Number of selected papers based on inclusion, exclusion and quality assessment criteria	65	40
Number of selected papers based on abstract	30	22
Number of selected papers based on content	13	10
In total:	23	

Within this section, various types of data on the collected works are presented using tables and graphs. The presentation of basic information about the selected papers is based on their publication year, source type, content, relevance to the research questions, and the mentioned technologies. The information was extracted from 23 selected papers.

The first graph (Fig. 2) shows the percentage of published papers based on inclusion criteria. The years of publication of the papers range from 2017 to the present. As observed, the highest percentage of selected works was published in 2019 and 2022, accounting for 23.1% of all selected papers.

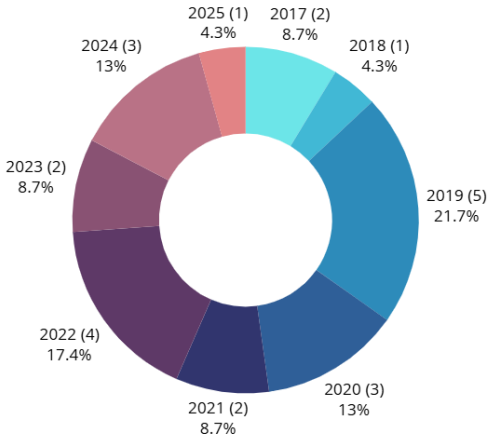


Figure 2. Publications pre year

Table 4 provides insight into the types of data sources, indicating whether the selected papers were published in conferences or scientific journals. The most of selected papers, nearly 77%, are classified as journal articles, while the remaining papers were published in conference proceedings.

Table 4. Source types of research

Source type	References	Percentage	Total
Conference proceeding	12, 14, 20, 21, 23	21.74%	5
Journal article	7, 8, 9, 10, 11, 13, 15, 16, 17, 18, 19, 22, 24, 25, 26, 27, 28, 29	78.26%	18

The subsequent division of selected papers is based on their content, that is, on the main topic of each research paper. Due to the diversity of topics, the papers are divided into three groups. The first group includes works dealing with migration between architectures, like migration from monolith to hybrid architecture as well as the reverse process. The second group consists of papers related to the integration of SQL and NoSQL databases, focusing on the integration of different database types through specific use cases. The third group refers to the papers that present theoretical explanation of polyglot persistence and multi-model databases. The papers are classified according to their primary topic and the results of this analysis are presented in the graph Fig. 3. It can be observed that the largest number of papers provides a theoretical introduction to the concepts of hybrid databases, which are the central focus of this research.

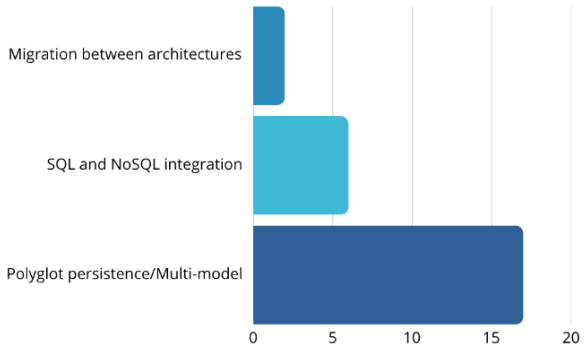


Figure 3. Research contexts

The following table presents the relevance of the selected papers to the research being conducted. Specifically, Table 5 provides an insight into which papers address specific research questions. The first column contains the serial number of the research question, which also refers to all sub-questions listed within that question. The second column lists the serial numbers of the papers that provide answers relevant to a particular research question.

Table 5. Answers to the research questions

Research question	References
RQ1	7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 21, 23, 24, 26, 27, 28, 29
RQ2	7, 9, 10, 11, 12, 13, 14, 15, 16, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
RQ3	7, 8, 10, 13, 15, 16, 17, 18, 20, 22, 24, 25, 26

The last graph (Fig. 4) presents the number of mentions of a particular technology within the selected works, as well as the context in which the technology appears. More precisely, specific types of databases may be mentioned independently as a tool, while others are referenced as part of a hybrid database. Graph presented in Fig. 4 displays the technologies that appear more than three times in the analyzed works, along with the number of occurrences for individual purposes, when the technology is part of the hybrid architecture, as well as the total number of occurrences in the papers. This approach can be useful for analyzing the frequency of technology utilization in research on polyglot persistence and multi-model concepts. The most often mentioned database in both approaches is MongoDB.

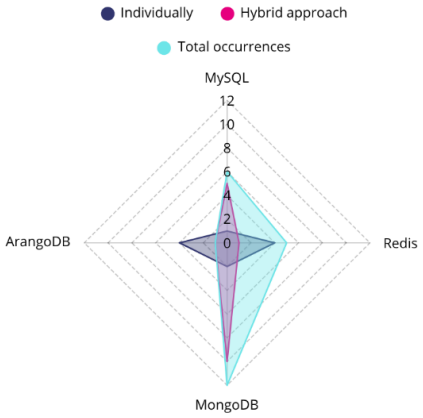


Figure 4. Number of technology occurrences

3 Discussion

The third phase of the SLR focuses on the results of the research. This part includes presenting the results in the form of a technical report. The aim of this paragraph is to answer the previously provided research questions.

RQ1 refers to the advantages and limitations of SQL and NoSQL databases. The important drawback of relational databases, mentioned by Lajam and Mohammed (2022), is the storage of exclusively structured data. Also, they have high costs of processing data and executing queries on tables due to joining tables and checking constraints. In research "*Recent trends in database technology*", (Lieponienė, 2020) concludes that relational databases prioritize data integrity, often at the expense of availability. Strict integrity enforcement prevents easy distribution of data. Data migrations are slower with these databases, but precisely because of the focus on data consistency and integrity, as stated by (Harezlak, Mermon, & Kasprowski, 2021).

In contrast, NoSQL databases have gained traction by addressing many of the limitations inherent in relational systems. Since data is kept in its original form, they offer improved performance and reduced programming costs. Through the study, (Lieponienė, 2020) it is concluded that the priority of NoSQL databases is data handling, not their security. They do not provide the possibility of built-in security mechanisms, but developers are responsible for their implementation in middleware. According to the research (Harezlak, Mermon, & Kasprowski, 2021), the main disadvantage of NoSQL databases is the absence of a standardized query language. Also, NoSQL databases cannot be optimized, which makes their use difficult, especially in a heterogeneous environment, as mentioned by (Kumar & Rampalli, 2019) in their research.

In (Zhang et al., 2022) autors discuss possible improvements within RDBMSs. Mainstream relational databases are working on optimizing their performance by adapting to big data through the JSON (Java ScripObject Notation) text storage format. In this way, SQL databases achieve efficiency by adopting NoSQL characteristics.

For that reason, the next research question (RQ2) logically focuses on the concept of polyglot persistence and multi-model databases, their characteristics and practical applications. Key differences between these two approaches are shown in Table 6.

To begin with, (Bimonte et al., 2023) discuss two possible approaches for integrating different types of data into one system. These are the parallel use of two or more database management systems (polyglot persistence) and the integration of all data into one DBMS (multi-model). Figure 5, as presented in (Mussbacher et al., 2024), clearly illustrates the hierarchy of multi-language development.

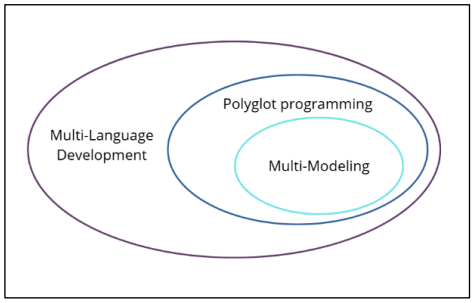


Figure 5. Multi-Language Development Hierarchy presented in (Mussbacher et al., 2024)

Table 6. Comparison of Polyglot Persistence and Multi-Model Databases

Characteristic	Polyglot persistence	Multi-model database
Definition	Multiple DBMSs	Single DBMS with multiple models
Architecture	Distributed	Unified
Complexity	High (integration, learning curve)	Low (single system)
Query language	Multiple	One (unified)
Scalability	Independent per DBMS	Centralized
Consistency	Hard to maintain	Easier
Cost	Higher (many tools, integration)	Lower (single platform)
Use cases	Heterogeneous data, e-commerce, analytics	Smart traffic, healthcare, Customer Relationship Management

Key characteristic of polyglot persistence is that an application uses more database systems of different types to solve conflicting requests. This allows different databases to complement each other and compensate for gaps. However, this architecture certainly increases programming complexity and requires knowledge of different types of databases. If implemented successfully, polyglot persistence brings a great advantage. This approach is presented from a competitive angle in (Peña, 2023). As the environment is dynamic and the company tends to react agilely and quickly to changes in business needs, properly implemented polyglot persistence can effectively support this adaptability. It optimizes data architecture, simplifies operations and enhances the efficiency and effectiveness of the decision-making process. Certainly, Lieponienė (2020) emphasizes that the implementation of this concept, including the choice of technologies, depends on how the stored data is utilized.

Polyglot data systems should keep important features of each database (like scalability), allow quick data changes using a global interface, automatica detection of changes between databases, run queries efficiently across systems, support real-time data processing, and multi-model schema management (Kiehn et al., 2022).

A good example that illustrates the concept of polyglot persistence is provided in (Lajam & Mohammed, 2022). The authors examine the functioning of e-commerce applications. Customer purchase data can be stored using a key-value NoSQL database.

However, if a client wants to know what their friends are buying, a graph database would be a more relevant option. Finally, for the execution of purchases and payments, a relational database is the most suitable choice. This is one example where polyglot persistence is a good choice to take advantage of different types of databases.

Polyglot persistence also has disadvantages. These relate to increased software complexity and a lack of support for maintaining data consistency. As databases are interconnected with mechanisms of other databases, it requires additional effort for their managing. The authors (Kazanavičius, Mažeika & Kalibatienė, 2022) notify that this concept is used only in cases where it is necessary to store different data models, otherwise there is a risk of overload. The disadvantages of polyglot persistence are the technical challenges of managing multiple databases, complex logic in applications, insufficient performance optimization, a steep learning curve for developers, and data consistency, according to (Bimonte et al., 2023). Using multiple databases makes data handling more complex due to differences in models, query languages, and systems, as noted by (Kumar & Rampalli, 2019). The hybrid database model produces different output formats, which directly increases complexity (Vyawahare et al., 2019). Recent studies focus on developing a hybrid framework that uses both databases to truly leverage the benefits of polyglot persistence.

Instead of dealing with the complexity of implementing the principle of polyglot persistence, there is the possibility of using multi-model databases. Multi-model databases simplify application maintenance and eliminate the need for different databases, as noted by (Subramanian & Saravanan, 2024). In their paper "*Holistic evaluation in multi-model databases benchmarking*", (Zhang & Lu, 2021) talk about multi-model databases as the next generation of DBMSs because they integrate flexibility, scalability and consistency. They predict that all leading data management systems will soon support multiple data models on a single platform.

The main advantage of the multi-model database is a unique system for all models, the maintenance of only one database, unique query language, architecture, adapter and topology, as stated by (Lieponienė, 2020). In addition, it is important to emphasize that the cost of scaling is lower because only one database is scaled, instead of several.

The authors of (Zhang & Lu, 2021) additionally note that multi-model databases also support agile development. However, these databases do not fully satisfy requirements such as scalability and performance. In cases where applications only need a limited set of data models, multi-model databases are recommended by (Kazanavičius, Mažeika & Kalibatienė, 2022).

Table 7. Commonly used technology combinations

SQL DB	NoSQL DB	Use Cases	Advantages	Ref.
MySQL	MongoDB, Redis	Real-time geospatial, IoT	Complex query optimization, real-time access, processing large amounts of data, reliable transactions, fast processing sensors data	8, 18, 22
	MongoDB	Web applications with user behavior tracking	Reliable transactions, flexibility with unstructured data (JSON format support)	20, 24
	HBase	High-volume data applications with transactions	High scalability, SQL query support over NoSQL	20
SQLite	MongoDB, Redis	E-commerce applications	Fast in-memory operations, easy integration, flexibility with unstructured data	16
Oracle	MongoDB		High security, support for large number of users, fast data insertion and updating, complex query optimization, data integration, data consistency	13, 17
PostgreSQL	Redis		Reliable transactions, availability-oriented	7
Any SQL	Cassandra	IoT systems (sensor + users and devices data)	Time-stamped data storage, processing large amounts of data	25

The authors (Zhang & Lu, 2021) point out that multi-model databases are useful in applications that operate in areas such as e-commerce, healthcare, online recommendation systems and smart traffic management.

A key part of the effectiveness of a polyglot system is the choice of the database, so RQ3 and its sub-questions are focused on frequently used technologies and their use cases. Table 7 present a summary of commonly used techology combinations in selected papers.

When asked which technology to choose for an application, one possible answer is often based on the most widely used databases.

The DB-Engines website has a table of database processing systems ranked by factors such as the number of mentions on websites, general interest in the system, number of active discussions, and number of job offers involving a given technology. According to the given criteria, Oracle (Oracle Corporation, 2025) relational and multi-model database is in the first place. Then, MySQL (MySQL, 2025), Microsoft SQL Server (Microsoft, 2025), and PostgreSQL (PostgreSQL Global Development Group, 2025) occupy the following positions. MongoDB (MongoDB, 2025) is the highest-ranked NoSQL database (DB-Engines, 2025). The technologies from the top of the ranking list are precisely those most frequently mentioned in selected papers.

Another way to select technologies, more complete than the previous one, involves matching the type of database with the purpose of the future application. In paper "*An integration approach of hybrid databases based on SQL in cloud computing environment*", (Li & Gu, 2019) deal with cloud services. In this area, hybrid databases are becoming a trend. NoSQL databases suitable for this architecture are Redis (Redis, 2025), MongoDB, HBase (Apache Software Foundation, Hbase documentation, 2025), Neo4j (Neo4j, 2025) and Memcached (Memcached, 2025). For MongoDB, HBase, and Neo4J systems, adding a SQL access layer is relatively easier because they have built-in Application Programming Interfaces (API) that support filtering and sorting records based on multiple conditions. In the case of multithreading and queries to multiple databases, MongoDB and Neo4j together perform better in execution time and resource usage, emphasized by (Ye et al., 2023). On the other hand, Redis and Memcached are characterized by difficult integration with SQL systems because they lack such APIs and require their construction first. A good practice when using a MongoDB in a hybrid architecture is to use it for user profile data or configuration data, while SQL databases handle the execution of e-commerce system transactions. The research by (Li & Gu, 2019) also points out the necessary use of SQL databases in banking systems for operational data, while HBase is used for analyzing large amounts of data. Redis and Memcached pair well with MySQL or PostgreSQL relational databases. Redis is presented by (Wu et al., 2017) as a tool suitable for high-traffic applications, i.e., those with many users accessing the system simultaneously. Cassandra (Apache Software Foundation, Hbase documentation, 2025) is a NoSQL database suitable for storing large volumes of time-stamped data, such as data collected from sensors. For this reason, it is often used in IoT (Internet of Things) systems in combination with SQL databases. In such architectures, SQL databases store data about users and devices.

A hybrid database composed of MySQL and MongoDB is discussed in (James & Asagba, 2017). The authors designed the system so that components can operate independently or together, providing flexibility in data management. The main task of MySQL is to handle transactions, while MongoDB plays a key role in data processing due to its efficient reading and writing characteristics.

According to a research evaluation conducted by (Li & Gu, 2019), the combination of MySQL, MongoDB, and Redis has competitive advantages in terms of optimizing complex queries, simplifying complex conditional clauses and certainly the types of

integrated data sources. This hybrid database is a good approach to solving the challenges of processing large amounts of real-time geospatial data. Also (Wu et al., 2017) in his work proposes the use of this combination of technologies.

A similar combination of technologies for applications dealing with e-commerce is presented in (Kazanavičius, Mažeika & Kalibatienė, 2022). A mentioned hybrid database is made up of Redis, MongoDB and SQLite (SQLite Consortium, 2025).

The combination of MySQL and HBase technologies is recommended by (Krishnapriya, Libin, & Gibin, 2021) in their work "*A study for integrating SQL and NoSQL Database*", when there is a need for high scalability to work with large amounts of data and process transactions. Using Apache Phoenix as a SQL translator for HBase allows SQL queries to be executed against a NoSQL database, facillitating integration with a MySQL database easier.

Also, (Krishnapriya, Libin, & Gibin, 2021) mention the MySQL and MongoDB hybrid database, which is well-known and frequently utilized in e-commerce web applications or for tracking user behavior. JSON is a commonly used format in web applications, and MongoDB is useful for storing data in that format.

Oracle is a relational database, suitable for a hybrid architecture alongside the MongoDB NoSQL database according to the research conducted by (Pokorný, 2019). This combination provides a higher level of business security when managing data and well supports systems with a large number of users. The authors (Bjeladinovic, Marjanovic, & Babarogic, 2020) also talk about this combination. They state that MongoDB enables faster data insertion and updating, especially when the data structure is variable. Oracle DBMS features exceptional join efficiency and high-performance multi-table query execution. Thus, the combination is characterized by good performance of complex queries, high data integrity and consistency.

(Bjeladinovic, 2025) in his research mentioned a previous database combination to which Cassandra was added. This improvement affects performance when working with large datasets, although that is not always the case with smaller amounts of data.

The last combination is mentioned by (Khine & Wang, 2019), which is a hybrid database consisting of PostgreSQL and Redis. The SQL part of this database is used for secure bank transactions in online shopping. NoSQL database Redis is responsible for implementing the user's cart due to its availability-oriented design.

At the end of discussion we would like to emphasize the main threat to validity. Namely, the presented analysis relies solely on data available in the Scopus and Google Scholar databases. In the future research more databases should be used to find relevant articles and in that way increase the number of selected articles. Besides, the usage of hybrid database system is evolving in time and we believe that the number of relevant articles will rise in further years.

4 Conclusion

The results of the conducted systematic literature review indicate a wide range of applications for the approaches of polyglot persistence and multi-model databases. Characteristics such as flexibility, scalability and data integrity represent the key requirements of modern information systems, and the analyzed approaches enable these goals to be achieved efficiently. The identified use cases and functionalities suggest significant potential for further development and wider adoption of these solutions in the future, despite the challenges associated with their implementation. Relational database management systems bring complex query optimization alongside with high data integrity and secure data transactions. On the other side, NoSQL databases can contribute to scalability, availability, flexibility with unstuctured data, and processing of large amounts of data. The diverse and fast-changing space of data systems technologies demands that designers and developers understand the concepts, pros and cons, and good practices of both of SQL and NoSQL worlds.

Future research should focus on proper implementation of polyglot persistence, in order to fully leverage its performance and improve the efficiency of business systems. Of particular importance is the development of recommendations for the optimal use of these concepts, tailored to the specific needs of users and the characteristics of information systems, which would represent a logical and valuable direction for further research in this area.

Acknowledgments

This research has been supported by the Ministry of Science, Technological Development and Innovation (Contract No. 451-03-137/2025-03/200156) and the Faculty of Technical Sciences, University of Novi Sad through project "Scientific and Artistic Research Work of Researchers in Teaching and Associate Positions at the Faculty of Technical Sciences, University of Novi Sad 2025" (No. 01-50/295).

References

Majhadi, K., & Machkour, M. (2021). The history and recent advances of Natural Language Interfaces for Database Querying. *E3S Web of Conferences (ICCSRE'2020)*, 229, 01039. https://doi.org/10.1051/e3sconf/202122901039

Ohlsson, A., & Persson, M. (2019). A Coparison in Performance Between a Selection of Databases (Using the benchmark tool YCSB). *Axis Communications AB* (Master's thesis).

Glake, D., Kiehn, F., Schmidt, M., Panse, F., & Ritter, N. (2022). Towards Polyglot Data Stores. *arXiv:2204.05779v1 [cs.DB]*. https://arxiv.org/abs/2204.05779

Nurhadi, R., Kadir, R. A., Mat Surin, E., & Sarker, M. R. (2024). A Systematic Review of Automated Classification for Simple and Complex Query SQL on NoSQL Database. *Computer Systems Science and Engineering, 48(6)*, 1434*, Tech Science Press.* https://doi.org/10.32604/csse.2024.051851

DB-Engines. (225). *Complete Ranking https://db-engines.com/en/ranking*

Khine, P. P., & Wang, Z. (2019). A Review of Polyglot Persistence in the Big Data World. *Information, 10*(4), 141. https://doi.org/10.3390/info10040141

Kitchenham, B. (2004). *Procedures for performing systematic reviews*. Keele University Technical Report TR/SE-0401, ISSN: 1353-7776, & NICTA Technical Report 0400011T.1. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=29890a936639862f45cb9a987dd599dce9759bf5

Li, C., & Gu, J. (2019). An integration approach of hybrid databases based on SQL in cloud computing environment. *Software: Practice and Experience*, 49, 401–422. https://doi.org/10.1002/spe.2666

Lajam, O., & Mohammed, S. (2022). Revisiting polyglot persistence: From principles to practice. *International Journal of Advanced Computer Science and Applications, 13*(5). https://doi.org/10.14569/IJACSA.2022.0130599

Kumar, K. P., & Rampalli, G. (2019). An efficient design for multiple data stores cloud applications. *International Journal of Innovative Technology and Exploring Engineering, 8*(9S3). https://doi.org/10.35940/ijitee.I3062.0789S319

Lieponienė, J. (2020). Recent trends in database technology. *Baltic Journal of Modern Computing, 8*(4), 551–559. https://doi.org/10.22364/bjmc.2020.8.4.06

Harezlak, K., Mermon, M., & Kasprowski, P. (2021). A comparison of the efficiency of eye movement data processing with the usage of a multi-model platform. *Procedia Computer Science, 192*, 3070–3078. https://doi.org/10.1016/j.procs.2021.09.079

Pokorný, J. (2019). Integration of relational and graph databases functionally. *Foundations of Computing and Decision Sciences, 44*(4). https://doi.org/10.2478/fcds-2019-0021

Bimonte, S., Gallinucci, E., Marcel, P., & Rizzi, S. (2023). Logical design of multi-model data warehouses. *Knowledge and Information Systems, 65*, 1067–1103. https://doi.org/10.1007/s10115-022-01788-0

Zhang, C., & Lu, J. (2021). Holistic evaluation in multi-model databases benchmarking. *Distributed and Parallel Databases, 39*, 1–33. https://doi.org/10.1007/s10619-019-07279-6

Kazanavičius, J., Mažeika, D., & Kalibatienė, D. (2022). An approach to migrate a monolith database into multi-model polyglot persistence based on microservice architecture: A case study for mainframe database. *Applied Sciences, 12*, 6189. https://doi.org/10.3390/app12126189

Bjeladinovic, S., Marjanovic, Z., & Babarogic, S. (2020). A proposal of architecture for integration and uniform use of hybrid SQL/NoSQL database components. *The Journal of Systems & Software, 164*, 110633. https://doi.org/10.1016/j.jss.2020.110633

Wu, C., Zhu, Q., Zhang, Y., Du, Z., Ye, X., Qin, H., & Zhou, Y. (2017). A NoSQL–SQL hybrid organization and management approach for real-time geospatial data: A case study of public security video surveillance. *ISPRS International Journal of Geo-Information, 6*(1), 21. https://doi.org/10.3390/ijgi6010021

Peña, L. (2023). Holistic approaches to strategically integrating SQL and NoSQL solutions in hybrid architectures for optimized performance and versatile data handling. *Journal of Artificial Intelligence and Machine Learning in Management*, 93. https://doi.org/10.1016/j.jaml.2023.06.004

Krishnapriya, V., Libin, S., & Gibin, G. (2021). A study for integrating SQL and NoSQL databases. *International Conference on Intellectual Property Rights*, 20/02/2021.

Kiehn, F., Schmidt, M., Glake, D., Panse, F., Wingerath, W., Wollmer, B., Poppinga, M., & Ritter, N. (2022). Polyglot data management: State of the art & open challenges*. Proceedings of the VLDB Endowment (VLDB 2022), 15(12), 3750–3753.* https://doi.org/10.14778/3554821.3554891

Zhang, L., Pang, K., Xu, J., & Niu, B. (2022). JSON-based control model for SQL and NoSQL data conversion in hybrid cloud database. *Journal of Cloud Computing*, 11(23). https://doi.org/10.1186/s13677-022-00302-9

Vyawahare, H. R., Karde, P. P., & Thakare, V. M. (2019). Hybrid database model for efficient performance. *Procedia Computer Science*, 152, 172–178. https://doi.org/10.1016/j.procs.2019.05.040

James, B. E., & Asagba, P. O. (2017). Hybrid database system for big data storage and management. *International Journal of Computer Science, Engineering and Applications (IJCSEA)*, 7(3/4), 15. https://doi.org/10.5121/ijcsea.2017.7402

Bjeladinovic, S. (2025). Extending hybrid SQL/NoSQL database by introducing statement rewriting component. *Computer Science and Information Systems*, 00(0), 0000–0000. https://doi.org/10.2298/CSIS123456789X

Ye, F., Sheng, X., Nedjah, N., Sun, J., & Zhang, P. (2023). A benchmark for performance evaluation of a multi-model database vs. polyglot persistence. *Journal of Database Management*, 34(3). https://doi.org/10.4018/JDM.321756

Zdepski, C., Bini, T. A., & Matos, S. N. (2020). New perspectives for NoSQL database design: A systematic review. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 68(1), 50–62.

Subramanian, S., & Saravanan, S. (2024). Modern trends in NoSQL databases. International Journal of Computer Trends and Technology, 72(9), 126–130. https://doi.org/10.14445/22312803/IJCTTV72I9P119

Mussbacher, G., Combemale, B., Kienzle, J., Burgueño, L., Garcia-Dominguez, A., et al. (2024). Polyglot software development: Wait, what? IEEE Software, pp. 1–8. https://doi.org/10.1109/MS.2023.3347875

MySQL (2025), MySQL Documentation. Retrieved from https://dev.mysql.com/doc/

MongoDB (2025), MongoDB Documentation. Retrieved from https://www.mongodb.com/docs/

Redis (2025), Redis Documentation. Retrieved from https://redis.io/docs/

The Apache Software Foundation (2025), HBase Documentation. Retrieved from https://hbase.apache.org/book.html

SQLite Consortium, SQLite Documentation. Retrieved from https://www.sqlite.org/docs.html

Oracle Corporation (2025), Oracle Database Documentation. Retrieved from https://docs.oracle.com/en/database/

The PostgreSQL Global Development Group (2025), PostgreSQL Documentation. Retrieved from https://www.postgresql.org/docs/

The Apache Software Foundation (2025), Cassandra Documentation. Retrieved from https://cassandra.apache.org/doc/latest/

Neo4j (2025), Neo4j Documentation Retrieved from https://neo4j.com/docs/

Memcached (2025), Memcached Documentation Retrieved from https://memcached.org/

Microsoft (2025), SQL Server Documentation Retrieved from https://learn.microsoft.com/en-us/sql/sql-server/

Appendix

No	Reference	No	Reference
1	(Majhadi & Machkour, 2021)	16	(Kazanavičius, Mažeika & Kalibatienė, 2022)
2	(Ohlsson & Persson, 2019)	17	(Bjeladinovic, Marjanovic, & Babarogic, 2020)
3	(Kitchenham, 2004)	18	(Wu et al., 2017)
4	(Glake et al., 2022)	19	(Peña, 2023)
5	(Nurhadi et al., 2024)	20	(Krishnapriya, Libin, & Gibin, 2021)
6	(DB-Engines, 2025)	21	(Kiehn et al., 2022)
7	(Khine & Wang, 2019)	22	(Zhang et al., 2022)
8	(Li & Gu, 2019)	23	(Vyawahare, Karde, & Thakare, 2019)
9	(Lajam & Mohammed, 2022)	24	(James & Asagba, 2017)
10	(Kumar & Rampalli, 2019)	25	(Bjeladinovic, 2025)
11	(Lieponienė, 2020)	26	(Ye et al., 2023)
12	(Harezlak, Mermon, & Kasprowski, 2021)	27	(Zdepski, Bini, & Matos, 2020)
13	(Pokorný, 2019)	28	(Subramanian & Saravanan, 2024)
14	(Bimonte et al., 2023)	29	(Mussbacher et al., 2024)
15	(Zhang & Lu, 2021)		

Unlocking Data Quality for Responsible Artificial Intelligence

Dajana Marnika
Gimnazija Vladimira Nazora, Perivoj Vladimira Nazora 3/2, 23000 Zadar, Croatia
Faculty of Informatics and Digital Technologies, University of Rijeka, Radmile Matejčić 2, 51000 Rijeka, Croatia
dajana.marnika@student.uniri.hr

Maja Gligora Marković
Faculty of Medicine, University of Rijeka
Department of Biomedical Informatics
Braće Branchetta 20, 51000 Rijeka, Croatia
majagm@medri.uniri.hr

Božidar Kovačić
Faculty of Informatics and Digital Technologies, University of Rijeka, Radmile Matejčić 2, 51000 Rijeka, Croatia
bkovacic@inf.uniri.hr

Abstract. *Artificial intelligence (AI) has been developing and advancing continuously for more than 70 years. The potential of AI is conditioned by the increase in computing power and memory, the progress of algorithms in machine and deep learning, but primarily - the availability of data. A particularly important role in ensuring data quality is played by metadata, which provide contextual information about the data, such as source, structure, creation time, format, and ownership. The aim of this paper is to investigate which types of metadata represent data quality parameters in intelligent systems and AI. The research is based on a theoretical analysis of relevant scientific literature indexed in the Web of Science (WOS) and SCOPUS databases, with the goal of identifying key metadata attributes that influence the quality and usability of data in AI contexts. This approach allows for a critical synthesis of existing knowledge and insights into how metadata contribute to improved outcomes in AI model development and application.*

Keywords. data, metadata, data quality, artificial intelligence

1 Introduction

Machines created by humans have long been capable of performing various types of more difficult and laborious human tasks. However, with the aim of greater productivity and greater ability to perform tasks, humans have been trying to inject human intelligence into machines, which is the original motive for AI. (Jiang et al., 2022)

There are many definitions of the term AI itself. Not so long ago in 1950, Alan Turing posed the philosophical question "Can machines think?" in his essay Computing Machinery and Intelligence, laying the foundations for the further development of computers. The pioneering definition of AI was devised by John McCarthy in 1955, stating that every aspect of learning or any other feature of intelligence can in principle be described so precisely that a machine can be built to simulate it. According to him, artificial intelligence can be broken down into algorithmic processes, i.e. a human precisely defines intelligent behavior, and a machine (AI) does not imitate a human but behaves as if it had intelligence and solves a given problem. A decade later, a similar definition was given by Marvin Minsky (Minsky, n.d.), according to whom AI is the science of how to make machines do things that would require intelligence if they were to be done by humans. Playing chess was the first practical application of AI. Initial forecasts predicted the rapid growth of artificial intelligence comparable to human intelligence, but the complexity of this process soon became apparent. (Leksikografski zavod Miroslav Krleža, n.d.)

With the development of machine learning, the focus has shifted from logical reasoning to statistics and data (Šuman, 2021), meaning that the goal was no longer for a machine to think like a human, but to learn from the data it receives. The definition has expanded over the years to include optimization of rational action, creativity, adaptation, social influence, and ethical use. It is expected that an AI system will be able to achieve human performance and be rational in doing the 'right thing' (Russell & Norvig, 2010) given the available data.

Because of this data dependency, artificial intelligence is inextricably linked to the development and spread of the phenomenon known as Big Data. The term Big Data (Cai & Zhu, 2015; Abdou, 2020) refers to data that is so large and complex that it exceeds the processing capabilities of traditional data management systems and software. The volume and diversity of data on which AI systems learn poses new challenges, not only in processing and storage, but also in the protection, interpretation, and ethical management of that data.

Doug Laney (Laney, 2001) first introduced the concept and described it with 3V characteristics: Volume, Velocity and Variety. With the development of technology and the growth of complexity of data systems, this model has been expanded and now includes 56V (Abdou, 2020) that reflect the comprehensiveness of modern data environments.

Following this development, the Organisation for Economic Co-operation and Development (OECD) (Grobelnik, 2024) defines artificial intelligence as a machine system that, for explicit or implicit purposes, infers from the input data it receives how to generate outputs such as predictions, content, recommendations, or decisions that can affect physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptability after implementation.

The OECD definition of AI reveals an important link between the definition of AI and the practical challenges in its application, particularly in the context of the importance of data quality management. This link is illustrated in Table 1.

Table 1. Relationship between elements of the OECD definition and the importance of data accuracy

Definition element	Importance of data
... from the input it receives it concludes ...	If the input data is incorrect, the system's inference may be incorrect or incomplete. Inconsistent data confuses models and makes it difficult to learn patterns.
... how to generate outputs such as predictions, content, recommendations or decisions ...	Inaccurate and inconsistent data reduce the reliability of system outputs. They can lead to poor or harmful decisions, especially in high-risk systems.
... can affect physical or virtual environments ...	The consequences of inaccurate, incomplete, and inconsistent data are becoming real and measurable – affecting people and systems.
... in levels of autonomy and adaptability ...	If trained on poor-quality data, adaptive systems can adopt incorrect or contradictory patterns of behavior. By adopting errors, they can reinforce them over time.”

The importance of data in information management has been recognized for decades, and it is precisely their massiveness, diversity, and accessibility that have enabled the development and training of advanced AI models in complex, real-world environments. Recent advances in AI have further emphasized the role of data quality in building sustainable *data ecosystems* (Mohammed et el., 2025). However, metadata management and quality parameters are still not systematically regulated or consistently applied in practice.

This paper consists of five parts: the first, introductory part with explanations of basic terms, the second part, which presents the methodology of the conducted research, the third and fourth parts represent a discussion based on selected reviewed papers in order to answer the research questions, while the last part presents the conclusions of the conducted research and future research.

2 Research methodology

The existing literature deals with the topics of data quality and metadata management in a fragmented manner, therefore there is a need for a systematic insight into the challenges, concepts and existing approaches related to data management, data quality and metadata within the context of AI systems. This research (Kitchenham & Charters, 2007) used a literature review in three phases: planning the review, conducting the review and discussing the studies. Through these phases, the results of the work will be presented.

Following the problem of lack of standardization and methodological uniformity that makes it difficult to evaluate the quality of datasets (which can have direct consequences on the performance and reliability of models), we posed the central research question in this article: What are the key metadata that define data quality in machine and deep learning?

In order to answer the question posed, additional research questions were also posed: 1. How is data quality in intelligent systems defined by academic papers and regulatory frameworks? and 2. What are the best practices for collecting, storing, and tracking metadata related to data quality in intelligent systems?

In the preliminary literature search, we used Scopus and Web of Science as primary scientific databases. We supplemented the publications by analyzing references and citations, with an additional search in Google Scholar. Official documents of international institutions - the European Commission and the OECD - were also used.

Thematic areas covered by the queries are: artificial intelligence, AI system lifecycle, data and metadata management, data quality, and ethical and regulatory aspects of AI. The keywords and combinations used in the search were:

- "data quality" OR "data accuracy" OR "data completeness" OR "data consistency" OR "data integrity" OR "data reliability" OR "data veracity"
- "metadata" OR "metadata management" OR "metadata standard*" OR "metadata quality" OR "semantic metadata" OR "provenance metadata"
- "data management" OR "data governance" OR "data stewardship" OR "data curation" OR "data lifecycle" OR "FAIR data" OR "data trustworthiness"
- "artificial intelligence" OR "AI" OR "machine learning" OR "ML" OR "deep learning" OR "neural network*" OR "AI system*" OR "AI model*".

By using the OR operator, we tried to include synonyms and related terms within each query, while by connecting them with the AND operator, we ensured that only papers were found that cover all four key areas, thus filtering out unnecessary results. The number of papers obtained by this search is 57 in the Scopus database and 20 in WOS. With the aim of further narrowing the search, we defined the criteria for inclusion and exclusion of papers for review as shown in Table 2.

Table 2. Inclusion and exclusion criteria for papers

Switching on	Shutdown
Publications between 2013 and 2025.	Publications older than 2013.
English language	They are not in English.
Paper or conference paper	Other types of documents (e.g. books)
Subject area: computer science	Other subject areas
Open access	Open access is not available.

The mentioned criteria reduced the number of papers to 28 in the Scopus database and 7 in WOS. By reviewing the titles and authors of the papers, duplication was observed, i.e. all 7 papers published in WOS were also published in Scopus, which resulted in a total number of papers for review, which is 28. The further selection process was carried out on the basis of the titles and abstracts, and irrelevant papers were excluded, which brought the number of papers to 23. Potentially relevant papers were then analyzed in detail, which reduced their number to 11.

In order to ensure methodological consistency, a formal protocol was developed that included: definition of key terms, database search strategies, list of used and excluded sources with associated justification, and description of the selection methodology. The protocol was formulated before the search itself began and used as a reference frame throughout the process. If necessary, it was subjected to minimal adaptations with clearly documented changes, in accordance with the recommendations of good practice in conducting systematic reviews.

3 Challenges in data preparation and quality

Data is a key aspect of achieving technological progress, but its use is hampered by the wide range of data formats, the lack of interoperability between tools, and the difficulty of discovering and combining data sets (Akhtar et al., 2024). In all three main phases of AI development – design, development, and production – data quality is the foundation of every step (Daswin & Alahakoon, 2022). Contradictory and inconsistent data can seriously undermine the reliability and effectiveness of models, directly affecting their ability to act responsibly and transparently. In order for AI systems to meet these requirements, it is crucial to ensure the consistency, clarity, and accuracy of data throughout the development lifecycle. In an attempt to provide a more comprehensive framework for solving AI tasks, the more precisely defined phases are: dataset selection, preprocessing, feature engineering, and deployment, with metadata recorded throughout all phases (Venkataramanan et al., 2024).

In order to address the challenges of data preparation and quality – the European Commission (2019) has adopted a framework for trustworthy artificial intelligence that includes seven key requirements: 1. human agency and oversight; 2. technical robustness and security; 3. privacy and data management; 4. transparency; 5. diversity, non-discrimination and fairness; 6. social and environmental well-being; and 7. accountability. In achieving the above requirements, especially those related to aspects of data management, transparency and accountability – metadata plays a key role. They enable documentation of the origin, context

and transformations of data, thus becoming an indispensable tool for ensuring the verifiability and monitoring of model's decisions. Without adequate metadata management, it is difficult to achieve full transparency in the operation of models and ensure responsible development and application of AI systems.

Following table 3 (based on (Gröge, 2022)) presents the European Commission's requirements for the reliable application of AI systematized through three basic categories of challenges: data management in the context of artificial intelligence, data governance, and democratization of data access. These categories are also conceptually linked to the Commission's requirements. The table is based on the model from (Gröge, 2022), but is supplemented with additional categories to more precisely capture the challenges arising from the application of AI systems in a real-world context.

Table 3. Data challenges of artificial intelligence

The challenge of data management in the context of artificial intelligence	**The challenge of data management**	**The challenge of democratizing data for AI**
1. Data modeling 2. Metadata management 3. Data architecture 4. Data quality 5. Integration of data from different sources 6. Methodology for managing text, images, video and sensor data	1. Data ownership 2. Data management 3. Privacy and data protection 4. Ethical use of data and algorithms 5. Management mechanisms for the life cycle of AI models	1. Data security 2. Data Engineering 3. Data discovery and exploration 4. User training 5. Access control 6. Personalization and contextualization of data access

The above challenges in data management are deeply interconnected and together form the basis for the successful development of artificial intelligence. Data quality can be described as *fitness for use* (Corte et al., 2024), where it depends on proper modeling, integration and clearly defined metadata. Assessing the level of data quality is crucial for deciding on its suitability in the process of making accurate and reliable decisions. The data architecture must enable the efficient management of all these components. The processing of different types of data (text, images, sensors) adds particular complexity, which requires harmonized methodologies that support all the previous aspects in a single system. The challenges listed in the second category form the regulatory and ethical framework for the responsible use of data and artificial intelligence. Metadata should not be seen only as technical additions to content (Zhan & Hai, 2024), they can in themselves threaten privacy if shared without appropriate protection. This raises important questions about how to design secure data exchange systems. Ultimately, the challenges listed in the third category form the operational and technical basis of data management in AI systems, and their harmonized connectivity enables data access that is efficient, secure, and user-friendly.

The FAIR principles (Findability, Accessibility, Interoperability and Reusability) (Jackson et al., 2024) should be the fundamental framework for metadata management. By using clear identifiers and descriptive labels, metadata must be findable and accessible through standardized protocols. They should be interoperable and reusable. Such compliance directly contributes to greater transparency, accountability and trustworthiness of AI systems, enabling not only the understanding of model outputs, but also the reuse and evaluation of the data on which the model was trained. Without the systematic application of the FAIR principles through metadata, it is difficult to imagine the realization of the principles of responsible and ethical artificial intelligence advocated by the European Commission.

Summarizing the dimensions of data quality, four dimensions can be distinguished – intrinsic, contextual, representational and accessibility dimensions. The intrinsic dimension can be assessed by measuring the internal attributes or characteristics of the data based on given references, and also measures missing values and redundant cases. The contextual dimension ensures that the data is aligned with the needs and objectives of the projects. The representational dimension assesses the formats and structures of the data, for example whether the data is concisely and consistently presented, but also interpretable. The accessibility dimension assesses the extent to which all or part of the data can be retrieved, and additionally allows users to use and share the data with security controls.

The lack of standardized metrics and overlapping dimensions stem from the variability and complexity of data – they are a challenge in data management. Defining standards is essential to ensure consistency, quality and reliability of data throughout the model lifecycle.

In the pursuit of standards, one solution is to introduce datasheets for datasets. The idea is that each dataset would be accompanied by a datasheet that documents its motivation, composition, collection process, recommended uses, etc. Their use would increase transparency and accountability within the machine learning community, mitigate unwanted social biases in machine learning models, facilitate greater

reproducibility of machine learning results, and help researchers and practitioners select more appropriate datasets for their chosen tasks. (Gebru et al., 2018) Ultimately, the adoption of dataset datasheets would foster a more responsible and ethical approach to machine learning research, ensuring that the datasets we use align with our values and objectives.

All the aforementioned dimensions and challenges in data management indicate that systematic metadata management is a key prerequisite for high-quality, transparent, and responsible use of AI systems. The implementation of standards such as the FAIR principles, the adoption of datasheets for datasets, and compliance with regulatory and ethical frameworks enable not only reproducibility and findability of data but also greater reliability and ethical behavior of algorithms themselves. This clearly links the need for strategic planning and data quality control, forming the foundation for the responsible deployment of AI, which is the main thesis of this work.

4 Metadata Management in Data Lakes: Ensuring Quality and Interoperability for AI Models

For a datasheet to be truly useful, it must contain clearly defined metadata. Their importance grows as the amount of data grows. It stems from the fact that AI models depend on the data from which they learn and on which they make decisions. Therefore – the data, i.e. its origin, quality, structure and meaning, must be precisely recorded and understood. In this context, the concept of a data lake emerges as a solution to the problems of big data heterogeneity – storing data in a way that prevents the lake from turning into a swamp, i.e. into an unusable data lake. (Sawadogo Pegdwendé et al., 2019)

The authors offer various strategies for metadata management. The underlying thesis is that metadata storage should be part of an organization's broader metadata management strategy, which in turn should be part of an organization's broader data management policy (Sheldon, 2023).

The characteristics of metadata quality, according to (Hillman & Thomas, 2004) are: completeness, accuracy, provenance, compliance with expectations, logical consistency and coherence, timeliness, and availability. Furthermore, it should be possible (although for most data it is not) to measure their attributes – semantic and syntactic structure, as well as the data values themselves – by automated systems (Hillman & Thomas, 2004).

Building upon this challenge, the FAIR principles (Findability, Accessibility, Interoperability, Reusability) provide a conceptual framework for assessing metadata quality, while the AIDRIN (Hiniduma et. al., 2024) framework translates these principles into measurable indicators. For instance, Findability can be quantified by metadata completeness, whereas Interoperability can be assessed by consistency and standardization across datasets. In this way, AIDRIN enables concrete, quantitative evaluation of the degree to which datasets adhere to FAIR principles, bridging the gap between abstract guidelines and practical implementation.

Semantic metadata enrichment is also highlighted by (Sawadogo Pegdwendé et al., 2019) as the first of six core functionalities that a metadata system within a data lake should provide. Semantic enrichment should enable the assignment of contextual tags to data using ontologies, thereby increasing their comprehensibility and revealing potential relationships between data sets. Other functionalities are: data indexing, link generation and preservation, polymorphism, versioning, and data tracking (Sawadogo Pegdwendé et al., 2019). Data indexing provides efficient access to data through structured indexes, which is particularly useful for managing textual and semi-structured data. Link generation and preservation identifies similarities and existing relationships between data sets and enables the discovery of clusters of related data. Data polymorphism refers to the storage of multiple versions of the same data in different formats, which facilitates analysis and avoids multiple preprocessing. Data versioning enables changes to be tracked, previous states to be preserved, and analysis to be repeated. Finally, usage tracking records user interactions with data, which contributes to security, monitoring, and anomaly detection.

In order to make certain data lake model as efficient as possible and to achieve the previously mentioned features, metadata should first be divided into three groups: intra-object, inter-object and global (Sawadogo Pegdwendé et al., 2019). Intra-object metadata is obtained from the file system: object title, size, last modified date, access path, etc., i.e. it is associated with a specific object. Inter-object metadata explains the relationships between at least two objects, where each object can simultaneously belong to multiple groups. Such groups can be automatically derived from semantic metadata such as tags and business categories. Global metadata potentially refers to the entire data lake.

To ensure consistent application of these principles in practice, it is recommended to establish a specialized team responsible for metadata management and storage, and to establish a centralized repository that allows easier monitoring and access to information or (primary) data. (Sheldon, 2023)

The Common Metadata Framework (CMF) proposes a metadata system that would record metadata for all stages and datasets, enabling the search for the optimal execution path (Koomthanam et al., 2024). One of the proposed solutions is the collection of metadata from open platforms such as Papers-with-Code, OpenML and

Hugging Face (Venkataramanan et al., 2024), although the proponents themselves see how integrating and unifying different terminologies and data formats from these diverse sources is a challenge. In this context, the Croissant method (Akhtar et al., 2024) offers a standardized, layered metadata format enabling consistent recording of basic data, resources, structure and semantic information in a way that promotes interoperability and easier unification of data from different sources. The Croissant represents a practical solution to overcome the problem of heterogeneity of formats and terminologies, facilitating the automated integration and search of metadata in AI ecosystems.

The added value of metadata exploitation is also demonstrated by the ProbSAP system (Wang et al., 2023), which uses metadata clustering to address the problems of unbalanced and multidimensional data sets in the context of predictive analytics. By integrating scalable metadata-based data clustering and an optimized predictive model (XGBoost), ProbSAP demonstrates how proper processing and structuring of metadata can directly contribute to increasing the accuracy and reliability of analytical results. Such an approach confirms the importance of metadata not only as a technical description of data, but also as an active component in improving the performance of intelligent decision-making systems.

These conceptual frameworks can be further illustrated with examples from different research domains, demonstrating how metadata are directly linked to data quality. Leipzig et al. (2020) emphasize that standards such as DICOM, EML, MIAME and CodeMeta underpin reproducible computational research by ensuring data provenance, context and structural integrity. Řezník et al. (2022) focus on geo(metadata) and repositories, demonstrating through examples such as OGC CSW, open data portals, Schema.org and Linked Open Data how proper documentation and semantic annotation directly improve data findability and reusability. Grant et al. (2024) show that spatial metadata, when integrated into machine learning models, enhance the objectivity and accuracy of geological drill core quantification, thus revealing a direct link between metadata and the reliability of analytical methods. Taken together, these contributions highlight that across domains metadata are consistently recognized as a cornerstone of data quality – whether through reproducibility, findability, or the precision of analytical processing.

It is the responsibility of the data author not only to enable the reuse of large-scale metadata but also to ensure its reproducibility. Emphasis must be placed on input data, as they carry most metadata standards, and on descriptive standards (metadata) to provide context, provenance, authenticity, and the data lifecycle. Sandve et al. (2013) identified the most common sources of reproducibility failures: lack of workflow frameworks, missing platform and software dependencies, manual data manipulation or web-based steps, lack of versioning, absence of intermediate or plotted data, and insufficient literate programming or context that can derail a reproducible analysis.

Despite the progress in the development of tools and standards, there is still no universal solution that would allow compiling a list of specific quality assurance techniques that would be applicable across a wide range of domains and data types. The large volume and different types of data are fundamental difficulties. Quality criteria (Gebru et al., 2018) must be considered based on the specific tasks of AI model development or different stages of the development process. Instead of a single comprehensive solution, the creation of a program roadmap for project managers has been proposed (Hillman & Thomas, 2004) but the time required and operational costs are also major constraints for selecting suitable roadmaps (Gebru et al., 2018). This challenge is further complicated by the fact that quality requirements vary not only across industries but also within different stages of the AI system lifecycle. Therefore, approaches must be flexible and context-specific rather than universally predefined.

5 Conclusion

Today, in a digital environment enriched by AI and Big Data, metadata is becoming an indispensable tool for managing complex data systems. With the exponential growth of data volume, AI systems increasingly depend on clearly defined metadata that enables understanding of the origin, structure, and quality of data. Without metadata, it is difficult to ensure the reliability, transparency, and ethical application of AI models in dynamic and distributed environments.

Theoretical analyses and applied studies confirm that metadata are essential for reproducibility, findability, and precise analytical outcomes across different domains. Moreover, the responsibility of data authors to ensure proper metadata management is critical for reproducibility and trustworthy AI outcomes.

With the sentence "An algorithm is only as good as the data it works with." (Corte et al., 2024) we can summarize the views presented in this work. Without reliable, representative and quality data, even the most advanced algorithms cannot deliver reliable results. We can conclude that adopting a strategic plan that includes analyzing the data infrastructure, understanding the methods and sources of data collection, and creating a plan for data cleaning is a fundamental step before launching an AI model.

The implementation of ethical standards and enforcement of data protection laws on a global scale

should be a fundamental demand of the world's leading statesmen and governments. In order to ensure the responsible use of AI, i.e. privacy and user trust in digital ecosystems, it is necessary to review it.

Systematic metadata management is a key prerequisite for quality data management, easier searching and finding of resources, and making informed decisions in conditions of exponential growth and complexity of data. Future research by the same author(s) will include a systematic review of existing data quality management models, with data storage as a fundamental activity in working with various AI models.

Acknowledgments

The work was created as part of the University of Rijeka project entitled “Learning analytics in e-learning systems based on interactive data visualization assisted by data mining”. The project ID is uniri-iskusni-drustv-23-270.

References

Abou-El-Ela Abdou, H. (2020.). Fifty-six big data V'scharacteristics and proposed strategies to overcome security and privacy challenges (BD2). *Journal of Information Security*, 11(4): 304–328.doi:10.4236/jis.2020.114019

Akhtar, M., Benjelloun, O., Conforti, C., Foschini, L., Gijsbers, P., Giner-Miguelez, J., … Zhang, L. (2024.). Croissant: A metadata format for ML-ready datasets. Workshop on Data Management forn End-to-End Machine Learning (DEEM 24), Santiago, Chile. ACM, NewYork, NY, USA, doi: 10.1145/3650203.3663326

Cai, L., & Zhu, Y. (2015.). The challenges of data quality and data quality assessment in the big data era. Data Science Journal, 14(2): 1–10.doi: 10.5334/dsj-2015-002

Corte, C., Sanz, C., Etcheverry, L., & Marotta, A. (2024.). Data quality management for responsible AI in data lakes. In Proceedings of the VLDB 2024 Workshop on Trust and Data Accountability (TaDA 2024). VLDB Endowment. Retrieved from https://www.vldb.org/workshops/2024/proceedings/TaDA/TaDA.13.pdf

Daswin, D. S., & Alahakoon, D. (2022.). An artificial intelligence life cycle: From conception to production. Patterns, 3(6), Article 100489.doi: 10.1016/j.patter.2022.100489

European Commission, High-Level Expert Group on Artificial Intelligence. (2019). Ethics and guidelines for trustworthy AI. Retrieved from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé, H., & Crawford, K. (2018.). Datasheets for datasets. arXiv. https://arxiv.org/abs/1803.09010

Grant, L. J. C., Massot-Campos, M., Coggon, R. M., Thornton, B., Rotondo, F. C., Harris, M., Evans, A. D., & Teagle, D. A. H. (2024). Leveraging spatial metadata in machine learning for improved objective quantification of geological drill core. Earth and Space Science, 11(3), e2023EA003220. doi: 10.1029/2023EA003220

Gröge, C. (2022.). Industry experiences on the data challenges of AI and the call for a data ecosystem for industrial enterprises [Lecture]. Retrieved from https://christophgroeger.de/download/Groeger_There_Is_No_AI_Without_Data.pdf

Grobelnik, M., Perset, K., & Russell, S. (2024.). What is AI? Can you make a clear distinction between AI and non-AI systems? OECD.AI. Retrieved from https://oecd.ai/en/wonk/definition

Hillmann, D., Thomas, B. (2004.). Metadata in Practice. Chicago: ALA Editions.

Hiniduma, K., Byna, S., Bez J. L., Madduri, R. (2024.).AI Data Readiness Inspector (AIDRIN) for Quantitative Assessment of Data Readiness for AI. SSDBM '24: Proceedings of the 36th International Conference on Scientific and Statistical Database Management Article No.: 7, Pages 1 – 12. doi: 10.1145/3676288.3676296

Jackson, S., Khan, S., Cummings, N., Hodson, J., & de Witt, S. (2024.). FAIR-MAST: A fusion device data management system. SoftwareX, 27: Article 101869. doi: 10.1016/j.softx.2024.101869

Jiang, Y., Li, X., Luo, H., Yin, S., & Kaynak, O. (2022.). Quo vadis artificial intelligence. Discover Artificial Intelligence, 2(4): 1–19. doi: 10.1007/s44163-022-00022-8

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering (EBSE Technical Report EBSE-2007-01). Keele University.

Koomthanam, A. J., Tripathy, A., Serebryakov, S., Nayak, G., Foltin, M., & Bhattacharya, S. (2024.). Common metadata framework: Integrated for

trustworthy artificial intelligence pipelines. IEEE Internet Computing, 28(3): 37–44. doi: 10.1109/MIC.2024.3377170

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. META Group Research Note, 6. Retrieved from https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--an...

Leipzig, J., Nüst, D., Hoyt, C. T., Soiland-Reyes, S., Ram, K., & Greenberg, J. (2021). The role of metadata in reproducible computational research. Patterns, 2(9), 100322. doi: 10.1016/j.patter.2021.100322

Leksikografski zavod Miroslav Krleža. Umjetna inteligencija. (n.d.). In Hrvatska enciklopedija. Retrieved from https://www.enciklopedija.hr/clanak/umjetna-inteligencija

Minsky, M. (n.d.). Marvin Minsky. In Encyclopaedia Britannica. Retrieved from https://www.britannica.com/biography/Marvin-Minsky

Mohammed, S., Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., & Harmouch, H. (2022.). The effects of data quality on machine learning performance on tabular data. Information Systems, 132: Article 102549. doi: 10.1016/j.is.2025.102549

Řezník, T., Raes, L., Stott, A., De Lathouwer, B., Perego,

A., Charvát, K., & Kafka, Š. (2022). Improving the documentation and findability of data services and repositories: A review of (meta)data management approaches. Computers & Geosciences, 169, 105194. doi: 10.1016/j.cageo.2022.105194

Russel, S. J., & Norvig, P. (2010). Artificial intelligence: A modern approach (3rd ed.). Pearson Sandve, G. K., Nekrutenko, A., Taylor, J., Hovig, E. (2013.). Ten Simple Rules for Reproducible Computational Research. PLoS Comput Biol 9(10): e1003285.doi: 10.1371/journal.pcbi.1003285

Sawadogo, P. N., Scholly, É., Favre, C., Ferey, É., Loudcher, S., & Darmont, J. (2019.). Metadata systems for data lakes: Models and features. In: Welzer, T., et al. New Trends in Databases and Information Systems. ADBIS 2019. Communications in Computer and Information Science, vol 1064. Springer, Cham.doi: 10.1007/978-3-030-30278-8_43

Sheldon, R. (2023.). 6 best practices for metadata storage and management. TechTarget. Retrieved from https://www.techtarget.com/searchstorage/feature/6-best-practices-for-metadata-storage-and-management

Šuman, S. (2021). Pregled metoda obrade prirodnih jezika i strojnog prevođenja. Zbornik Veleučilišta u Rijeci, 9(1), 137–152. Retrieved from https://hrcak.srce.hr/file/377721

Venkataramanan, R., Tripathy, A. K., Serebryakov, S., Annmary, J., Shah, A., Bhattacharya, S., Foltin, M., Faraboschi, P., Roy, K., & Sheth, A. (2024.). Constructing a metadata knowledge graph as an atlas for demystifying AI pipeline optimization. Front. Big Data, 7: Article 1476506.doi: 10.3389/fdata.2024.1476506

Wang, X. N., Zhao, Y. B., Li, C., & Ren, P. (2023.). ProbSAP: A comprehensive and high-performance system for student academic performance prediction. Pattern Recognition, 137: Article 109309.doi: 10.1016/j.patter.2022.100489

Zhan, D., & Hai, R. (2024.). Will sharing metadata leak privacy? IEEE 40th International Conference on Data Engineering Workshops (ICDEW), Utrecht, Netherlands, 2024, 317–323. doi: 10.1109/ICDEW61823.2024.00047

Building Trustworthy Knowledge: Data Quality and Validation Protocols for Educational Knowledge Graphs

Ana Pasarić, Danijela Jakšić, Patrizia Poščić
University of Rijeka
Faculty of Informatics and Digital Technologies
Radmile Matejčić 2, 51000 Rijeka, Croatia
ana.pasaric1@student.uniri.hr, {danijela.jaksic, patrizia}@inf.uniri.hr

Abstract. *Todays education struggles with rapid technological change, the need for interdisciplinary learning, and demands for personalization. Knowledge Graphs (KGs) offer a crucial AI solution, organizing vast information and enabling advanced reasoning. In education, KGs are transformative, forming the foundation of smart learning by enhancing personalized learning, curriculum development, and assessment. However, smart learning's effectiveness relies entirely on KG data quality. The diverse and dynamic nature of educational data highlights a critical need for rigorous data validation. Populating KGs often faces a knowledge acquisition bottleneck, impeding consistent, high-quality information. This review therefore assesses current educational data management methods, evaluating existing protocols to determine if new or improved standards are essential for ensuring a continuous flow of accurate knowledge into KGs, thereby fostering more effective intelligent educational systems.*

Keywords: knowledge graphs, education, AI

1 Introduction

The digital age is defined by an unprecedented explosion of data, with an estimated 149 zettabytes created, captured, copied, and consumed globally in 2024, a figure projected to surge to 394 zettabytes by the end of 2028[1]. This exponential growth, largely fuelled by the proliferation of Internet of Things (IoT) devices, the demand for real-time processing, and widespread cloud adoption, underscores a critical need for advanced information organization (Qin et al., 2020). A substantial portion of this burgeoning data, estimated at 80% remains unstructured, presenting both significant challenges and opportunities for effective utilization[2].

The modern education landscape faces complex challenges driven by rapid technological advancements, the growing need for interdisciplinary connections, and an escalating demand for personalized learning experiences (Fettach et al., 2022, Hu et al., 2023). Keeping pace requires continuous adaptation from educators and students, demanding ongoing skill development, frequent curriculum updates, and significant investment in new resources (Paneque et al., 2023, Yang Y. et al., 2024). As technology and industry standards shift, educators must promptly adjust content, develop innovative tech-driven approaches, and integrate ethical considerations (Abu-Rasheed et al., 2025, Fan et al., 2024, Yan et al., 2024). In response to these pervasive data and educational complexities, Knowledge Graphs (KGs) have emerged as indispensable structures within Artificial Intelligence (AI), expert systems, data management, and intelligent applications (Abu-Salih & Alotaibi, 2024, Deng et al., 2021).

KGs serve as structured repositories that formalize domain expertise and facilitate sophisticated reasoning, playing a critical role across diverse fields such as medicine, e-commerce, and, particularly, education (Qu et al., 2024). Within the educational sector, KGs underpin advancements in personalized learning, curriculum design, content recommendation systems, innovative teaching methods and muti-dimensional assessment driving the increasing global attention given to smart learning (Li et. al., 2022, Chen Z. et al., 2024).

Smart learning, or smart education, has rapidly evolved since the 2010s, shifting from traditional paradigms to emphasize technology-embedded, self-directed, motivated, adaptive, and resource-rich learning experiences (Qin et al., 2020, Qu et al., 2024). Its key features include being learner-centric, personalized, adaptive, interactive, collaborative, context-aware, and ubiquitous, enabling anytime,

[1] https://www.statista.com/statistics/871513/worldwide-data-created/

[2] https://www.forbes.com/sites/traceywelsonrossman/2019/01/28/i-see-data-forge-ai-mines-the-worlds-unstructured-data/?ctpv=searchpage

anywhere knowledge acquisition via smart devices (Mzwri & Turcsányi-Szabo, 2023).

However, smart learning's effectiveness critically depends on the quality of knowledge within KGs. This presents a significant challenge: knowledge acquisition (Tian et al. 2022). Populating KGs, whether through laborious manual encoding or automated extraction from diverse data, often leads to a knowledge acquisition bottleneck (Paneque et al., 2023). Therefore, robust protocols are essential to manage the quality, consistency, and validity of knowledge within KGs, as these directly impact intelligent system reliability.

Our objective is to evaluate current methodologies' efficacy in addressing the unique data challenges of educational contexts. This review will assess if existing protocols adequately handle the scale, diversity, and dynamic nature of educational data, and determine the need for new or improved protocols to ensure the consistent acquisition of high-quality, verifiable knowledge within educational KGs. This work aims to contribute to the development of more accurate, adaptive, and impactful intelligent educational systems.

2 Methodology

This systematic literature review (SLR) critically evaluates how well current methodologies address the unique data challenges in education. Our main goal is to see if existing approaches are robust enough to handle the sheer scale, diversity, and dynamic nature of educational data. We're also assessing whether new or significantly improved protocols are needed to ensure we consistently acquire high-quality, verifiable knowledge within educational Knowledge Graphs (KGs). Ultimately, this work aims to contribute to more accurate, adaptive, and impactful intelligent educational systems.

To gather relevant articles, we initiated an extensive search on Scopus, limiting results to English-language publications (2020-2025). Our initial keyword query, (knowledge AND graph AND in AND education), yielded approximately 22 articles. Recognizing that a single database search can limit comprehensive coverage and introduce bias, we expanded our screening process to include Google Scholar. For instance, the exact phrase allintitle: "knowledge graph education" was utilized to specifically target articles where "knowledge graph" and "education" appeared together in the title, thereby enhancing the precision and relevance of the retrieved results. This expansion aimed to cast a wider net for relevant literature and ensure a more robust and diverse dataset. By thoroughly reviewing citations and references from the initially retrieved papers on Google Scholar, we added another 75 articles to our dataset. This preliminary phase resulted in a total of 97 records.

A rigorous screening process followed to eliminate redundant or irrelevant articles. We carefully examined the titles and abstracts of all 97 records against our pre-defined inclusion criteria, leading to the exclusion of 65 records. In the final stage of the literature review, we identified 32 papers that met our qualifications for inclusion in this comprehensive review of Knowledge Graph use in an educational context.

3 The Expansive Role of Knowledge Graphs in Modern Education

Knowledge graphs (KGs) are fundamentally changing education, proving themselves to be indispensable tools for teaching and learning. They are increasingly employed to enhance personalized learning experiences, optimize curriculum design, facilitate concept mapping, and power educational content recommendation systems. KGs further support diverse pedagogical approaches, including personalized and collaborative learning, by recommending suitable activities and resources, thereby improving group knowledge building (Chen L. et al., 2024, Edge et al., 2024, Gligorea et al., 2023).

KGs are highly versatile in education, applied across disciplines from history to medicine, and spanning all levels from primary to higher education (Agrawal et al., 2022, Fan et al., 2024, Hou et al., 2025, Qu et al., 2024). Their utility includes recommending personalized learning resources, predicting performance, generating questions, retrieving materials, and improving course management and curriculum coherence (Bernasconi et al., 2025, Deng et al, 2021, Qi et al., 2025). KGs dynamically enhance education by acting as continuously updated information hubs. They provide real-time access to current trends and foster interdisciplinary connections in course design. Crucially, KGs facilitate personalized learning by analyzing student interactions and feedback to tailor content, optimize materials, and boost engagement and innovative thinking (Qin et al., 2020, Fan et al., 2024, Hu et al., 2023, Li et al., 2022). Ultimately, KGs are proving transformative across diverse educational domains. They provide structured, interconnected knowledge that addresses key challenges, with applications ranging from data integration in Learning Management Systems (LMSs) to specialized pedagogical approaches in subjects like history, computer science, and creative arts (Mzwri & Turcsányi-Szabo, 2023, Xue, 2023, 32).

A core similarity among many proposed KG-based educational models is their focus on semantic enrichment and the creation of structured representations of knowledge. For instance, the e-LION model (Paneque et al., 2023) addresses the challenge of heterogeneous LMS data by transforming

raw information into a standardized RDF knowledge graph, enabling richer data analysis and supporting AI algorithms to identify implicit student interaction patterns. Similarly, in primary and secondary history education, a proposed model (Yu et al., 2021) uses a semantically enriched KG to define historical events by attributes like time and place, fostering deeper understanding and cross-disciplinary exploration. The cybersecurity education model (Deng et al., 2021) also relies on a semantically enriched KG built from public information (Wikipedia articles, YouTube videos) and advanced word embedding to guide students through problem-based learning scenarios. This shared emphasis on building well-defined, interconnected knowledge structures underpins their ability to provide context-aware and personalized learning experiences.

The scope of these models extends across various learning experiences and addresses different educational gaps. While e-LION focuses on administrative support and data analytics within the broader e-learning ecosystem (Paneque et al., 2023), other models directly enhance student learning outcomes. The history education model (Yu et al., 2021) aims for improved retention and deeper understanding of complex historical concepts. The cybersecurity model (Deng et al., 2021) targets practical problem-solving skills and increased engagement in a highly specialized technical field. Furthermore, the MEduKG model (Li et al., 2022) pushes the boundaries by constructing a multi-modal educational knowledge graph, integrating not just text but also images and teacher speech to create a more comprehensive and engaging learning experience, addressing the limitation of predominantly text-based KGs. This multi-modal approach is a significant step towards covering the richness of real classroom interactions that other models, focused primarily on textual data or structured content, might not fully capture. Another model (Hu et al., 2023) for professional introduction courses exemplifies systematizing complex content into hierarchical knowledge trees and then KGs, demonstrating applicability beyond traditional academic subjects to vocational training.

Finally, while proposed models show significant advancements, they also indicate opportunities for further integration. KG construction varies, from multi-agent crawling in cybersecurity (Deng et al., 2021) to combining machine learning and human expertise for text-based KGs (Agrawal et al., 2022). The integration of Large Language Models (LLMs) with KGs, exemplified by Graphusion for NLP education (Yang R. et al., 2024) and IntelliChain for Socratic math teaching (Qi et al., 2025), powerfully enhances content generation and dialogue-based learning, an area some earlier models underutilized. Similarly, models like Optimal Knowledge Component Extracting (OKCE) (Choi et al., 2023) and KG-PLPPM (Hou et al., 2025) for personalized learning paths complement existing KGs by focusing on knowledge component relationships and addressing specific learning gaps.

The overarching theme is clear: KGs, by offering versatile applications from data integration to highly personalized and multi-modal learning experiences, are continuously expanding their footprint across the educational landscape.

4 Data Quality and Validation in Educational KGs: Addressing Challenges of Sparse and Incomplete Data

While KGs are undeniably powerful tools for revolutionizing education, their effective implementation faces several key challenges. These include insufficient data for construction, difficulties in extending applications across diverse subject areas, and the critical need for more comprehensive user feedback and evaluation processes.

One primary challenge lies in KG construction, which typically involves both top-down (domain modeling and rule definition) and bottom-up (knowledge acquisition and extraction) approaches. Building high-quality educational KGs necessitates a huge quantity of clean, prepared course-related data (Agrawal et al., 2023, Qin et al., 2020, Elkaimbillah et al., 2021). However, acquiring and integrating this data, especially from diverse sources like unstructured texts, can be complex and resource-intensive. Existing methods for knowledge extraction heavily rely on Natural Language Processing (NLP), text mining, and machine learning techniques, including Conditional Random Fields (CRF), Support Vector Machines (SVM), Bi-directional Long Short-Term Memory (BiLSTM) networks, and Hidden Markov Models (HMM) (Abu-Salih, 2021, Bernasconi et al., 2025, Elkaimbillah et al., 2021, Fettach et al. 2022). Despite these advanced techniques, the sheer volume and heterogeneity of educational content often result in insufficient data for fully comprehensive and automated KG development. Furthermore, while generic KGs like Wikidata provide broad encyclopedic content, developing domain-specific KGs tailored for particular educational problems, requires specialized data and expertise, which might not always be readily available or easily scalable across numerous subject areas (Abu-Salih & Alotaibi, 2024, Agrawal et al., 2023, Chen Z. et al., 2024).

Another significant challenge is extending KG applications across diverse subject areas and ensuring their quality and reliability. While knowledge graph embeddings (KGEs) like TransE, TransH, and TransR effectively transform entities and relationships into low-dimensional vector representations for numerical analysis, ensuring their relevance and accuracy across subjects like history, mathematics, or art and design remains complex (Bernasconi et al., 2025,

Elkaimbillah et al. 2021, Yu et al., 2021). More critically, the implementation of robust data check and validation protocols is paramount for controlling knowledge input into KBs. These protocols encompass syntactic and structural validation (e.g., OWL ontology axioms, RDF graph well-formedness), semantic consistency checks (e.g., logical consistency, redundancy detection), and plausibility checks (leveraging heuristics or external knowledge) (Qu et al., 2024, Weichselbraun et al., 2022, Yang R. et al., 2024).

Table 1. Educational Knowledge Graphs: Usage, Technical Means, and Evaluation Protocols

Ref.	KG usage	Technical means	Evaluation protocol
(22)	Smart education applications	NER, BILSTM-CRF, Word2vec, Neo4j	Teaching quality and student learning outcomes
(32)	Personalized learning	Semi-automatic annotation and extraction	N/A
(10)	Problem-based learning (PBL)	XLNet, KG, interactive GUI	Student learning outtcome
(31)	Recommender systems	LightGCN	Recall@K, NDCG@K
(4)	Personalized learning using chatbot for question answering	Bi-LSTM-CRF, BERT), KG, SPARQL	Completeness consistency, accuracy, application impact
(18)	Intelligent tutoring system	EduBERT, BiLSTM-CRF, speech-fusion	Precision, recall, F1-score f
(17)	Learning Resources and Guiding Recommendations	TransE Neo4j	Sudent learning outcomes
(27)	intelligent tutoring systems in e-learning	OWL 2 DL, RDF, SPARQL, KG	Accuracy, precision, recall, F1 and interpretability
(20)	Inteligent tutoring systems in e-learning	OWL2, RDF, SWRL	Accuracy, precision, recall, F1 and interpretability
(9)	Personal learning path reccommendation	LASSO, Random Forest, DKT, LSTM	AUC
(30)	Knowledge recommendation	MKR	LRS, LKG, LREG, L2
(29)	TutorQA (Question Answering)	Graphusion	TutorQA benchmark
(7)	Collaborative problem solving (CPS)	Knowledge extraction & structural analysis	Student knowledge identification, visualization, and activation patterns
(13)	Learning resource, curriculum design and educaional assesment	Node feature extraction method	Curriculum, personalized learning, teaching methods, assessment
(6)	Personalized learning	IQASs	Multi-faceted quantitative & user-centered qualitative assessment
(1)	Personal learning path reccommendation	N/A	Relevance, accuracy, ADC, Modularity
(21)	Personalized learning via chain-of-thought dialogue	LLMs, knowledge graphs, multi-agent system	Teaching outcomes, pedagogical dialogue, alignment with objectives
(16)	Personal learning path reccommendation	TransR Euclidean distance V-DINA	Learning path recommendation accuracy via KG semantic and user history score similarity
(11)	Global QA (Question Answering)	GraphRAG	RAG answer evaluation metrics

Furthermore, process-oriented controls like comprehensive review and approval workflows, versioning, access control, and robust feedback mechanisms are essential for continuous quality enhancement. The emergence of "hallucination" in Large Language Models (LLMs) underscores this challenge, driving research into KG augmentation strategies—Knowledge-Aware Inference, Learning, and Validation—to enhance LLM performance and ensure factual accuracy by providing external,

structured knowledge and robust fact-checking (Abu-Salih, 2021, Agrawal et al., 2023, Qi et al., 2025). However, despite these advancements, ensuring consistent application and comprehensive coverage of these validation protocols across every unique educational context and domain still represents a considerable implementation hurdle.

The effective implementation of KGs in education faces persistent challenges related to data quality, including issues of completeness, consistency, and noise. A significant hurdle is the management of sparse and incomplete data within many educational KGs, coupled with the persistent challenge of ensuring truthfulness and semantic accuracy. This necessitates robust protocols for verifying the factual correctness and meaningfulness of knowledge extracted from potentially noisy or ambiguous sources, such as diverse free-text curricula or unvalidated online resources (Mzwri & Turcsányi-Szabo, 2023, Waagmeester et al., 2020).

Human expert validation is a consistent requirement across proposed models. For example, some approaches (Abu-Salih & Alotaibi, 2024, Agrawal et al., 2022) implicitly integrate data validation through rigorous ontology development and schema constraints, where human domain experts refine extracted entities and relations, define types, and apply validation rules to ensure information accuracy and consistency. Similarly, other models (Abu-Rasheed et al., 2025, Qi et al., 2025) explicitly employ a human-in-the-loop approach, with teachers validating LLM-extracted and classified topics and sub-topics. This expert verification, combined with quantitative accuracy data, establishes high-quality and reliable semantic relations, effectively acting as a robust fact-checking mechanism.

Beyond human oversight, the proposed models integrate robust computational and experimental validation to ensure the quality and impact of their KGs. For instance, the MEduKG model (Li et al., 2022) employs rigorous experimental evaluation, utilizing standard metrics such as Accuracy, Precision, Recall, and F1-Score to validate concept recognition and relation extraction, further predicting entities in unlabelled text as a quality assurance measure. Similarly, the proposed KG model (Hu et al., 2023) for professional introduction courses leverages the transE algorithm to achieve over 90% accuracy in predicting relationships, with its educational impact validated by significant improvements in student scores. The e-LION model (Paneque et al., 2023) prioritizes extensive data consolidation and cleaning to resolve semantic inconsistencies, validating its predictive capabilities through rigorous metrics for student grade prediction and time-series forecasting. While the cybersecurity model (Deng et al., 2021) lacks explicit systematic fact-checking for all KG output, it does incorporate content vetting for appropriateness and factual correctness during resource selection.

Finally, emerging research, particularly Knowledge-Aware Validation (Yang R. et al., 2024) for LLM augmentation, explicitly leverages KGs as a robust fact-checking mechanism for generated information, a strategy that could significantly bolster the validation protocols of educational KG models currently lacking explicit fact-checking components.

5 Addressing Critical Gaps in Educational Knowledge Graph Protocols

Despite their significant potential, existing educational KGs are hindered by critical limitations, primarily concerning data quality, including completeness, consistency, and noise. A persistent challenge is maintaining truthfulness and semantic accuracy, especially when dealing with large, continuously evolving datasets common in education. This necessitates urgent development of standardized protocols for KG creation and management to prevent misinformation, ensure comprehensive coverage, and enable effective scalability.

The adaptability to dynamic domains presents another critical limitation. Educational landscapes are constantly evolving with new pedagogical theories, research findings, and curriculum updates (Hu et al., 2023, Mzwri & Turcsányi-Szabo, 2023, Yang R. et al., 2024). Current protocols often fail to easily accommodate these rapid changes, resulting in KGs that quickly become outdated or incomplete. Furthermore, the difficulty in integrating heterogeneous and multi-modal knowledge is pronounced. Educational KGs draw from disparate sources, including structured course syllabi, unstructured research papers, and even multimedia content (Qin et al., 2020, Waagmeester et al., 2020). Ensuring semantic consistency across these varied formats, particularly when concepts are described differently across sources, presents a considerable technical hurdle.

A significant gap observed in many proposed KG models pertains to the explicit implementation of data quality and truthfulness protocols. While efforts are often directed towards robust construction methods or validating the model's utility through learning outcomes (Choi et al., 2023, Deng et al., 2021, Hu et al. 2023, Yang Y. et al. 2024), there is a frequent absence of systematic, granular fact-checking mechanisms for the KG's output itself. For instance, some models might mention general "vetting" of source material, but lack detailed, transparent frameworks for how the factual accuracy of each knowledge triplet or inferred relationship is rigorously verified against external ground truths. This omission is particularly concerning given that educational KGs frequently draw from diverse, potentially noisy, or unvalidated online sources, creating a vulnerability for

the introduction and propagation of misinformation (Li et al., 2022, Mzwri & Turcsányi-Szabo, 2023, Waagmeester et al., 2020). The lack of a universal, standardized protocol for assessing factual correctness across varied educational KGs not only leads to inconsistencies but also undermines the verifiable trustworthiness of these crucial learning resources.

Furthermore, proposed models often struggle with the inherent challenges of sparse and incomplete data in educational contexts, alongside the broader issue of scalability. Educational content can be highly specialized and fragmented, making the comprehensive construction of KGs incredibly resource intensive (Abu-Rasheed et al., 2025, Bernasconi et al. 2025, Yang Y. et al., 2022, Yang Y. 2024). While techniques like knowledge graph completion are employed (e.g., TransE in Hu et al., 2023), many models do not fully detail scalable, systematic strategies for identifying and filling data gaps in a way that can transcend specific domains or large datasets. This problem is compounded when considering the semantic accuracy across diverse subject areas as KGs grow. Maintaining consistent definitions and relationships (e.g., how "concept" is understood in mathematics versus fine arts) becomes exponentially more complex with scale. The reliance on human experts for validation (Abu-Rasheed et at., 2025, Agrawal et al., 2022, Qi et al., 2025, Yu et al., 2021), while crucial for accuracy, can become a bottleneck for achieving both comprehensive data completeness and maintaining semantic consistency in vast, evolving KGs.

From a practical standpoint, cost-effectiveness is a major impediment. The extensive manual effort required for data collection, curation, and validation—an inherently continuous process for a comprehensive educational KG—renders thorough quality control prohibitively expensive and often impractical for many institutions. Finally, a significant systemic issue is the lack of standardized metrics and benchmarks. The absence of universally accepted formats or ontologies for educational KGs, unlike in more established domains such as medicine, means there is no common ground to objectively compare the effectiveness of different input control protocols or to seamlessly integrate KGs across diverse educational platforms. This fundamental lack of standardization not only hinders the development of a cohesive educational knowledge ecosystem but also contributes to broader challenges of limited interoperability.

These identified gaps underscore the critical need for standardized protocols for educational KGs. Such a framework should provide clear guidelines for data collection and ingestion, including handling multi-modal content and varied sources; robust ontology and schema design to manage sparsity and ensure semantic consistency; and rigorous knowledge extraction and fusion methodologies to improve completeness and accuracy. Crucially, these protocols must encompass multi-layered validation and fact-checking processes, explicitly integrating mechanisms for external ground-truth verification and for combating issues like LLM hallucination, drawing on advancements such as Knowledge-Aware Validation (Yang R. et al., 2024).

Finally, a comprehensive standard would address scalability considerations from initial design to ongoing maintenance and integrate clear user feedback loops that directly contribute to continuous data refinement and quality enhancement, thereby fostering truly reliable and effective educational KGs.

6 Conclusion

This paper presented a systematic literature review that highlighted key insights from our analysis, particularly regarding the need for high-quality, verifiable knowledge within educational Knowledge Graphs (KGs). Our comprehensive evaluation of current methodologies reveals significant limitations in addressing the unique data challenges inherent in educational contexts. Existing protocols frequently fall short in adequately handling the sheer scale, remarkable diversity, and dynamic nature of educational data.

The full promise of educational KGs is currently constrained by several critical weaknesses, primarily concerning data quality. Key challenges include ensuring the completeness, consistency, and accuracy of information, particularly when managing the extensive and dynamically evolving datasets characteristic of educational environments. KGs frequently contend with sparse and incomplete data, where some learning concepts are exhaustively detailed while others lack any coverage, directly compromising the efficacy of downstream applications such as personalized learning paths. Furthermore, a persistent challenge lies in verifying the truthfulness and semantic accuracy of knowledge extracted from potentially noisy or unvalidated sources.

These pervasive data quality issues, compounded by the computational burden of applying rigorous quality controls to large and dynamic KGs, underscore a pressing need. To fully unlock the transformative potential of KGs in education, the development and adoption of standardized protocols are urgently required. Such protocols would guide the entire lifecycle of educational KGs, ensuring their truthfulness, mitigating the risks of misinformation, enhancing data quality and completeness, and ultimately improving their overall utility and scalability.

Acknowledgements

The research has been co-funded by the University of Rijeka under projects “uniri-drustv- 23-230 3224.

References

Abu-Rasheed, H., Jumbo, C., Al Amin, R., Weber, C., Wiese, V., Obermaisser, R. and Fathi, M. (2025, April). Llm-assisted knowledge graph completion for curriculum and domain modelling in personalized higher education recommendations. In *2025 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1-5). IEEE.

Abu-Salih, B. and Alotaibi, S. (2024). A systematic literature review of knowledge graph construction and application in education. *Heliyon*, *10*(3).

Abu-Salih, B. (2021). Domain-specific knowledge graphs: A survey. *Journal of Network and Computer Applications*, *185*, p.103076.

Agrawal, G., Deng, Y., Park, J., Liu, H. and Chen, Y.C. (2022). Building knowledge graphs from unstructured texts: Applications and impact analyses in cybersecurity education. *Information*, *13*(11), p.526.

Agrawal, G., Kumarage, T., Alghamdi, Z. and Liu, H. (2023). Can knowledge graphs reduce hallucinations in llms?: A survey. *arXiv preprint arXiv:2311.07914*.

Bernasconi, E., Redavid, D. and Ferilli, S. (2025). Enhancing Personalised Learning with a Context-Aware Intelligent Question-Answering System and Automated Frequently Asked Question Generation. *Electronics*, *14*(7), p.1481.

Chen, L., Li, G., Ma, B., Tang, C. and Yamada, M., (2024). A Three-Step Knowledge Graph Approach Using LLMs in Collaborative Problem Solving-Based STEM Education. *International Association for Development of the Information Society*.

Chen, Z., Zhang, Y., Fang, Y., Geng, Y., Guo, L., Chen, X., Li, Q., Zhang, W., Chen, J., Zhu, Y. and Li, J. (2024). Knowledge graphs meet multi-modal learning: A comprehensive survey. *arXiv preprint arXiv:2402.05391*.

Choi, H., Lee, H. and Lee, M. (2023). Optimal knowledge component extracting model for knowledge-concept graph completion in education. *IEEE Access*, *11*, pp.15002-15013.

Deng, Y., Zeng, Z., Jha, K. and Huang, D. (2021). Problem-based cybersecurity lab with knowledge graph as guidance. *Journal of Artificial Intelligence and Technology*.

Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., Metropolitansky, D., Ness, R.O. and Larson, J. (2024). From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*.

Elkaimbillah, Z., Rhanoui, M., Mikram, M. and El Asri, B. (2021, September). Comparative study of knowledge graph models in education domain. In *Proceedings of the Internationa L Conference on Bigdata, Modelling and Machine Learning*.

Fan, X., Sun, Y., Wang, Q. and Wang, X. (2024, December). Research on the Application of Knowledge Graph in Higher Education of Art and Design. In *2024 6th International Conference on Literature, Art and Human Development (ICLAHD 2024)* (pp. 494-501). Atlantis Press.

Fettach, Y., Ghogho, M. and Benatallah, B. (2022). Knowledge graphs in education and employability: A survey on applications and techniques. *IEEE Access*, *10*, pp.80174-80183.

Gligorea, I., Cioca, M., Oancea, R., Gorski, A.T., Gorski, H. and Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. *Education Sciences*, *13*(12), p.1216.

Hou, B., Lin, Y., Li, Y., Fang, C., Li, C. and Wang, X. (2025). KG-PLPPM: A Knowledge Graph-Based Personal Learning Path Planning Method Used in Online Learning. *Electronics*, *14*(2), p.255.

Hu, X., Zhou, Y. and Qi, J. (2023, May). Research on the teaching model of knowledge graph for professional introduction course based on OBE. In *2023 35th Chinese Control and Decision Conference (CCDC)* (pp. 1928-1933). IEEE.

Li, N., Shen, Q., Song, R., Chi, Y. and Xu, H. (2022). MEduKG: A deep-learning-based approach for multi-modal educational knowledge graph construction. *Information*, *13*(2), p.91.

Mzwri, K. and Turcsányi-Szabo, M. (2023). Internet wizard for enhancing open-domain question-answering chatbot knowledge base in education. *Applied Sciences*, *13*(14), p.8114.

Paneque, M., del Mar Roldán-García, M. and García-Nieto, J. (2023). e-LION: Data integration semantic model to enhance predictive analytics in e-Learning. *Expert Systems with Applications*, *213*, p.118892.

Qi, C., Jia, L., Wei, Y., Jiang, Y.H. and Gu, X. (2025). IntelliChain: An Integrated Framework for Enhanced Socratic Method Dialogue with LLMs and Knowledge Graphs. *arXiv preprint arXiv:2502.00010*.

Qin, Y., Cao, H. and Xue, L. (2020, August). Research and Application of Knowledge Graph in Teaching: Take the database course as an example. In *Journal of Physics: Conference Series* (Vol. 1607, No. 1, p. 012127). IOP Publishing.

Qu, K., Li, K.C., Wong, B.T., Wu, M.M. and Liu, M. (2024). A survey of knowledge graph approaches and applications in education. *Electronics*, *13*(13), p.2537.

Tian, L., Zhou, X., Wu, Y.P., Zhou, W.T., Zhang, J.H. and Zhang, T.S. (2022). Knowledge graph and knowledge reasoning: A systematic review. *Journal of Electronic Science and Technology*, *20*(2), p.100159.

Waagmeester, A., Stupp, G., Burgstaller-Muehlbacher, S., Good, B.M., Griffith, M., Griffith, O.L., Hanspers, K., Hermjakob, H., Hudson, T.S., Hybiske, K. and Keating, S.M. (2020). Wikidata as a knowledge graph for the life sciences. *Elife*, *9*, p.e52614.

Weichselbraun, A., Waldvogel, R., Fraefel, A., van Schie, A. and Kuntschik, P. (2022). Building knowledge graphs and recommender systems for suggesting reskilling and upskilling options from the web. *Information*, *13*(11), p.510.

Xue, H. (2023, September). Development History and Frontier Trends of Natural Language Processing Technology in Education: A Knowledge Graph-Based Visualization Analysis. In *Proceedings of the 15th International Conference on Education Technology and Computers* (pp. 436-443).

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y. and Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), pp.90-112.

Yang, R., Yang, B., Ouyang, S., She, T., Feng, A., Jiang, Y., Lecue, F., Lu, J. and Li, I. (2024). Graphusion: Leveraging large language models for scientific knowledge graph fusion and construction in nlp education. *arXiv preprint arXiv:2407.10794*.

Yang, Y. (2024). Research on the digital construction mode of college music teaching curriculum based on knowledge mapping. Applied Mathematics and Nonlinear Sciences. 9. 10.2478/amns-2024-0571.

Yang, Y., Huang, C., Xia, L. and Li, C. (2022, July). Knowledge graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 1434-1443).

Yu, H., Chen, Y., Wang, X., Hu, S. and Chang, L. (2021, May). A capability-oriented method for integration of liberal arts knowledge graph in primary education. In *2021 International Conference on Computers, Information Processing and Advanced Education (CIPAE)* (pp. 342-346). IEEE.

Designing Innovation through Productive Disruptions

Tamara Ćurlin, Jasmina Pivar
Virtual Reality as a Strategic Tool: Adoption Patterns and Distribution Channels in Croatian Hotels

Metodija Mircev, Violeta Cvetkoska, Igor Ivanovski
Smart Synergies: Open Innovation and Predictive Models in the Telecom-Insurance Ecosystem

Aldo Valencia, Violeta Cvetkoska, Bojan Kitanovikj, Alison Pearce
Tactile Collaboration and Backcasting to Design Meaningful Futures and Strategies

Domagoj Tuličić, Snježana Križanić
Digital Twin-Enhanced Fire Response Through Unmanned Aerial Reconnaissance System

Virtual Reality as a Strategic Tool: Adoption Patterns and Distribution Channels in Croatian Hotels

Tamara Ćurlin, Jasmina Pivar
Faculty of Economics and Business Zagreb
Trg J.F. Kennedy 6,
10 000 Zagreb, Croatia
{tcurlin, jpivar}@efzg.hr

Abstract. *Hotel industry is one of the first to adopt virtual reality into websites to enhance customer experience, gain competitive advantage and strengthen their position relative to online travel agencies by increasing the share of direct bookings. The purpose of this research is to provide an overview of how Croatian hotels are utilizing VR on their websites. Based on sample of 201 hotels, the study examined hotel features to gain deeper insights into adoption patterns. Chi-square analysis was used to assess the relationships between hotel characteristics and VR usage. Findings confirm significant relationship between the direct booking ratio and VR implementations.*

Keywords. hotel industry, booking strategies, VR in hotels, ICT, data visualization, Chi square, hotel business strategy

1 Introduction

In recent years, the hospitality industry has undergone rapid digital transformation, driven by the increasing demand for personalized, immersive, and interactive online experiences (Collins, 2023). Among emerging technologies, virtual reality (VR) has gained significant attention as a tool for enhancing customer engagement and influencing booking behavior (Fransisco, 2024).

VR offers potential guests an experience which mimics experience in the hotel that can influence perception, trust, and purchase intention by allowing users to explore hotel environments virtually, through realistic, three-dimensional simulations (Israel et al., 2019). The competitive nature of the hotel industry forces hotels to adopt innovative technologies, especially in destinations such as Croatia where tourism is a major economic sector (Slivar et al., 2016). While many global hotel brands have already embraced VR, little is known about its adoption in regional markets or how implementation varies across different types of hotels (Lodhi et al., 2024). Moreover, limited empirical research has examined how VR usage relates to booking strategies, especially direct bookings, which are increasingly prioritized by hotels aiming to reduce commission costs and improve customer relationships (Zeqiri, 2025).

The purpose of this paper is to provide a comprehensive overview of the use of VR in Croatian hotel websites. The study examines 201 hotels across Croatia analyzing various hotel features such as rating, location, ownership, and size with the purpose to explore potential correlations between these characteristics and the adoption of VR technology on their websites.

To the authors' knowledge, this is the first investigation of its kind, offering a foundation for future research that could be extended to other countries and regions for comparative analysis.

The research employs descriptive statistics, and Chi-square analysis to examine the dataset and extract meaningful insights. Additionally, the study investigates the relationship between hotel distribution channel strategies with the focus on the dominant distribution channel and direct booking ratio and VR usage. The findings of this study not only contribute to academic knowledge but also provide practical implications for hotel management, serving as a basis for developing effective strategies to leverage VR technology in enhancing direct bookings and competitive positioning.

The paper is organized as follows: After the Introduction section where papers objectives and context were described, the Methodology section is presented. Methodology section describes research instrument, research sample and the methods used for achieving the research goal. Furthermore, the Result section provide a statistical investigation output which were explained in the Discussion section. Lastly, the Conclusion section presents main contributions, further investigations and limitations of the research.

2 Literature Review

2.1 Virtual reality

The concept and significance of virtual reality (VR) have been widely discussed since the early 1990s, when the most used definition was first developed. Steuer (1992) defines VR as an environment that induces telepresence through stimulation, where vividness and interactivity enhance the observer's sense of presence. More recent definitions describe VR as the navigation and interaction within a computer-generated, three-dimensional environment, designed to simulate real-time sensory experiences (Yung & Khoo-Lattimore, 2019). A key feature of VR is its ability to allow users to interact with and navigate the virtual space, enhancing immersion (Kim et al., 2025)

The reason of VR technology's prevalence is because it engages users cognitively, sparking their interest and providing useful knowledge that can aid in selling products and services (De Gauquier et al., 2019). The immersive nature of VR often leads to stronger emotional responses, as users feel more involved in the experience (Riva et al., 2007). Additionally, VR enhances sensory experiences, particularly through movement and immersion, which contribute to more engaging bodily sensations. These deeper experiences tend to result in positive attitudes and behaviors toward the content (Gibson, 2017).

2.2 Virtual reality in hotel industry

In the hotel industry, travellers often rely heavily on visual content and online reviews on their websites (Fidzani & Read, 2014). Images of rooms and key services and amenities play a critical role in shaping perceptions and influencing booking decisions (Bogicevic et al., 2017). Consequently, how hotels visually communicate their brand identity, style, and service offerings has become a vital source of competitive advantage (Chung, 2017). Therefore, VR has emerged as a powerful marketing tool that allows potential guests to engage experiences of hotel environments before making a reservation (Mc Lean et al., 2022). Besides, a significant challenge for the hotel industry is its reliance on Online Travel Agencies (OTAs), which charge high commissions, reduce brand visibility, and weaken direct customer relationships and loyalty (Ćurlin, 2022). This dependence forces hotels to compete digitally while maintaining profitability and brand strength. At the same time, the intangible and perishable nature of hospitality services, such as product that cannot be evaluated before consumption, makes effective online presentation crucial (Kandampully & Solnet, 2024). The rise of e-commerce has expanded these opportunities, yet consumer expectations have evolved, demanding personalized, multichannel digital experiences (Ntumba et al, 2023). Users now expect not only platforms that offer more than just basic information and simple applications, but sophisticated, consumer-centred applications across multiple channels and technology platforms (Lam & Law, 2019).

VR has been increasingly incorporated into the marketing strategies of many hotels, leading to its growing popularity within the hospitality and tourism sectors (Guerra et al., 2015). Marriott International was among the first hotel chains to adopt VR for marketing purposes, successfully integrating it into their product development and brand promotion efforts (Mandelbaum, 2018). Recent studies have identified key VR applications in hospitality and tourism, including promotion and advertising, planning, management, entertainment, education, accessibility, and heritage conservation (Siddiqui et al., 2022). VR is widely recognized as a major technological advancement influencing the travel and tourism industry (Tussyadiah et al., 2018). Numerous leading travel companies, such as Virgin, Disney, Emirates, and hotel brands like Marriott and Hyatt Regency, have implemented VR to allow customers to "experience before they buy" offering significant advantages by engaging consumers prior to their actual visit. In the hotel industry, VR has emerged as a powerful tool for enhancing customer engagement by offering immersive and interactive visual experiences. VR enables precise control over key visual elements such as lighting, object placement, and spatial layout, allowing hotels to create realistic and emotionally compelling environments (Soranzo & Wilson, 2014). Unlike traditional 2D or 3D platforms that limit user interaction to basic clicks or rotations (Choi & Taylor, 2014), VR immerses users in hotel spaces through multisensory experiences, including spatial navigation and sound (De Gauquier et al., 2018). This heightened sense of presence helps potential guests explore rooms, lobbies, or amenities virtually, encouraging emotional connection and stronger purchase intent (Brakus et al., 2009). In particular, 360° immersive VR (IVR) has become widely used in hotel marketing strategies, offering panoramic, photo- and video-based virtual tours of hotel interiors and surroundings. These tours provide credible and detailed representations of real-world environments, improving perceived authenticity and engagement (Wu & Lai, 2022). Consequently, 360° videos are now a key promotional asset for hotels aiming to differentiate themselves and attract bookings in an increasingly competitive digital marketplace.

Scholars have explored the potential of VR in hotel industry and hospitality for nearly thirty years. First researches were focused on highlighting VR's ability to present visually rich destination landscapes and support sustainable tourism. For instance, Cheong (1995) did a conceptual study on how far VR technology can actually deliver as a tool to increase destinations appeal, and Hobson & Williams (1995) highlighted VR as the most important technology breakthrough for hotel and hospitality industry in late 20-th century. Following research demonstrated VR's positive influence on destination image, branding,

advertising effectiveness, and visit intention (Berger et al., 2006; Giordimaina, 2008). Early empirical studies employed VR to simulate tourism settings and examine consumer responses, such as reactions to crowding and emotional responses to multistage waiting in virtual environments (Hwang et al., 2012; Kim & Kim, 2016). As VR technology advanced, researches it also used in combination with eye-tracking to study consumer attention and brand perception in hotels (Bigne et al., 2016). Moreover, VR has been shown to enhance servicescape design and support more immersive decision-making processes by offering experiential and realistic information (Han et al., 2018). With the commercialization of VR devices like Oculus, attention shifted toward user acceptance and the technology's influence on behavioral intentions, including purchase and recommendation (Griffin et al., 2017). While many studies affirm VR's ability to foster a sense of presence and increase travel intentions, relatively few have examined the mechanisms by which technical and sensory features shape user experiences (Kim et al., 2020). Recent reviews have identified significant gaps, particularly the need for more theory-driven investigations into usability, psychological outcomes, and the experiential qualities of VR travel (Shukla et al., 2024, Lampropoulos et al., 2024).

3 Methodology

Descriptive statistical analysis was employed to provide deeper insight into the data structure. Furthermore, Chi-square analysis was conducted to examine the statistical relationships between variables. The aim of this paper is to investigate how many Croatian hotels have adopted VR technologies on their websites, the distribution of adoption across different hotel types, and whether there is a significant relationship between hotel features and the VR adoption rate. Additionally, the study explores the relationship between preferred distribution channels, direct booking ratios, and VR implementation to determine whether hotel booking strategies are associated with the adoption of VR technologies on hotel websites.

The study was conducted on a sample of 201 hotels in Croatia, utilizing data on VR usage on hotel websites, hotel features, and booking strategies. Data collection was carried out in two stages: first, through a questionnaire administered to hotel employees, followed by an in-depth analysis of hotel websites. The data used in this paper were collected as part of a doctoral dissertation (Ćurlin, 2022).

Table 1 presents the research instrument, including all data utilized in this study. The variables listed in the table were coded and referenced using these codes throughout the paper. Additionally, the modalities of each variable are also provided. The variable code VR indicates whether a hotel's website includes virtual reality content, where a value of 0 denotes the absence of VR and 1 indicates the presence of VR. Following hotel features were analysed: hotel rating, measured on a scale from two to five stars; hotel location, where inland hotels are coded as 0 and seaside hotels as 1; and ownership, with domestically owned hotels coded as 1 and foreign-owned hotels as 2. The variable size refers to hotel size, categorized into four groups ranging from the largest (more than 250 rooms) to the smallest (fewer than 10 rooms). Regarding distribution channels, the variable code channel represents the preferred booking channel, categorized into direct bookings, online travel agencies, foreign agencies, and domestic agencies. Finally, the direct variable refers to the direct booking ratio, expressed in percentages and grouped into ten categories.

Table 1. Research instrument

Variable code	Variable description	Modalities
VR	VR present on hotel website	0-no 1-yes
	Hotel features	
RATING	Number of stars	1-2*, 2-3*, 3-4*, 4-5*
LOCATION	Hotel location	0-land 1-sea
OWNERSHIP	Hotel ownership	1-domestic 2-foregin
SIZE	Total number of rooms	1->250, 2-100-250, 3-51-100, 4-11-50, 5-<10
	Distribution channels	
CHANNEL	Preferred distribution channel	1- direct 2-OTA 3-foregin agencies 4-local agencies
DIRECT	Direct booking ratio (%)	1-0-10, 2-11-20, 3-21-30, 4-31-40, 5-41-50, 6-51-60, 7-61-70, 8-71-80, 9-81-90, 10-98-100

Source: Authors work (2025)

The research sample included 201 Croatian hotels, along with the variables described in the previous section. Fig. 1 presents data on the presence of VR technology on hotel websites within the sample. Out of the total 201 hotels, 135 (67.16%) do not include VR on their websites, while 66 hotels (32.84%) have implemented VR technologies.

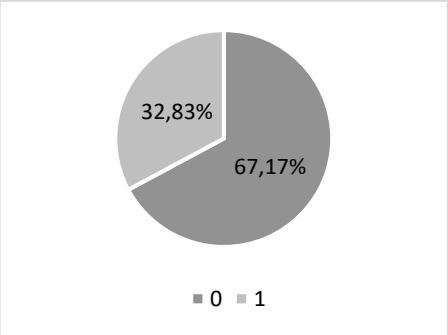


Figure 1. VR usage in Croatian hotels website; Source: Authors work (2025)

Table 2 provides an overview of the hotel features included in this study. Variable frequencies, structure, and cumulative distributions are presented. The largest group of hotels is categorized as four-star, comprising 55.33% of the sample (a total of 109 hotels). The second most represented category is three-star hotels, accounting for 23.35%, followed by five-star hotels at 17.77%. Only seven two-star hotels are included in the study. Most hotels are located by the seaside, totaling 73.60%, which aligns with the overall structure of the Croatian hotel population. The remaining 26.40% of hotels are located inland. Most hotels in the sample are domestically owned, with 66.50% owned by domestic individuals or groups, while 33.50% are foreign-owned. Regarding hotel size, 48.76% of the sample comprises large hotels with more than 250 rooms, 45.27% have between 100 and 250 rooms, and only 5.96% have fewer than 100 rooms.

Table 2. Hotel features

Rating	Frequency	Structure (%)	Cumulative (%)
1	7	3,553	3,553
2	46	23,350	26,904
3	109	55,330	82,234
4	35	17,766	100,000
Location	Frequency	Structure (%)	Cumulative (%)
0	52	26,396	26,396
1	145	73,604	100,000
Ownership	Frequency	Structure (%)	Cumulative (%)
0	131	66,497	66,497
1	66	33,503	100,000
Size	Frequency	Structure (%)	Cumulative (%)
1	98	48,756	48,756
2	91	45,274	94,030
3	5	2,488	96,517
4	7	3,483	100
Total	201	100	

Source: Authors work (2025)

Table 4 presents the structure of hotel distribution channels. Most hotels in the sample identified online travel agencies as their dominant distribution channel, accounting for 93 hotels or 46.27%. Direct booking was the primary distribution channel for 27.86% of the hotels, while a slightly smaller proportion, 24.38%, reported foreign agencies as their main channel. Only 1.49% of the hotels included in the study listed local agencies as their dominant distribution channel. The direct booking ratio refers to reservations made through a hotel's official website, by phone, email, or on-site. Within the sample, 45 hotels (22.39%) reported that direct bookings formed between 21–30% of their total reservations. Moreover, 39 hotels (19,40%) had a direct booking share between 11–20%, and 32 hotels (15.92%) reported between 31–40%. Collectively, 57.51% of the hotels surveyed had direct booking rates ranging from 11% to 40%. Furthermore, 9,45% of hotels reported 61–70% of their bookings coming through direct channels, and a total of 32 hotels (8.96% each) had either 0–10% or 41–50% direct reservations. Another 14 hotels (6.97%) fell within the 51–60% range. Only 8 hotels (3.98%) reported direct bookings in the ranges of 71–80% and 81–90%, respectively. Notably, none of the hotels in the sample reported more than 90% of bookings via direct channels.

Table 3. Hotel distribution channel structure

Channel	Frequency	Structure (%)	Cumulative (%)
1	56	27,861	27,861
2	93	46,269	74,129
3	49	24,378	98,507
4	3	1,493	100
Direct	Frequency	Structure (%)	Cumulative (%)
1	18	8,955	8,955
2	39	19,403	28,358
3	45	22,388	50,746
4	32	15,920	66,667
5	18	8,955	75,622
6	14	6,965	82,587
7	19	9,453	92,040
8	8	3,980	96,020
9	8	3,980	100
Total	201	100	-

Source: Authors work (2025)

4 Results

The following section presents a descriptive analysis alongside the results of a Chi-square test. This analysis aims to investigate the relationships between VR utilization in Croatian hotels and various hotel characteristics, as well as the distribution channel structure, with the objective of uncovering new insights on the topic.

4.1 Descriptive statistics results

Table 4 presents the results of the descriptive statistics for the use of VR on hotel websites, hotel feature variables (rating, location, ownership, and size), and distribution channel structure variables (dominant distribution channel and direct booking ratio). Descriptive statistics reported include the mean, median, mode, standard deviation, variance, skewness, kurtosis, and their respective standard errors. The results suggest an adequate dispersion of the data, with standard deviation values ranging from 0.473 to 2.16, indicating the representativeness of the sample. The mean and median values suggest that most hotels in the sample do not include VR on their websites, are large, domestically owned, located on the seaside, and categorized with four stars. The results also indicate that online travel agencies (OTAs) are the dominant distribution channel, and that the average direct booking ratio for Croatian hotels falls between 32% and 40%.

4.2 Chi-square analysis results

Finally, the results of the Chi-square analysis are presented. The Chi-square test is a nonparametric statistical method commonly employed in research

involving nominal variables (Adhikari et al., 2023). This analysis not only assesses the significance of differences between variables but also identifies the specific categories contributing to these differences (Beillouin et al., 2021). In this study, the Chi-square test is utilized to examine the relationships between VR usage on hotel websites, hotel characteristics, and distribution channels.

Table 5 presents the results of the Chi-square analysis examining the relationship between VR usage on Croatian hotel websites and various hotel features. The analysis reveals a statistically significant association between the rating variable and VR usage at the 10% significance level, with a Chi-square value of 6.270. Cramér's V coefficient of 0.560 supports a strong relationship between these variables. Additionally, a statistically significant relationship was found between ownership and VR usage at the 5% significance level, with a Chi-square value of 4.700 and a Cramér's V coefficient of 0.154, suggesting a weak association. In contrast, no statistically significant relationship was identified between location and VR usage, as evidenced by a Chi-square value of 1.370 and a Cramér's V coefficient indicating no meaningful association. Similarly, the analysis did not reveal a statistically significant relationship between hotel size (measured by number of rooms) and VR usage, with a Chi-square value of 3.110. However, the corresponding Cramér's V value of 0.126 suggests a potential weak relationship, which may warrant further investigation in future research.

5 Discussion

The following section summarizes the key findings presented in this paper. Table 6 provides an overview of the Chi-square analysis results related to VR usage on hotel websites, hotel features (including rating, location, ownership, and size), and hotel distribution channel strategy (including dominant distribution channel and direct booking ratio). The analysis confirmed statistically significant relationships between hotel rating and ownership with VR usage in the hotel features category. Regarding hotel rating, the findings suggest that higher-category hotels are more likely to implement VR on their websites. For hotel ownership, when normalized values are considered, it can be concluded that foreign-owned hotels are more likely to adopt VR technologies, implying that domestic hotel owners should consider adopting modern technological solutions to remain competitive. The variables *location* and *hotel size* did not show statistically significant associations with VR usage, suggesting that the adoption of VR is not dependent on whether the hotel is located inland or on the seaside, nor on the number of rooms. However, further investigation into hotel size is warranted, as the Cramér's V coefficient indicates a potential weak relationship between hotel size and VR adoption.

Table 4. Descriptive analysis results

Variable	N	Mean	Median	Mode	Standard deviation	Variance	Skewness	Std. error skewness	Kurtosis	Std. error kurtosis
VR	201	0,34	0	0.00	0,473	0,224	0,704	0,173	-1,52	0,345
Rating	201	2,86	3	3.00	0,735	0,540	-0,309	0,172	-0,0507	0,341
Location	201	0,73	1	1.00	0,444	0,197	-1,05	0,172	-0,903	0,341
Ownership	201	1,25	1	1.00	0,433	0,188	1,17	0,172	-0,635	0,341
Size	201	1,61	2	1.00	0,707	0,500	1,33	0,172	2,40	0,341
Channel	201	2,00	2	2.00	0,765	0,585	0,21	0,172	-0,737	0,341
Direct	201	3,99	3	3.00	2,16	4,680	0,67	0,172	-0,456	0,341

Source: Authors work (2025)

Table 5. The Chi square analysis results for VR and hotel features

Variable	Modality	VR		n	Pearson Chi-square	df	p-value	Phi-coefficient	Cramer's V
		0	1						
rating	1	3,08%	4,48%	197	6,270	3	0,089	/	0,560
	2	24,62%	20,90%						
	3	54,62%	56,72%						
	4	17,69%	16,42%						
location	0	29,23%	22,39%	197	1,370	1	0,241	0,0835	0,083
	1	70,77%	77,61%						
ownership	0	79,23%	64,18%	197	4,700	1	0,03	0,154	0,154
	1	20,77%	35,82%						
size	1	53,85%	40,30%	197	3,110	3	0,375	/	0,126
	2	42,31%	49,25%						
	3	2,31%	2,99%						
	4	1,54%	5,97%						

Source: Authors work (2025)

Regarding the distribution channel strategy and VR usage on Croatian hotel websites, the statistical analysis confirmed a significant relationship between the direct booking ratio and the presence of VR, representing one of the most important contributions of this study. The visual representation of variable structure indicates that hotels which have implemented VR on their websites tend to have a higher proportion of direct bookings compared to those that have not. This insight may be valuable for both future research and hotel managers aiming to increase their share of direct bookings. The visual data also showed that hotels with VR tend to prefer online travel agencies (OTAs), direct bookings, and foreign agencies as their primary distribution channels. However, the dominant distribution channel variable did not show a statistically significant relationship with VR usage. Nonetheless, the Cramér's V coefficient suggests a possible weak association, indicating that further research in this area may be warranted.

Table 6. Summary results of the Chi square analysis

Variable	VR in hotel websites
hotel features	
rating	10%
location	∅
size	∅
ownership	5%
distribution channel	
channel	∅
direct	5%

Source: Authors work

6 Conclusion

Virtual reality technologies have become omnipresent across various industries, significantly enhancing business operations and marketing strategies. The hotel industry, as a leader in adopting advanced technologies, continuously seeks novel approaches to gain competitive advantage. One of the key interests in hotel industry is to increase their direct booking in order to be less dependable of the online travel agencies. Consequently, many hotel strategies prioritize strengthening direct distribution channels due to the high commission fees associated with OTAs. Previous research has established links between website quality and direct bookings; however, most studies focus on website functionality and utility. Recently, hotels have begun incorporating VR on their websites to offer potential guests an immersive preview of the hotel experience.

This paper aimed to explore the adoption of VR technologies on Croatian hotel websites, with particular emphasis on the relationship between VR usage and hotel characteristics. Chi-square analysis was employed to investigate potential associations between specific hotel features and VR presence on websites, as well as to examine the connection between hotel distribution channel strategies and VR adoption.

The paper makes several valuable contributions to both academia and hotel management. Firstly, it provides a comprehensive overview of VR implementation in Croatian hotels, representing the first study of its kind to the authors' knowledge, thus laying the groundwork for future research. Nearly one-third of the sampled hotels have implemented VR on their websites, a substantial adoption rate with room for growth. Hotels rated four stars, located on the seaside, domestically owned, and with hundred or more rooms showed the highest VR adoption. However, only hotel rating and ownership demonstrated statistically significant associations with VR usage. Higher-rated hotels and foreign-owned properties were more likely to adopt VR, suggesting that better-resourced or internationally affiliated establishments invest more in advanced digital technologies.

The most significant finding is the statistically significant relationship between VR adoption and direct booking ratio. Hotels utilizing VR tended to achieve higher direct booking levels, indicating that VR serves not only as a marketing enhancement but also as a strategic asset to reduce reliance on third-party booking platforms. Although the dominant distribution channel did not show a statistically significant link to VR adoption, observed weak associations warrant further investigation. Visual analysis confirmed that hotels with VR on their websites enjoy higher direct reservation ratios, a critical advantage for hoteliers aiming to increase direct bookings and reduce OTA dependency. Thus, VR integration on hotel websites emerges as a strategic tool for boosting direct booking performance.

These findings enrich the growing literature on digital innovation in hospitality and offer practical implications for hotel managers. By adopting VR, particularly domestically owned and mid-tier hotels can strengthen their competitive position, enhance user engagement, and potentially increase direct sales. Future research should explore the impact of hotel size on VR adoption and conduct longitudinal studies to assess whether VR continues to influence booking behavior over time. Additionally, further investigation is needed to examine other hotel features and distribution channels, ideally using larger and more diverse samples across regions and market segments to improve representativeness. Moreover, a promising direction for further research is also the potential of augmented reality (AR), which may offer more accessible and interactive experiences for users and could complement or even enhance the role of VR in hotel marketing and customer engagement.

References

Adhikari, A., Gorain, S. C., Gayen, P., Pal, I., & Sen, S. (2023). Studying the differences: A review on t-test. *International Research Journal of Education and Technology*, 5(5), 338-349.

Beillouin, D., Ben-Ari, T., Malézieux, E., Seufert, V., & Makowski, D. (2021). Positive but variable effects of crop diversification on biodiversity and ecosystem services. *Global change biology*, 27(19), 4697-4710.

Berger, H., Dittenbach, M., Merkl, D., Bogdanovych, A., Simoff, S., & Sierra, C. (2006). Opening new dimensions for e-Tourism. *Virtual Reality*, 11(2–3), 75–87.

Bigne, E., Simonetti, A., Guixeres, J., & Alcaniz, M. (2024). Visual attention and product interaction: a neuroscientific study on purchase across two product categories in a virtual store. *International Journal of Retail & Distribution Management*, 52(4), 389-406.

Bogicevic, V., Seo, S., Kandampully, J. A., Liu, S. Q., & Rudd, N. A. (2019). Virtual reality presence as a preamble of tourism experience: The role of mental imagery. *Tourism Management*, 74, 55-64.

Brakus, J. J., Schmitt, B. H., & Zarantonello, L. (2009). Brand Experience: What is It? How is it Measured? Does it Affect Loyalty? *Journal of Marketing*, 73(3), 52–68.

Cheong, R. (1995). The virtual threat to travel and tourism. *Tourism management*, 16(6), 417-422.

Choi, Y. K., & Taylor, C. R. (2014). How do 3-dimensional images promote products on the Internet? *Journal of Business Research*, 67(10), 2164–2170.

Chung, S. Y. (2017). *Managing Brand Identity Through Visual Elements: A Longitudinal Case Study of Five-Star Hotels* [Doctoral dissertation, The University of Manchester, United Kingdom].

Collins, G. (2023). Impact of Digitalization on Customer Experience Management in the Hospitality Industry. *Hospitality and Tourism Journal*, 1(1), 12-23.

Ćurlin, T. (2022). *Povezanost Organizacijskih Aspekata Poslovanja, Kvalitete Hotelskih Mrežnih Stranica i Kanala Distribucije u Hotelijerstvu* [Doctoral dissertation, University of Zagreb, Faculty of Economics and Business]. REPEFZG. https://urn.nsk.hr/urn:nbn:hr:148:724041

De Gauquier, L., Brengman, M., Willems, K., & Van Kerrebroeck, H. (2019). Leveraging advertising to a higher dimension: experimental research on the impact of virtual reality on brand personality impressions. *Virtual Reality*, 23, 235-253.

Fidzani, L. C., & Read, M. A. (2014). Identity expression and bedroom personalization by urban adolescents in Botswana. *Journal of Adolescent Research*, 29(6), 691-715.

Fransisco, T. A. (2024). The Impact of Virtual Reality (VR) Technology on Enhancing Customer Engagement in the Online Travel Industry. *Cebong Journal*, 3(3), 122-130.

Gibson, R. (2017). Full Circle: Towards VR. In *Digital Research in Humanities and Arts: Data Ache*.

Giordimaina, J. G. (2008). Multi sensory marketing and its application in tourism. *Journal of the Institute of Tourism Studies*, 1(2), 42-45.

Griffin, T., Giberson, J., Lee, S. H., Guttentag, D., Kandaurova, M., Sergueeva, K., & Dimanche, F. (2017). Virtual reality and implications for destination marketing. *Travel and Tourism Research Association: Advancing Tourism Research Globally*, 29, 34.

Guerra, J. P., Pinto, M. M., & Beato, C. (2015). Virtual reality-shows a new vision for tourism and heritage. *European Scientific Journal*, 11(9), 49-54.

Han, J., Kang, H. J., & Kwon, G. H. (2018). A systematic underpinning and framing of the servicescape: Reflections on future challenges in healthcare services. *International journal of environmental research and public health*, 15(3), 50

Hobson, J. P., & Williams, A. P. (1995a). Virtual reality: A new horizon for the tourism industry. *Journal of Vacation Marketing*, 1(2), 124–135.

Hwang, J., Yoon, S. Y., & Bendle, L. J. (2012). Desired privacy and the impact of crowding on customer emotions and approach-avoidance responses: waiting in a virtual reality restaurant. International *Journal of Contemporary Hospitality Management*, 24(2), 224-250.

Israel, K., Zerres, C., & Tscheulin, D. K. (2019). Presenting hotels in virtual reality: does it influence the booking intention?. *Journal of Hospitality and Tourism Technology*, 10(3), 443-463.

Kandampully, J., & Solnet, D. (2024). *Service Management Principles for Hospitality & Tourism in the Age of Digital Technology*. Oxford: Goodfellow Publishers.

Kim, J., Ahn, J. H., & Kim, Y. (2025). Immersive Interaction for Inclusive Virtual Reality Navigation: Enhancing Accessibility for Socially Underprivileged Users. *Electronics*, 14(5), 1046.

Kim, M. J., Lee, C. K., & Preis, M. W. (2020). The impact of innovation and gratification on authentic experience, subjective well-being, and behavioral intention in tourism virtual reality: The moderating

role of technology readiness. *Telematics and Informatics,* 49, 101349.

Kim, Y. R., & Kim, G. J. (2016). Hovr-type: Smartphone as a typing interface in VR using hovering. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (pp. 333-334).

Lam, C., & Law, R. (2019). Readiness of upscale and luxury-branded hotels for digital transformation. *International Journal of Hospitality Management*, 79, 60-69.

Lampropoulos, G., & Kinshuk. (2024). Virtual reality and gamification in education: a systematic review. *Educational technology research and development*, 72(3), 1691-1785.

Lodhi, R. N., Asif, M., Del Gesso, C., & Cobanoglu, C. (2024). Exploring the critical success factors of virtual reality adoption in the hotel industry. *International Journal of Contemporary Hospitality Management*, 36(11), 3566-3586.

Mandelbaum, E. (2018). Seeing and conceptualizing. *Philosophy and Phenomenological Research*, 97(2), 267-283.

Ntumba, C., Aguayo, S., & Maina, K. (2023). Revolutionizing retail: a mini review of e-commerce evolution. *Journal of Digital Marketing and Communication*, 3(2), 100-110.

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., ... & Alcañiz, M. (2007). Affective interactions using virtual reality: the link between presence and emotions. *Cyberpsychology & behavior*, 10(1), 45-56.

Shukla, V., Rana, S., & Prashar, S. (2024). Examining the potential of virtual and augmented reality in enhancing tourism experiences. *The Bottom Line*. doi: 10.1108/BL-05-2023-0139

Siddiqui, M. S., Syed, T. A., Nadeem, A., Nawaz, W., & Alkhodre, A. (2022). Virtual tourism and digital heritage: an analysis of VR/AR technologies and applications. *International Journal of Advanced Computer Science and Applications*, 13(7), 303-315.

Slivar, I., Božić, S., & Batković, A. (2016). Innovation in tourism: Perception of tourism providers from Croatia and Serbia. *Review of Innovation and Competitiveness: A Journal of Economic and Social Research*, 2(4), 37-56.

Soranzo, A., & Wilson, C. (2014). Virtual environments in visual perception: applications and challenges. In: Cipresso, P. and Serino, S., (eds.) *Virtual Reality: Technologies, Medical Applications and Challenges. Psychology Research Progress*. Nova Science, 73-80. [Book Section]

Steuer, J. (1992). De ning virtual reality: Dimensions determining telepresence. *Journal of communication*, 42(4), 73-93.

Tussyadiah, I. P., Wang, D., Jung, T. H., & Tom Dieck, M. C. (2018). Virtual reality, presence, and attitude change: Empirical evidence from tourism. *Tourism management*, 66, 140-154.

Wu, X., & Lai, I. K. W. (2022). How A 360° virtual tour is more effective than photographs on communication effects: the roles of mental imagery processing and a sense of presence. *Current Issues in Tourism*, 26(23), 3813–3830.

Yung, R., & Khoo-Lattimore, C. (2019). New realities: a systematic literature review on virtual reality and augmented reality in tourism research. *Current issues in tourism*, 22(17), 2056-2081.

Zeqiri, A., Ben Youssef, A., & Maherzi Zahar, T. (2025). The Role of Digital Tourism Platforms in Advancing Sustainable Development Goals in the Industry 4.0 Era. *Sustainability*, 17(8), 3482.

Smart Synergies: Open Innovation and Predictive Models in the Telecom-Insurance Ecosystem

Metodija Mircev
Chief Executive Officer, A1 Makedonija
Presveta Bogorodica Square, No. 1, 1000 Skopje,
North Macedonia
mircev.metodija@gmail.com

Violeta Cvetkoska, Igor Ivanovski
Ss. Cyril and Methodius University in Skopje, Faculty of Economic-Skopje
Blvd. Goce Delchev 9V, 1000 Skopje,
North Macedonia
{vcvetkoska,
igor.ivanovski}@eccf.ukim.edu.mk

Abstract. *This study examines the convergence of the telecommunications and insurance industries through the lens of open innovation and digital transformation, using predictive analytics to uncover customer behavior patterns. The aim is to understand how integrating telecom services with financial and insurance products impacts customer churn, revenue generation, and market segmentation. We employ a quantitative methodology on a real-world convergent telecom dataset, applying logistic regression to predict customer churn, multiple linear regression to identify revenue drivers, and cluster analysis to segment customers, all implemented in Microsoft Power BI. The findings reveal that customer churn is significantly influenced by service bundling depth, acquisition channels, and use of value-added services, while monthly revenue is driven primarily by base service pricing, multi-service adoption, and customer tenure. Unsupervised clustering uncovered three distinct customer segments with varying risk profiles: a segment of high-value yet high-churning customers expecting greater value, a segment of minimally engaged customers with extreme churn rates, and a segment of long-tenured, mid-/high-tier customers lacking cross-service engagement. These results highlight critical implications for industry convergence strategy: multi-service bundles and open innovation partnerships can enhance value but must be aligned with customer expectations to improve retention and lifetime value. The study contributes to literature by providing an empirical, data-driven perspective on cross-industry digital ecosystem strategies. It offers actionable insights for practitioners on tailoring retention efforts and product offerings in an era of blurring industry boundaries.*

Keywords. industry convergence, telecommunications, insurance services, open innovation, predictive analytics.

1 Introduction

The telecommunications and insurance sectors are increasingly intertwining as firms extend beyond traditional industry boundaries. Spurred by digital transformation, financial services are now delivered through broad ecosystems of diverse players co-evolving via collaboration and competition (Baumann et. al., 2023). Telecom operators and banks, for example, have been converging services for years, leveraging technological advancements and regulatory trends (e.g. open banking) that lower barriers to cross-industry offerings. This phenomenon of industry convergence is characterized by telecom companies offering financial and insurance products, and conversely, banks and insurers embracing telecom-style digital platforms. The drivers include shared technological trends (cloud, mobile, data analytics) and common challenges around data privacy and customer experience (Research and Markets, 2024). In essence, digital transformation has blurred industry lines, pushing firms to form new partnerships and value networks.

Open innovation provides a strategic framework for this convergence. Coined by Chesbrough (2006), open innovation involves leveraging purposive inflows and outflows of knowledge to accelerate internal innovation and expand markets. In practice, this means companies partner with external organizations to co-create services and access new capabilities. In the telecom and finance context, open innovation manifests as telecoms collaborating with fintech or insurtech startups, and banks opening their APIs to third-party developers. Industry analysts argue that banks must adopt an *open innovation model with third parties*, analogous to how telecoms innovated with over-the-top (OTT) partners (Kazmi, 2021). By teaming up, each sector can remain innovative and relevant in the digital era. Indeed, much of the recent innovation in telecom is introduced via partnerships rather than in-house development (Kazmi, 2021). This shift is exemplified by initiatives like open banking and open insurance, where sharing data and services across firm boundaries is encouraged to spur innovation and new business models (Baumann et. al., 2023). Such cross-sector collaboration is rapidly becoming imperative rather than optional.

Market trends underscore the momentum of convergence. Telecom operators are expanding their digital portfolios to include financial and insurance services, via both proprietary platforms and strategic partnerships (Research and Markets, 2024). Banks and insurers seek telecom partners to enhance digital distribution and access real-time data streams (for example, telecom data used in credit scoring or pricing insurance). Convergence offers mutual benefits: telcos gain new revenue streams, and banks/insurers gain innovative channels and customer insights. However, it also introduces strategic challenges, telecoms must manage services outside their core expertise, and financial institutions must ensure regulatory compliance and customer trust when partnering with non-bank actors (Kazmi, 2021).

Despite the growing prevalence of such partnerships, academic research has only begun to explore their implications. Prior studies have extensively examined customer analytics within individual industries, for instance, churn prediction in telecom or customer segmentation in banking, but integrated analyses across telecom and insurance remain scarce. The convergence of these industries raises new questions: *How do bundled cross-industry services affect customer loyalty (churn) and revenue? Can data analytics identify distinct customer segments in a combined service ecosystem?* Addressing these questions is important for both scholars (to understand emerging business ecosystem dynamics) and practitioners (to optimize convergence strategies).

This paper aims to fill this gap by studying a real-world case of industry convergence through a data-driven approach. We leverage customer data from a telecom operator in North Macedonia that has incorporated financial/insurance offerings into its service bundle, a prime example of open innovation in action. Using predictive modeling techniques (logistic regression, linear regression, and clustering) within a business intelligence platform (Power BI), we analyze customer churn drivers, revenue factors, and segmentation in this converged-service context.

The contributions of this paper are threefold: 1) Conceptual: we bridge the literature on open innovation and digital transformation with industry convergence, proposing a framework in which telecom and insurance form a collaborative ecosystem to co-create value. We highlight how predictive analytics serves as a linking mechanism between strategic convergence initiatives and customer-level outcomes; 2) Methodological: we demonstrate an integrated application of predictive modeling in an *open innovation ecosystem context*. Specifically, we show how logistic regression (for churn prediction), multiple linear regression (for revenue estimation), and cluster analysis (for customer segmentation) can be applied using Power BI's AI-powered tools. This approach illustrates the feasibility of advanced analytics in accessible BI software, which is valuable for practitioners outside traditional data science environments; 3) Empirical: we provide insights based on a unique dataset spanning telecom services and insurance add-ons. The analysis identifies key factors that influence churn and revenue in a converged setting (e.g. product bundling, service tenure, channel of acquisition, regional differences), and uncovers distinct customer segments that inform targeted retention and marketing strategies. These findings offer actionable guidance for companies pursuing convergence strategies, such as the importance of aligning new services with customer needs to truly realize retention benefits.

In the following sections, we first review relevant literature on industry convergence, open innovation, and predictive analytics in telecom/finance (Section 2). We then describe our data and methodology (Section 3), followed by the results of the predictive models and clustering (Section 4). Section 5 discusses the implications of the findings for theory and practice, and Section 6 concludes the paper with recommendations and future research directions.

2 Literature Review

2.1 Industry Convergence and Open Innovation

Industry convergence refers to the blurring of boundaries between formerly distinct industries, often driven by technological innovation and changing consumer demands. Scholars note that digital technologies create common platforms enabling firms from different sectors to offer overlapping services, leading to *"competitive encroachment"* and collaboration across industries (Research and Markets, 2024). In the context of telecom and insurance, convergence is evident in examples such as telecommunication companies offering mobile payment and insurance products, or insurers leveraging telecom data (e.g., telematics for auto insurance). The convergence trend is underpinned by the rise of digital ecosystems: networks of firms that co-evolve their capabilities around a shared platform or customer base. Deloitte and the Institute of International Finance observe that financial services are extending into broad ecosystems with new actors, where value is co-created through complex collaboration and competition (Baumann et. al., 2023). In these ecosystems, roles are fluid, and firms integrate via API and data sharing agreements, going beyond traditional industry silos (Kazmi, 2021).

Open innovation theory provides a lens to understand how and why firms engage in such cross-industry integration. Open innovation is defined by Chesbrough as "the use of purposive inflows and outflows of knowledge to accelerate internal innovation and expand the markets for external use of innovation" (Chesbrough et al., 2006). Rather than relying solely on internal R&D, companies embrace external ideas and paths to the market. In telecom and

financial services, this might involve joint ventures, platform sharing, or enabling third-party developers. By adopting open innovation, incumbent firms can respond to digital disruption more rapidly, for example, banks partnering with fintech startups to offer innovative services, or telecoms co-developing products with insurance companies. A 2019 industry report noted that organizations across sectors, including telecom operators, have widely accepted the open innovation paradigm to harness external expertise and technology. This approach has proven effective in many cases (with oft-cited examples like P&G's connect-and-develop, or LEGO's user co-creation) and is increasingly seen in telecom and finance. Open banking (the regulatory-driven opening of bank APIs) and emerging "open insurance" initiatives are concrete manifestations of this trend, intended to spur service convergence through shared data and collaborative product offerings (Kazmi, 2021). However, convergence via open innovation is not without challenges. Regulatory and security concerns arise when sharing sensitive financial data across industries (Kazmi, 2021). Differences in organizational culture between, say, a telecom and a bank can hinder seamless collaboration. Moreover, delivering an integrated customer experience requires significant IT integration and coordinated innovation. When executed well, convergence can create win-win scenarios, e.g., telecoms can leverage their customer reach to distribute banking services in underserved markets, improving financial inclusion, while banks gain new customers and data sources (Baumann et. al., 2023). The convergence of telecom and insurance can be viewed as part of a broader shift toward open data ecosystems, wherein data flows enable new business models, partnerships, and customer value propositions (Baumann et. al., 2023).

2.2 Digital Transformation and Customer Analytics

Digital transformation in all three industries has elevated the importance of data-driven decision making. In telecommunications, firms have long used analytics for network optimization and customer relationship management. Banking and insurance, too, increasingly relies on big data and AI for risk assessment, personalized marketing, and operational efficiency. A notable area of analytics application is customer churn prediction, identifying customers likely to leave (churn) so that preventative retention actions can be taken. The telecom industry has a rich history of research in churn modeling due to the high competitiveness of the market and the substantial cost of customer acquisition. Early studies applied statistical methods and decision trees, while more recent approaches leverage machine learning ensembles and feature selection techniques to improve predictive accuracy. For example, Idris et al. (2012) used an ensemble (RotBoost) with maximum relevance feature selection to achieve high churn prediction performance. Likewise, Lemmens and Croux (2006) demonstrated that bagging and boosting decision trees can significantly enhance churn classification accuracy. These advances have made churn prediction a well-established practice in telecom management.

Insurance companies use predictive models to identify policyholder lapses or cross-selling opportunities. What is novel in a convergence context is analyzing customer behavior across service types, for instance, how usage of an add-on insurance product might correlate with telecom service churn. Bundling of services is a related concept: telecom providers often bundle services (e.g. phone, internet, TV) to increase customer stickiness. Research indeed shows that multi-service bundles can reduce churn; Prince and Greenstein (2011) found that offering *"triple-play"* telecom bundles significantly lowered churn rates for all included services. By analogy, one might expect that bundling financial or insurance services with core telecom offerings could also improve retention by increasing the overall value proposition and switching costs. At the same time, poorly executed bundling could have the opposite effect if customers do not perceive added value.

Another pertinent area is customer segmentation. Clustering techniques are widely used to segment a customer base into homogeneous groups for targeted marketing. In telecommunications, clustering can reveal profiles such as "high-usage business users" vs. "low-ARPU prepaid users," enabling differentiated retention strategies. In a converged setting, segmentation might separate customers by their adoption of cross-industry services (e.g., those who embrace the full bundle vs. those who only use core services). Prior work has applied clustering in telecom for usage-based segmentation, often using k-means or hierarchical clustering on usage metrics (Ren et al., 2009, Edwine et al., 2022, Sharaf Addin et al., 2022). With the addition of financial product variables, new hybrid segments may emerge (for example, a cluster of tech-savvy users who buy device insurance and use mobile payments heavily, versus a cluster of traditionalists who only use basic telecom services).

Despite notable advances in analytics within individual sectors, a significant gap remains in the literature regarding integrated analytics across the telecom and insurance industries. Most existing studies focus either on two-sector convergence (e.g., telecom-fintech) or remain conceptual in their approach to digital ecosystem strategies. A SCOPUS database search using the keywords "telecom," "insurance," and "prediction" confirms this gap, revealing an absence of empirical research that integrates all three service sectors using predictive analytics, a foundational pillar of modern business intelligence.

Several studies have explored predictive modeling within single industries. For instance, Deepthu and Raju Ramakrishna (2021) investigate various data mining and machine learning techniques to forecast

customer switching behavior between insurance companies. Their models include logistic regression, decision trees, and artificial neural networks. Similarly, Mukhopadhyay et al. (2021) apply machine learning algorithms to reduce churn risk within service industries, emphasizing the strategic importance of early detection. More recently, Wagh et al. (2024) conducted a comprehensive evaluation of classification algorithms, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) for predicting customer churn in the telecom sector. Their findings highlight the superior performance of the Random Forest model (92.31% accuracy and 91.57% precision) and the importance of demographic data, usage behavior, and customer feedback in enhancing prediction outcomes.

Building on these insights, our study addresses the identified research gap by applying multiple predictive analytics techniques to an integrated dataset that spans telecom and insurance services. Specifically, we investigate customer churn and segmentation within a cross-sector context, evaluating the influence of bundled service offerings (e.g., telecom + insurance) on customer behavior and lifetime value. In doing so, we extend previous research and contribute to both academic discourse on digital ecosystem management and the practical design of open-innovation strategies aimed at maximizing customer value through convergence.

3. Methodology and Data

3.1. Data Source and Description

To investigate the convergence of telecom and insurance services, we utilized an anonymized customer dataset from a telecommunications operator that has diversified its portfolio by offering insurance products (and/or related financial services) to its subscribers. This dataset represents a *converged service environment*, capturing information on customers' telecom usage as well as their uptake of optional services like device insurance and travel insurance. The data can be seen as a manifestation of open innovation strategy, the telecom partnered with insurance providers to bundle insurance with its core offerings, and the resulting customer data reflects behaviors across these combined services.

The dataset contains customer-level records with a variety of attributes, totaling 20+ variables. Key variables include:

- Customer demographics: age, and region (city). For instance, AGE (in years) and City (which we encoded via a marketing region code, e.g., BAN_City_Marketing for billing city) (*Lemmens & Croux, 2006; Idris et al., 2012*).
- Telecom service tenure: how long the customer has been with the provider, measured in months on both the mobile subscription level (Months_in_onevip_MSISDN) and account level (Months_in_onevip_BAN).
- Account/service features: number of mobile lines on the account (NUMBER_OF_MOBILE_LINES), account type (ACCOUNT_TYPE, e.g., prepaid vs postpaid), and the primary product or tariff plan (ProductName and associated base fee).
- Financial metrics: the monthly subscription fee, including any discounts (Monthly_fee_incl_dsc_tariff), and the base tariff fee without discounts (Monthly_fee_tariff). The difference indicates promotional discounts applied.
- Insurance and value-added service flags: binary indicators for whether the customer has certain add-on services, such as device insurance (DEVICE_INSURANCE_FLAG), screen insurance, an internet security service (A1_NET_PROTECT_CURRENT_FLAG and a flag if they ever had it, A1_NET_PROTECT_EVER_FLAG), and travel assistance insurance (TRAVEL_ASSISTENCE_CURRENT_FLAG and ..._EVER_FLAG). These variables link to the insurance industry and represent cross-selling within the telecom's open-innovation ecosystem.
- Acquisition channel: how the customer was acquired or purchased services, denoted by CHANNEL (e.g., retail store, online, or D2D – door-to-door sales) (*Neslin et al., 2006*).
- Contract timing: variables like TIME_OF_ACTIVATION (possibly the time of day or a code for when service was activated) and SAME_DAY_CONVERGENCY_FLAG (which might indicate if multiple services were adopted on the same day, a proxy for bundle purchase behavior).

The dependent variables for our analysis are derived as follows:

- Churn: We define a binary churn outcome indicating whether the customer has churned (i.e., discontinued service). Due to data limitations (no direct churn label), churn was operationalized from the service expiration date. Specifically, we created a field Churn such that Churn = 1 if the customer's subscription expiration date precedes the analysis date (meaning the service would have expired or not renewed), and Churn = 0 if the expiration date is in the future or the service is still active. In practice, this was implemented with a formula in Power BI:

Churn = IF([EXPIRATION_DATE] < TODAY(), 1, 0). This proxy aligns with common telecom churn definitions (non-renewal of service by a certain date) and was necessary in the absence of an explicit churn field. We also created a more readable categorical variable Churn_Status (values "Yes" for churned and "No" for active) for use in analysis and visualization.

- Monthly Revenue: As a continuous outcome, we use the net monthly fee paid by the customer (Monthly_fee_incl_dsc_tariff). This is effectively the customer's average monthly revenue to the company after any discounts, and reflects the financial value of that customer. It incorporates the effect of any cross-sold products if their fees are included in the billing, thus serving as a holistic revenue measure.
- Customer Segments: For clustering analysis, there is no single target variable; instead, we use a set of key features to discover natural groupings. The clustering will consider variables that reflect customer profile and behaviors, such as churn status, number of lines, tenure, spending level, and use of insurance services, to form segments.

3.2 Analytical Approach

We adopted a quantitative analytical methodology comprising three complementary techniques: logistic regression, multiple linear regression, and cluster analysis. Each technique addresses a specific aspect of the research questions: churn prediction, revenue estimation, and customer segmentation, respectively. All analyses were conducted in *Microsoft Power BI Desktop*, leveraging its data modeling capabilities and AI-enhanced visuals (Key Influencers and segmentation tools, i.e., Top Segments). Power BI provides a user-friendly environment to perform advanced analytics on business data; in particular, the Key Influencers visual uses machine learning (logistic/linear regression) to identify and rank the impact of variables on a chosen outcome, and a built-in clustering tool (via the "Top Segments" feature) to automatically segment data. Using Power BI allowed us to integrate data preparation, modeling, and visualization in one platform, exemplifying how practitioners can apply predictive modeling in a BI setting.

The overall workflow was as follows:

- Data Preparation: We imported the raw data into Power BI and performed cleaning and transformation. Categorical variables (like Channel or City) were encoded as needed (Power BI typically handles categories internally). We ensured that the Churn field was correctly derived as described, and filtered the data to exclude any obviously irrelevant or erroneous records (for example, customers with missing key data fields). We also created any necessary calculated fields (such as Churn_Status). Descriptive statistics were reviewed to understand distributions (e.g., churn rate was about 25.9% overall, as gleaned from the data, and various service adoption rates for add-ons).
- Logistic Regression (Churn Prediction): We set Churn_Status (Yes/No) as the target and used Power BI's Key Influencers visual to run a logistic regression model. The independent variables provided to the model included all relevant customer attributes: age, tenure, number of lines, account type, base tariff, actual fee, presence or absence of each insurance/add-on service, channel, city/region, etc. The Key Influencers analysis then evaluated which factors had the most significant influence on the probability of churn, outputting the top influencers and an approximate odds ratio or impact magnitude for each. This approach identifies both increasing factors (features whose presence or range raise churn likelihood) and decreasing factors (features associated with retention). We paid special attention to factors reflecting convergence (e.g., the insurance flags, multi-line subscriptions, etc.) to see how they correlate with churn in this cross-industry context.
- Multiple Linear Regression (Revenue Modeling): We next examined drivers of Monthly Revenue (Monthly_fee_incl_dsc_tariff) using a multiple linear regression approach. Again, we used the Key Influencers visual but for a continuous outcome, which in Power BI uses linear regression and decision tree analysis to find key factors that increase or decrease the value. The independent variables considered were similar to the churn model (customer tenure, number of lines, product type, region, insurance flags, etc.), excluding any that were redundant or highly collinear with revenue (for instance, the base tariff Monthly_fee_tariff is naturally a strong determinant of net fee, but we included it to quantify its effect). The model provides insight into how much each factor contributes to changes in monthly revenue. We treat the results as a linear approximation of revenue drivers, recognizing that some relationships might be non-linear or interacting (Power BI's tool can capture simple interactions by segmenting the data as well).
- Cluster Analysis (Customer Segmentation): Finally, we performed an unsupervised clustering to identify homogeneous customer segments. We used Power BI's built-in

clustering on the dataset, instructing it to find an optimal segmentation based on features like churn status, number of lines, tenure, age, monthly fee, and presence of add-on services. The algorithm behind this is akin to k-means clustering (partitioning customers into k clusters by maximizing similarity within clusters and differences between clusters). We experimented with different numbers of clusters and found that three clusters provided a meaningful and interpretable segmentation of the customer base. The "Top Segments" feature in Power BI was then used to profile each cluster, revealing which attributes most distinguish each segment. This helped relate the clusters back to business characteristics (e.g., high churn group vs low churn group, etc.). Each cluster's profile (in terms of average attributes and churn rate) was documented for analysis.

Throughout the modeling, we took care to avoid overfitting and to validate results. Given that Power BI's visuals do not output traditional statistical metrics (like R-squared or p-values) directly to the user, we interpreted the results primarily in terms of practical significance. For logistic regression, we note which factors have the strongest influence on churn probability and cross-check that they align with domain expectations or patterns observed in the data. For linear regression, we consider the directional effects and relative magnitudes to ensure they make sense (for example, higher base tariff should correspond to higher revenue a sanity check that the model is logical). The clustering was validated by examining cluster statistics (size of each cluster, churn rate within cluster, etc.) and ensuring they were distinct. We also reviewed a sample of data points in each cluster to ensure no obvious miscoding.

By using three different techniques on the same dataset, we triangulate our understanding of the converged customer base. The logistic model answers "who is likely to leave and why," the linear model answers "what drives more revenue per customer," and the clustering answers "what distinct groups of customers exist and what characterizes them." The integration of these results provides a comprehensive view that can inform strategic decisions in an open innovation context, such as which segment to target with a new bundled offering or which factor to improve to reduce churn.

4 Results and Analysis

4.1 Logistic Regression: Churn Prediction

The logistic regression analysis for churn yielded clear insights into the factors that increase or decrease the likelihood of customer attrition in the converged telecom-financial service context (Fig. 1). The most influential factor was the channel of acquisition: customers acquired through the door-to-door sales channel ("D2D") showed a dramatically higher propensity to churn. In fact, the model indicates that being acquired via D2D is associated with about 3.10 times greater odds of churning compared to customers acquired through other channels. This suggests that aggressively acquired customers (e.g., through direct sales campaigns) may be less loyal or of lower long-term fit, possibly because they were attracted by a short-term promotion or had less intrinsic interest in the service. It underscores the risk of certain customer acquisition tactics without strong ongoing engagement; these customers tend to disconnect early.

The analysis also found that customers with minimal service engagement are much more likely to churn. Subscribers who have zero mobile lines on their account (for instance, perhaps only using a fixed service or having an account structure with no active line) were highly prone to churn. This intuitively aligns with the idea that the more services a customer has (and the more entrenched they are), the less likely they are to leave. A customer with no mobile lines (or essentially very low usage) is likely to have weak attachment to the provider. This result reinforces the strategy of service bundling to increase stickiness with each additional service (e.g., adding a mobile line, or an insurance product) can deepen the customer's commitment.

Interestingly, the presence of certain insurance add-on services correlated with higher churn likelihood. Customers who had purchased device insurance were more likely to churn than those who hadn't, all else being equal. At first glance, this is counter-intuitive: one might expect that if a customer bought device insurance (an indicator of engagement and an additional service), they would be more loyal. Two possible explanations are: (1) Timing effect, customers may tend to buy device insurance near the end of a device contract or when concerns about device damage rise, which might coincide with the end of their service contract as well. Thus, the insurance purchase could be a last act before churn (for instance, they insure a device but then churn when their contract term is up). (2) Satisfaction issues it's conceivable that some customers who filed a claim or interacted with the insurance service became dissatisfied (perhaps due to claim process hassles or coverage limitations) and that contributed to their decision to leave the telecom provider entirely. While our data cannot pinpoint the reason, the association is a valuable insight: simply selling an add-on service does not guarantee retention, and in some cases might cluster with higher churn, meaning companies should monitor customer satisfaction with those add-ons closely.

Price and tariff factors also emerged as significant. The model highlighted that customers in a certain mid-range of monthly fee were more likely to churn: specifically, those paying a monthly fee (after discount) in the range of about 635 to 677 MKD had

an elevated churn probability. This segment of upper-mid-tier spenders might feel they are not getting sufficient value for the cost, as they are not the lowest budget customers, yet not the absolute top-tier who might receive premium perks. Additionally, the analysis flagged low base tariff levels (≤ 592) as another churn-prone category, these could be low-paying customers who are always hunting for cheaper deals and thus have low loyalty. Taken together, it seems both ends of the spectrum need attention: low-value customers churn perhaps due to price sensitivity, whereas mid-value customers churn possibly due to value perception issues (feeling "for what I pay, I deserve better").

Another demographic insight was that senior customers (age above 67) showed higher churn probability. Older customers might be churning due to different reasons (e.g., less need for service, moving to family plans, or even age-related factors). It's a reminder that different age segments have different churn drivers; retention approaches must be tailored (for example, tech-savvy youth might churn if services are not digitally advanced, whereas seniors might churn due to complexity or life stage changes).

Crucially, the analysis unveiled some regional patterns in churn. Customers residing in particular cities: Gostivar, Kochani, Prilep, and Tetovo were more likely to churn than those in the capital or other regions. For instance, being in one of those medium-sized cities correlated with higher churn risk. This could reflect local competitive dynamics (perhaps a strong competitor presence or network issues in those areas), or economic factors unique to those regions. It signals that the company may need targeted retention efforts or market research in those cities to address the root causes (be it service quality, localized marketing, or distribution issues).

Service tenure had a nuanced effect. Medium tenure ranges (customers around 1.5-2 years with the company, e.g. 17-23 months) as well as very long tenure (e.g. > 47 months) showed higher churn propensity than the baseline. The medium-tenure spike likely represents contract cycle churn many contracts are 12 or 24 months, so churn often occurs at those renewal points. The higher churn among ultra-long tenured customers might indicate that after many years, some customers eventually do leave possibly due to complacency by the provider or attractive offers elsewhere targeting long-time customers of competitors. It's an important reminder that loyalty is not forever without continued value delivery, and even long-time clients should not be neglected.

The absence of certain cross-service engagements was telling. Customers who never subscribed to the Net Protect service (a cybersecurity/value-added service) were more likely to churn. This could mean that those who engage with such supplementary services end up having more reasons to stay (the service itself providing stickiness), or conversely, those who churn might generally be less engaged and thus never bothered with extra services. In either case, it suggests that encouraging adoption of value-added services (security, insurance, etc.) as part of an open innovation bundle can help reduce churn, but only if those services truly add perceived value.

Churn in this converged service scenario is multifactorial, driven by a combination of how the customer was acquired, how many and what types of services they use, how much they pay relative to perceived value, where they live, and who they are (tenure and age). Notably, no single factor dominates churn risk, instead, it's the interplay: for example, a door-to-door acquired customer with one line, who never took an add-on service and lives in a smaller city, is an archetype of a high-risk churner. In contrast, a customer acquired through a stable channel (store or online), with multiple lines and one or two add-on services, living in the capital and on a fitting plan, would be relatively secure. These insights confirm that in an open innovation ecosystem, retaining customers requires a holistic strategy addressing product bundle value, customer experience across channels, and targeted interventions for specific segments prone to churn.

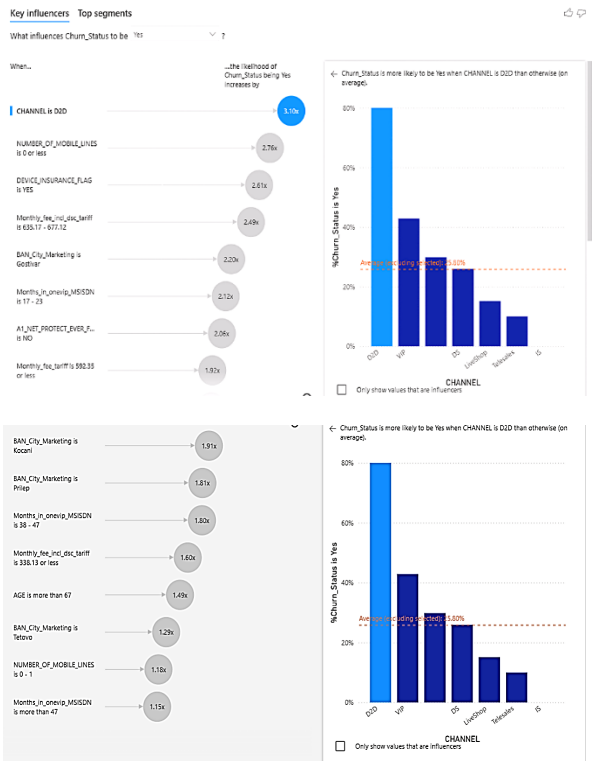


Figure 1. Key influencers of Churn Status to be YES

4.2 Multiple Linear Regression: Revenue Drivers

The multiple linear regression (Key Influencers) analysis on monthly revenue revealed critical insights into the drivers of customer-level revenue contribution, identifying which elements of customer profiles and

service usage behaviors most significantly influence monthly billing outcomes (Fig. 2). As expected, the base service plan price emerged as the strongest predictor. This near-linear relationship confirms that customers on higher-tier plans consistently generate greater revenue, even after discounts or promotional offers are applied. Quantitatively, the model shows an approximate 0.74:1 ratio of net revenue gain for each unit increase in plan tariff, highlighting the enduring value of premium subscribers. In essence, even when adjusted for incentives, higher-paying customers remain substantially more profitable. This underlines the strategic importance of upselling, positioning premium plans as a high-impact lever to increase average revenue per user (ARPU). In addition to plan pricing, the breadth of services subscribed also significantly influenced revenue. Customers with multiple lines (e.g., family or multi-device plans) and those who opt for premium product bundles demonstrated substantially higher total monthly charges. While some multi-line discounts may apply, the aggregate bill still increases with each additional line, contributing positively to overall revenue. Furthermore, premium plan users frequently subscribe to add-on packages such as extended data, entertainment bundles, or advanced features which incrementally raise their monthly payments. Of note is the impact of value-added services such as device protection or insurance, which showed a clear positive correlation with revenue contribution. These findings emphasize the effectiveness of cross-selling strategies in boosting ARPU. However, this must be balanced with the earlier churn analysis, which suggests that aggressive cross-selling could risk increasing customer attrition. From a revenue-centric standpoint, though, both upselling and cross-selling remain powerful tools for revenue enhancement when executed strategically and with customer satisfaction in mind.

Another influential factor identified was customer tenure particularly very long tenure. The model highlighted that when a customer's Months_in_onevip_MSISDN (mobile tenure) exceeded around 89 months (roughly 7.5 years), their average monthly revenue was about 137.7 units higher than that of shorter-tenured customersfile. This is a significant jump, showcasing that long-term customers tend to accumulate more services or stick to higher plans, thereby paying more. There are a few plausible reasons: long-term customers might have gradually added services over time (a second line for a family member, an upgrade to a better plan as their usage grew, etc.), or prices might have risen and they've remained on older, pricier legacy plans. It could also reflect loyalty in terms of opting into various company offerings. In any case, it underlines the point that customer lifetime value increases with retention not just because of duration, but also because retained customers often deepen their relationship and revenue contribution. This aligns with classic CRM theory that retaining customers can yield more revenue over time via cross-selling and upselling.

Beyond these top factors, some other variables have smaller yet notable effects. For instance, account type can influence revenue: postpaid customers usually have higher ARPU than prepaid. Geography might play a role, customers in urban centers or certain regions could opt for more expensive plans (though our model didn't call out city as a top factor for revenue explicitly, possibly because income levels or competition cause differences). Also, time of activation could indirectly reflect if a customer joined during a particular promotion (which might affect their bill structure). However, the model's primary findings revolve around the scope of services and longevity.

The revenue model's implications for strategy are straightforward: to boost revenue, the provider should focus on increasing the adoption of multi-line plans and value-added services, and migrating customers to higher tiers where appropriate. This is a classic upsell approach, but our data provides evidence of its magnitude and importance. Additionally, the tenure effect suggests a payoff in cultivating long-term relationships. It's cheaper to retain than acquire, and as we see, those retained customers often become more valuable. Therefore, investments in customer retention programs not only save the cost of churn but also likely increase ARPU over the years, a double benefit.

The revenue analysis further confirmed that customer tenure is positively associated with higher monthly revenue contributions. Customers with longer relationships tend to migrate to more advanced plans, accumulate additional services, and display generally higher engagement levels. However, this finding must be viewed in the context of earlier churn results, which indicated that very long-tenured customers may exhibit a renewed propensity to churn. This nuance suggests that while tenure correlates with increased revenue, it does not guarantee retention. Consequently, companies should proactively re-engage their long-standing customers through periodic plan refreshes, personalized communications, or loyalty rewards that reinforce the value of staying, preventing stagnation or the lure of competitor offers.

In addition, the bundling of multiple services such as telecom plans with insurance products emerged as a significant driver to revenue growth. Customers subscribing to device protection, for instance, contributed meaningfully more to average revenue per user. However, such revenue gains are not without risk. If bundling is perceived as non-transparent, overly aggressive, or misaligned with customer needs, it may result in dissatisfaction and elevated churn. For example, if device insurance is sold without clearly communicating its value, limitations, or benefits, customers may see it as an unwelcome charge rather than a value-added service. Similarly, mandatory bundles or upsell pressure may erode trust over time. Therefore, while cross-selling remains a powerful revenue lever, it must be executed with a customer-

centric approach. Companies should evaluate not just the short-term revenue boost from bundled offers, but also their impact on long-term customer satisfaction and retention. Enhancing the perceived value of bundled services such as offering the first year of device insurance for free, simplifying claims processes, or clearly explaining benefits can reduce the likelihood of negative customer reactions. Ideally, cross-selling should feel like a solution to the customer's needs, not a revenue tactic. Balancing profitability with perceived fairness and value is key to maintaining both high ARPU and low churn.

The linear regression confirms that "revenue follows value": the more value (services, premium features) a customer receives (and is willing to pay for), the higher their contribution. It validates the convergence strategy from a financial perspective, customers who embraced convergent offerings (like insurance or multi-service bundles) did have higher ARPUs. The task for management is to encourage more customers to find value in multiple services (thereby raising revenue) while also using the churn insights to ensure those customers remain happy and loyal.

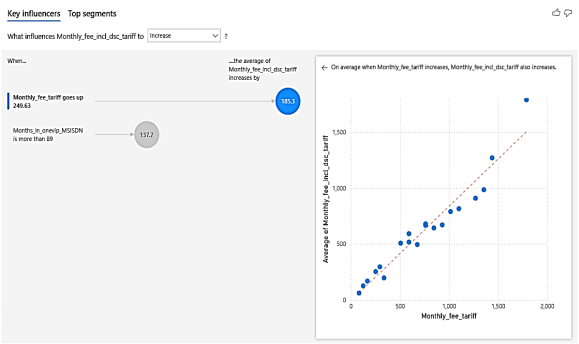


Figure 2. Key influencers of Revenue to Increase

4.2 Cluster Analysis: Customer Segmentation

The clustering analysis revealed three distinct customer segments in our data, each with unique profiles relating to churn behavior, service usage, and value indicators. These segments provide a more refined understanding beyond individual variables, as they capture combinations of characteristics that tend to occur together in the real world. We label them Segment 1, Segment 2, and Segment 3 for reference, and describe each below.

4.2.1 Segment 1: High-Value, High-Churn Customers

This segment is characterized by customers who have above-average spending and multiple services, yet exhibit a high churn rate (Fig. 3). Their typical monthly fee (including discounts) falls in an upper-mid tier (around 635-677 in billing units) these are not the absolute lowest spenders, indicating they likely subscribe to decent-tier plans (and possibly some add-ons). They also tend to maintain more than one mobile line on their account, reflecting a degree of service complexity and family or multi-device bundles. Their tenure with the company is usually moderate (over ~17 months), meaning they have experience with the provider (not brand new customers). Despite these seemingly positive attributes (multi-line, moderate tenure, mid-high spend), Segment 1 customers have the highest churn rate among all segments. In our data, a significant portion of churned customers fell into this group. Geographically, this segment is disproportionately present in the capital city Skopje (around 33% of this segment) and in some smaller cities like Kochani (~15%) and Bitola (~13%). These three cities together accounted for a large share of churn within Segment 1, suggesting possible regional dissatisfaction or competition hotspots.

The paradox of Segment 1 is that these customers are valuable (higher revenue contribution, multi-service) yet they are leaving at high rates. This implies a mismatch between expectations and delivered value. They pay not the lowest price, likely expecting premium service quality or more benefits, and if those expectations are unmet, they may be more resentful and quicker to churn than a low-budget customer. The presence of multiple services indicates they gave the provider a broader mandate (e.g., mobile + perhaps fixed line or insurance), so churn here is especially concerning it's a loss of a high-value relationship. *Strategic implications:* Segment 1 should be a top priority for retention efforts. These customers might respond well to loyalty programs, personalized offers, or "VIP" treatment to make them feel they are getting value commensurate with what they pay. For example, offering exclusive discounts, faster support, or bundling additional perks could entice them to stay. Proactive engagement is key: surveys to identify pain points, targeted service improvements in the flagged regions (Skopje, Kochani, Bitola) to address any local issues, and possibly repackaging of services to ensure they perceive a clear advantage in staying bundled rather than seeking alternatives.

4.2.2 Segment 2: Low-Engagement, At-Risk Newcomers

Segment 2 is a relatively small segment (only about 8% of the entire customer base), but it has an exceedingly high churn rate in our data, 62.6% of Segment 2 customers were churned, which is over 36 percentage points higher than the overall average churn of ~25.9% (Fig. 4). The defining feature of this segment is that customers have zero mobile lines associated with their accounts. In other words, they either subscribed to a service that doesn't include a traditional mobile line (perhaps only an insurance product or a data-only service, or they registered but never activated a line) or their accounts are essentially inactive from the start.

They also show negligible adoption of any add-ons or bundles essentially no deep engagement with any core offerings. We interpret Segment 2 as likely including very new customers who never fully onboarded, one-time sign-ups that didn't stick, or administrative accounts with no real usage.

The behavior of Segment 2 (quick and high churn) suggests issues in the onboarding process or the initial value proposition. These customers leave almost as soon as they arrive. They might represent those who *subscribe on a trial or promotion and then cancel*, or whose service activation failed or never truly started (hence zero lines active). They could also include accounts that were opened for a secondary product but never resulted in continued telecom usage. From an open innovation perspective, if these were customers who perhaps only wanted an insurance product through the telco channel and then canceled, it indicates a challenge in cross-selling they didn't become telecom loyalists. The churn here might not be very costly revenue-wise (since they were low spenders), but it distorts retention metrics and points to a potential waste in acquisition costs. *Strategic implications:* Even though Segment 2 customers are low-value individually, their extreme churn rate deserves attention because it might signal structural problems. The company should investigate why these sign-ups didn't convert into active customers. Are sales teams bringing in "junk" contracts just to meet targets? Is the onboarding experience confusing or is the product not as advertised? By addressing these issues, e.g., improving the onboarding communication, setting proper expectations, or offering a small incentive for new customers to actually start using the service, the firm could reduce churn in this segment. Moreover, we might decide to not focus heavy retention resources on them (since they have low revenue) but rather fix upstream processes (acquisition and onboarding) to prevent the phenomenon. In the spirit of open innovation, learning from these quick failures (possibly via feedback or exit surveys) could inform better partnership offerings or more suitable customer targeting.

4.2.3 Segment 3: Established Multi-Service Users with Selective Engagement

Segment 3 customers have a profile somewhat in between Segment 1 and the average, but with distinct quirks (Fig. 5). About 46.3% of Segment 3 customers were churned, which is ~20 points above average. So churn is high, though not as extreme as Segment 2. Characteristics of Segment 3 include: medium-to-long tenure (most have been with the provider > 17 months), more often more than one mobile line (so they are not single-line users), and they fall into a bimodal spending pattern – their monthly fees are either very low (≤ 338) or quite high (≥ 804), with fewer in the middle range. Additionally, a common attribute is the absence of the A1 Net Protect service virtually none of them have ever had this value-added security service. They also typically are on mid- to high-tier base tariffs (somewhere in ~592-1100 range), and geographically, this segment is overrepresented in Kochani (~9.7%), Prilep (~10.7%), and Bitola (~7.3%), which interestingly overlaps partially with the regions from Segment 1's concerns (except Skopje is not highlighted here).

Interpreting Segment 3: these customers are reasonably well-established with the company (not new, many have multiple lines), and many pay for higher-tier services, yet a large chunk leaves. Their distinguishing feature is not using the optional security service and having a polarized spending (some very budget-conscious, some premium). One way to see this is that Segment 3 could consist of (a) price-sensitive customers on the low end who maybe have multiple connections (e.g., a family plan with minimal usage) but churn because even slight price issues or lack of value-added services make them switch, and (b) premium customers who use multiple lines and pay a lot but possibly expect comprehensive service (including things like security) the fact they *haven't* taken the security service might indicate they didn't see value in it, or it wasn't offered effectively, and their churn might be due to unmet service expectations or finding better packages elsewhere. The regional aspect suggests that in those specific cities, even substantial customers are prone to switch there could be strong competitor promotions in those areas or local service quality issues. *Strategic implications:* For Segment 3, the lack of engagement with the security add-on (and possibly other add-ons) signals an opportunity. The company could market value-added services more effectively to these customers, showing how services like A1 Net Protect provide benefit this could both generate new revenue and increase stickiness if customers adopt them. Additionally, since these customers are split between very low and very high spenders, a two-pronged strategy is needed: ensure the low-end users feel they are getting a good deal (maybe small perks or reassurance of value), and ensure the high-end users feel they are getting premium treatment. The mid-to-high base tariff range indicates they invest in good plans; bundling an included security service or other perks into those plans might prevent churn by increasing perceived value. Regionally, special retention campaigns in Kochani, Prilep, and Bitola for at-risk high-end customers (maybe dedicated account managers or service clinics) could help. Segment 3 highlights that even a seemingly stable group (long-tenured, multi-line) can hide vulnerabilities in this case, a lack of cross-service engagement and localized issues lead to churn.

Across all three segments, certain themes emerge. First, having multiple services (lines or add-ons) generally correlates with being in a higher-value segment (Segments 1 and 3), but it does not guarantee low churn. The quality of those services and alignment with needs are crucial. Second, perceived value and

expectation management are central, both Segment 1 and Segment 3 churn in large part due to perceived gaps in value (Segment 1 felt they paid a lot and possibly didn't get enough; part of Segment 3 might feel similarly at the high end, while others at the low end are very price-driven). Third, execution in acquisition and onboarding stands out: Segment 2's extreme churn is a clear call to improve those early stages, whereas Segment 1 and 3's issues often manifest later in the customer life cycle (post-promotion or during contract renewals).

It's instructive to note that churn is not monolithic across the customer base; it manifests differently in different clusters. Segmenting the base allowed us to see that one-size-fits-all retention solutions won't work, each segment has distinct churn triggers. Our converged industry context adds layers to this: for instance, a typical telecom segmentation might not have considered an insurance or security service variable, but here it proved informative. Also, by identifying high-risk segments that coincide with high value, the company can prioritize retention resources more effectively (e.g., focus on Segment 1 first).

The cluster analysis provides a "market map" for the telecom's converged strategy. Segment 1 is a *high-risk, high-reward* group needing immediate retention focus. Segment 2 is a *red-flag group* signaling issues in customer acquisition/onboarding perhaps needing a process fix more than a marketing fix. Segment 3 is a *mixed-bag group* that requires boosting engagement (through cross-sell of value-adds) and careful attention to both ends of its customer spectrum. These insights validate that convergence (telecom + insurance) strategies must be nuanced and segment-specific. The mere presence of multiple services isn't enough companies must deliver consistent value and address the unique needs of different customer segments in the converged ecosystem.

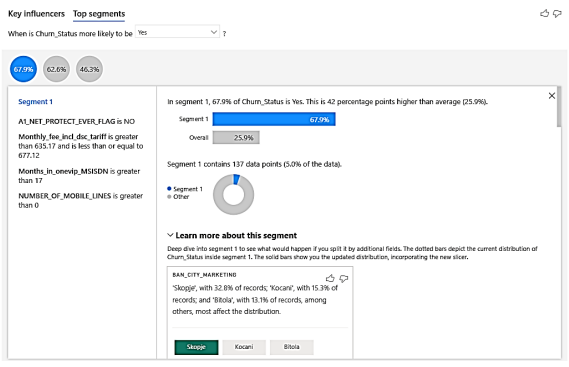


Figure 3. Cluster Analysis: Segment 1

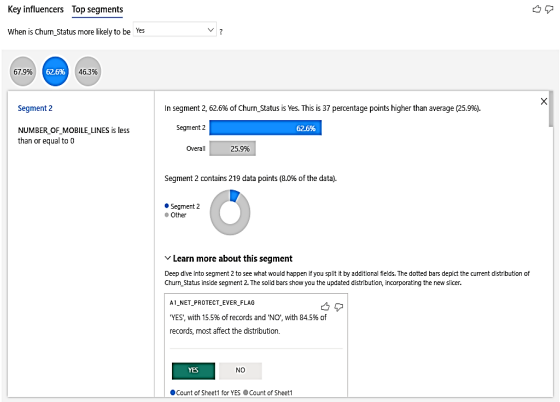


Figure 4. Cluster Analysis: Segment 2

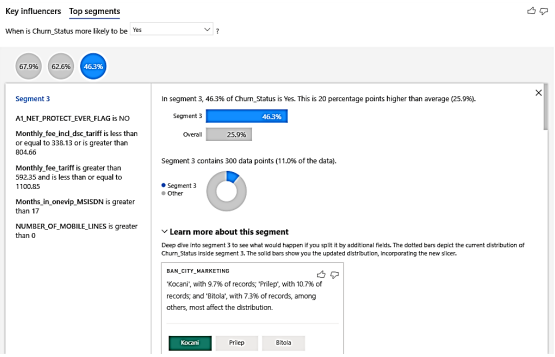


Figure 5. Cluster Analysis: Segment 3

5 Implications

5.1 Theoretical Implications

From a theoretical standpoint, this research enriches the discourse on business ecosystem convergence and open innovation by providing empirical evidence of how cross-industry integration affects customer behavior. Prior conceptual work has posited that converging industries (telecom, finance, etc.) can create synergistic value and new business models (Baumann et. al., 2023). Our findings give granular support to these claims: for instance, we show that a telecom offering insurance (open innovation in action) can increase customer revenue (through cross-selling) and potentially retention (if value is delivered), but it also introduces new complexity in managing customer satisfaction. This underscores the idea that open innovation is a double-edged sword, it opens avenues for value co-creation (more revenue, diversified offerings) but also demands higher coordination and customer-centricity to ensure the innovation truly benefits the user.

We also contribute to the customer analytics literature in telecom and services marketing by examining traditional metrics (churn, ARPU) in an unconventional context (multi-industry service bundles). The results suggest that classic drivers of

churn (like price, tenure, service quality perceptions) still apply, but new drivers emerge in a converged setting (like the effect of add-on services or acquisition channel nuances). This indicates that analytical models need to evolve as industries converge; researchers and analysts should incorporate variables that capture cross-industry engagements when predicting outcomes. Our study provides a framework for doing so and could be a reference for similar analyses in other convergent contexts (e.g., tech companies offering banking services, etc.).

Furthermore, our segmentation analysis offers a practical application of segmenting in an ecosystem context. Traditionally, segmentation might be siloed (telco segments vs. bank segments). We demonstrate a holistic segmentation that spans service types, revealing, for example, a segment of customers who engage with multiple services yet churn due to value gaps. This introduces the concept of a "convergence segment" groups of customers defined by their multi-service usage patterns and corresponding behaviors. Future research can build on this by exploring if such segments exist in other markets and how stable they are over time. It also raises interesting questions for theory: Do converged service consumers exhibit fundamentally different loyalty drivers than single-service consumers? Our work suggests they might, as their expectations are broader.

Another theoretical implication relates to innovation strategy: our study shows tangible outcomes of an open innovation strategy (the telecom's partnership with insurance providers) at the micro level. It provides a case for the often-theorized link between open innovation and performance. Specifically, open innovation in this case led to new products (insurance) that increased ARPU (positive outcome), but also had nuanced effects on churn (some negative signals). This nuanced outcome aligns with the complexity suggested in open innovation literature, where not all collaborations yield straightforward benefits. It suggests that the success of open innovation strategies should be evaluated on multiple metrics, revenue, retention, customer satisfaction, not just one.

Our use of Power BI as an analytical tool in academic-style research is a methodological implication but also a theoretical one for the information systems field: it exemplifies the increasing capability of self-service BI tools to perform advanced analytics, blurring the line between business analyst and data scientist roles. It raises the notion that analytical innovation (in terms of tools and democratization of AI) is part and parcel of digital transformation in organizations. Researchers studying technology adoption could glean from this that advanced analytics are no longer confined to R&D labs; they can be executed in mainstream business software, potentially accelerating data-driven decision-making as part of digital transformation strategies.

5.2 Practical Implications

For practitioners, particularly managers in telecom, insurance, or any combination thereof this study offers several actionable insights grounded in customer data from North Macedonia. While many findings reflect localized realities, they provide useful guidance that may be adapted to other contexts with proper cultural and market calibration.

Segment-Specific Retention Strategies: One size does not fit all in convergent service retention. We identified concrete customer segments (e.g., Segment 1: high spend, high churn; Segment 2: low engagement, quick churn; Segment 3: established but at-risk). Managers should adopt tailored retention approaches for each segment. For high-value customers who are churning (Segment 1), invest in premium loyalty programs, solicit feedback, and address service quality or feature gaps. These customers are worth the extra effort due to their revenue contribution. For new customers who drop off immediately (Segment 2), improve onboarding: simplify setup, follow up early, and ensure expectations match deliverables, perhaps offer a "welcome" gift or personal call to increase engagement in the first month. For the mixed segment (Segment 3), a combination of upselling value-added services and targeted communication in regions with issues is needed. For instance, if Bitola has network complaints, they address those issues to improve satisfaction, while also educating customers on the benefits of services like A1 Net Protect to increase their service stickiness.

Optimize Acquisition Channels: The analysis, based on customer behavior in North Macedonia, clearly showed that door-to-door (D2D) acquisition channels experienced higher churn rates. While this insight is highly relevant to the local market, its applicability in other regions may vary depending on cultural norms, consumer expectations, and regulatory factors. This finding suggests that marketing and sales teams should critically reassess aggressive acquisition methods. While D2D may drive customer volume, its effectiveness in retaining loyal customers is limited. Companies might consider improving onboarding for D2D-acquired customers, perhaps by pairing the sale with a strong follow-up program or pivoting toward acquisition channels with higher customer retention rates such as online platforms, retail stores, or referral-based programs. Additionally, if companies engage third-party dealers or agents (in line with open innovation strategies), incentive schemes should reward not only activations but also post-sale customer retention (e.g., 3-month or 6-month loyalty metrics).

Bundle Design and Value Communication: Bundling telecom and insurance services is a core element of convergence. Our results highlight both the potential and the risk of such bundling strategies. Customers who subscribed to multiple services, including device protection, tended to spend more and exhibited signs of increased engagement. However, a

subset of these customers still churned, likely due to perceived low value or dissatisfaction with the bundled services. The implication is that companies must design bundles with real, clearly communicated benefits. Simply adding services for an extra charge is insufficient if customers don't understand or appreciate the value. For example, one possible approach might be to explore embedding device insurance into a premium plan as a built-in benefit, rather than offering it as a separate add-on. This *could* elevate perceived value and potentially reduce dissatisfaction, though further validation would be needed. Moreover, if analytics indicate a correlation between certain add-ons and churn, companies should revisit those offerings, for example, by improving the insurance claims process or selecting more trusted underwriters.

Leverage Data for Personalized Upselling: The revenue model revealed opportunities to increase ARPU through upselling strategies, including multi-line plans, premium tariffs, and value-added services. These can be operationalized via internal analytics tools such as Power BI or CRM systems to identify promising customer segments. For instance, a mid-tier customer with one line and no add-ons could be targeted with a family plan upgrade or device protection offer. However, such strategies must reinforce perceived value to avoid triggering churn. If a customer is offered a security app, ensure they understand, activate, and benefit from it. This reinforces satisfaction while increasing spend.

Geographical Focus: The study identified several high-churn regions in North Macedonia, such as Gostivar, Kochani, Prilep, and Tetovo. Practical next steps include localized customer surveys, network quality audits, and regional engagement campaigns. For instance, telecom providers might host local customer appreciation events, offer limited-time regional promotions, or invest in infrastructure upgrades in areas with known service complaints. Ensuring regional consistency in service quality is vital for the success of convergence strategies, especially when expanding into financial services through telecom partnerships. A local churn spike can undermine broader ecosystem innovations if left unaddressed.

Integration of Customer Data and Experience: A key enabler of this study was the integration of telecom and insurance usage data, which is often siloed in real-world partnerships. For convergence to succeed, firms must invest in shared, secure data platforms that allow for unified customer views. Such integration enables better targeting, churn prediction, and offer personalization. For example, if a telecom operator and bank share mobile usage and payment data, they can jointly identify patterns that trigger churn or upsell opportunities. Open data collaboration, governed by privacy standards, should be a priority in converged business models.

Long-Term Customer Management: While tenure was positively associated with revenue, a mild uptick in churn among very long-tenured customers suggests that these customers should not be overlooked. Beyond first-year engagement, companies *may consider testing* "late-tenure" strategies, such as sending personalized thank-you messages, trialing surprise upgrades, or exploring transitions to newer bundled service plans. The potential impact of these approaches should be assessed through targeted pilot programs or experiments. This can refresh the value proposition and reduce complacency-driven churn.

The overarching practical message is that data-driven open innovation strategies offer significant promise but require thoughtful execution. Convergence is not just about adding more services it is about adding the right ones and delivering them in a way that aligns with customer expectations and cultural norms. Managers at the telecom-insurance nexus must foster a culture of experimentation and feedback, pilot offers regionally, closely tracking key performance indicators like churn, and continuously refining their strategies.

Finally, it is important to note that while our insights stem from the North Macedonian market, the practical implications are not intended as one-size-fits-all prescriptions. Factors such as acquisition effectiveness (e.g., D2D), bundling perceptions, or service loyalty may differ significantly across countries and cultures. Therefore, we encourage practitioners and researchers to adapt these strategies in line with local consumer behavior, regulatory constraints, and partner ecosystem maturity.

6 Conclusion

This paper presented an in-depth analysis of industry convergence between telecommunications and insurance through the twin lenses of open innovation strategy and digital transformation, using predictive analytics as a unifying tool. By examining a unique telecom dataset enriched with convergent service information (insurance add-ons), we gained insights into how cross-industry offerings impact classic telecom performance indicators like customer churn and revenue, as well as how they shape the customer segments within the base.

Our findings affirm that convergence strategies when implemented via open innovation can unlock new revenue streams and customer value, but also introduce new challenges in customer retention and experience management. On one hand, customers who embraced multiple services (telecom + insurance) demonstrated higher average revenues, validating the synergy and cross-selling potential of convergence. On the other hand, the churn analysis revealed that simply adding services does not guarantee loyalty; in fact, poorly delivered or misaligned services can contribute to churn. The most at-risk customers included those acquired through aggressive channels and those whose service bundle value did not meet their expectations.

Meanwhile, long-tenured and multi-service customers proved to be the most valuable in revenue terms, underscoring the importance of long-term relationship management in a converged ecosystem.

From a strategic perspective, the research highlights the need for a holistic, data-driven approach to managing convergent business models. Telecom operators venturing into finance or insurance must extend their analytical toolkit and business processes to monitor not just their core metrics, but also the metrics emerging from the interplay of industries (e.g., adoption rates of financial products, their effect on telecom churn, etc.). Open innovation should be complemented by what we might call "open analytics" sharing data and insights across partner organizations to jointly optimize outcomes. The results encourage companies to invest in integrated customer data platforms and AI tools (like the use of Power BI demonstrated) to continually learn from customer behavior and adapt their offerings.

Limitations and future research: While our study provides valuable insights, it has limitations that open avenues for further research. First, the data is from a single company operating in a specific market context; generalizing to other markets or different scales should be done with caution. Future studies could compare multiple cases of convergence to see if the patterns we observed hold broadly, for example, do all telco-finance bundles see a D2D channel issue, or was that specific to this case? Second, our analysis was largely observational and correlational. We identified associations (like device insurance linked to higher churn) but cannot conclusively prove causation. Experimental or longitudinal studies (e.g., A/B testing an intervention where some customers get a bundle and others don't) could strengthen causal inference about the impact of convergence offerings on churn and revenue. Third, we focused on a subset of possible analytics; other techniques like decision trees, survival analysis for churn timing, or advanced machine learning (random forests, gradient boosting) could be applied to possibly improve predictive power. Incorporating social network analysis (using call/SMS graphs) or text mining of customer feedback could further enrich churn models in future research, especially in a converged context where reasons for churn might be multifaceted. Additionally, future research could explore the customer perspective qualitatively: our data tells the "what" but not the "why" from the customer's mouth. Interviews or surveys could reveal, for instance, why those high-paying customers churn, is it network issues, better competitor offers, or dissatisfaction with the insurance service? Such insights could feed back into refining open innovation strategies (e.g., maybe the insurance product needs redesign or the partnership governance needs adjustment).

References

Baumann, N., Renier, J., Tang, M., French, C., & Sanchez Soriano, G. (2023). *The ecosystem imperative: Digital transformation of financial services and moving from Open Banking to Open Data*. Institute of International Finance (IIF) and Deloitte.

Chesbrough, H. W. (2006). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston, MA: Harvard Business School Press.

Deepthu, D. & Raju Ramakrishna (2021). Machine learning models for customer churn prediction in insurance sector. 12th International Conference on Advances in Computing, Control, and Telecommunication Technologies, pp. 69-73.

Edwine, N., Wang, W., Song, W., & Ssebuggwawo, D. (2022). *Detecting the risk of customer churn in telecom sector: A comparative study*. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2022/8534739

Idris, A., Khan, A., & Lee, Y. S. (2012). Intelligent churn prediction in telecom: Employing mRMR feature selection and RotBoost-based ensemble classification. *Applied Intelligence, 39*(3), 659–672. https://doi.org/10.1007/s10489-013-0440-x

Kazmi, S. (2021). Refine convergence of banking and telecom. *Pakistan & Gulf Economist*. Retrieved from https://www.pakistangulfeconomist.com/2021/07/26/refine-convergence-of-banking-and-telecom/#google_vignette

Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research, 43*(2), 276–286. https://doi.org/10.1509/jmkr.43.2.276

Mukhopadhyay, D., Malusare, A., Nandanwar, A., & Sakshi, S. (2021). An approach to mitigate the risk of customer churn using machine learning algorithms. In A. Joshi, M. Khosravy, & N. Gupta (Eds.), Machine Learning for Predictive Analysis (pp. 133–142). Springer. https://doi.org/10.1007/978-981-15-7106-0_13

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research, 43*(2), 204–211. DOI: 10.1509/jmkr.43.2.204

Prince, J., & Greenstein, S. (2011). Does service bundling reduce churn? [Working Paper]. Bloomington, IN: Kelley School of Business, Indiana University.

Ren, H., Zheng, Y., & Wu, Y. (2009). *Clustering analysis of telecommunication customers*. *The Journal of China Universities of Posts and Telecommunications*, 16(2), 114–116, 128. https://doi.org/10.1016/S1005-8885(08)60214-9

Research and Markets. (2024). *Telco, Fintech, and Insurtech Portfolio Evolution and Positioning Strategies* (Report No. 6029446). Globe Newswire Press Release: The convergence of telecommunications, fintech, and insurtech as key driver for digital services portfolio expansion.

Sharaf Addin, E. H., Admodisastro, N., Mohd Ashri, S. N. S., Kamaruddin, A., & Chong, Y. C. (2022). *Customer mobile behavioral segmentation and analysis in telecom using machine learning*. *Applied Artificial Intelligence, 36*(1), 1–20. https://doi.org/10.1080/08839514.2021.2009223

Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. *Results in Control and Optimization, 14*, Article 100342. https://doi.org/10.1016/j.rico.2023.100342

Tactile Collaboration and Backcasting to Design Meaningful Futures and Strategies

Aldo Valencia
Northumbria University
Newcastle Business School
Falconar Street, Newcastle Upon Tyne, NE83QX, United Kingdom
aldo.valencia@northumbria.ac.uk

Violeta Cvetkoska, Bojan Kitanovikj
Ss. Cyril and Methodius University in Skopje
Faculty of Economics - Skopje
bul. Goce Delcev 9V, 1000 Skopje, North Macedonia
{violeta.cvetkoska, bojan.kitanovikj}@eccf.ukim.edu.mk

Alison Pearce
Northumbria University
Newcastle Business School
Falconar Street, Newcastle Upon Tyne, NE83QX, United Kingdom
alison.pearce@northumbria.ac.uk

Abstract. *This paper introduces a novel framework integrating backcasting with LEGO® Serious Play to foster radical innovation of meanings within organisations, specifically in higher education. Backcasting, by starting with a desired future state, provides a powerful lens for strategic planning. Combining it with the tangible and imaginative nature of LEGO® Serious Play enhances cognition and ownership, enabling participants to map their values onto the strategic vision. The framework's effectiveness was validated through three participatory workshops. Findings demonstrate the approach can reframe assumptions and stimulate innovative strategic pathways, presenting a human-centred focus to enhance the robustness, adaptability, and future-readiness of organisational strategies.*

Keywords. Design thinking, backcasting, strategy, future of higher education.

1 Introduction

Today's world is characterised by technological disruptions, geopolitical shifts, and evolving societal values, the need for adaptable strategic thinking has never been greater for organisations (Vuorinen et al., 2017). As Greenberg et al., (2024) suggests, we stand at a crossroads between economic stagnation and productivity-driven abundance, making it crucial for organisations to proactively shape their future rather than merely reacting to external forces. One of the key challenges highlighted by van Vliet & Kok (2013) is the need for strategies that can withstand uncertainty and embrace a multitude of potential futures.

Traditional strategic planning models, often rooted in predicting a single, definitive future, prove inadequate in this context (Millar et al., 2018). Instead, organisations need to adopt a more agile and dynamic approach to strategy, one that acknowledges the inherent unpredictability of the future and emphasizes continuous learning and adaptation (Garavan et al., 2024). This shift necessitates a move away from static, long-term plans towards more flexible frameworks that allow for adjustments as the environment evolves. Vuorinen et al., (2017) suggests that organisations should maintain a portfolio of strategic options, each designed to address different potential future scenarios. This approach, often referred to as scenario planning, encourages organisations to envision a range of plausible futures and develop corresponding strategies, ensuring they are prepared to adapt to a variety of potential outcomes. Furthermore, Lee (2018) emphasizes the importance of integrating foresight into strategic thinking, enabling organisations to anticipate potential disruptions and opportunities, and proactively shape their future rather than simply reacting to it.

This is where backcasting becomes particularly relevant. Backcasting, as described in Bibri, (2018) and van Vliet & Kok, (2013), is a planning methodology that starts with defining a desired future state and then works backward to identify the steps needed to achieve that vision. By grounding strategic thinking in a clearly articulated desired future, organisations can make more informed decisions today, ensuring their actions align with their long-term aspirations. Strategy can learn from backcasting the importance of intentionality and a long-term perspective, moving beyond simply reacting to current circumstances and instead actively shaping a more desirable future.

Navigating uncertainty and ambiguity in strategic decision-making is not ideal. When faced with a lack of clarity about the future, organisations need access to rich information and diverse perspectives to make informed choices (Thorén & Vendel, 2018). Workshops provide a valuable platform for gathering such information. By bringing together individuals with different backgrounds, experiences, and

viewpoints, organisations can foster a more comprehensive understanding of the challenges and opportunities that lie ahead. Through facilitated discussions, brainstorming sessions, and collaborative activities, workshops can help uncover blind spots, challenge assumptions, and generate innovative solutions that might not emerge from individual contemplation.

LEGO® Serious Play® (LSP), as discussed in Hadida (2013), offers a particularly engaging and effective approach to facilitating such workshops. By using LEGO® bricks as a shared language, participants can express their ideas and perspectives in a tangible, three-dimensional form. This hands-on, playful approach helps to break down barriers, encourage creative thinking, and foster a more inclusive and collaborative environment where all voices can be heard. Jerome et al. (2024) emphasizes the importance of creating a level playing field in collaborative settings, ensuring that all participants feel comfortable sharing their insights and perspectives. LSP achieves this by shifting the focus away from verbal communication, which can sometimes be dominated by a few vocal individuals, towards a more visual and tactile form of expression.

LSP excels in fostering a deeper understanding of complex topics, particularly in the context of envisioning future scenarios. Magnanini et al., (2021) Fischer & Marquardt (2022) and Manogue et al. (2014) highlight the power of tangible metaphors in facilitating this understanding. When participants construct physical models representing abstract concepts or future scenarios using LEGO® bricks, they engage in a process of externalizing their thoughts and assumptions, making them visible and shareable. This tangible representation allows for a more intuitive and multi-sensory engagement with complex ideas, fostering a deeper level of understanding and shared meaning among participants. The act of building and manipulating physical models helps to break down complex concepts into smaller, more manageable chunks, making them easier to grasp and discuss.

While LSP provides a dynamic medium for collaborative exploration, the true power of this approach lies in its ability to facilitate rich information exchange and collective envisioning. By grounding strategic conversations in a tangible, shared understanding of aspirations, we move beyond the limitations of traditional strategy models, often confined to incremental advancements and reactive responses to current realities. This shift towards a future-oriented, value-driven approach aligns with the principles of meaningful innovation (Verganti, 2016). By starting with a shared vision of a desirable future, we unlock the potential for transformative change, driven not by short-term gains but by enduring human values.

Radical innovation of meanings, as described in (Verganti & Öberg, 2013), goes beyond incremental improvements to existing products or services. It seeks to redefine the very meaning of those offerings, creating new value propositions that resonate with deep-seated human needs and aspirations. This often involves disrupting established industries, shifting paradigms, and creating new markets driven by novel understandings of what is **desirable and meaningful**. Radical Innovation of Meanings is also referred as the fourth type of design thinking, characterized by critique and reflection (Dell'Era et al., 2020).

Similarly, preferable futures are not mere extrapolations of current trends. They represent a conscious departure from the status quo, envisioning a world where societal challenges are addressed, and human well-being is prioritized. This requires a willingness to question existing systems, structures, and beliefs, and to imagine alternative realities that embody desired values and principles. Therefore, both radical innovation of meanings and preferable futures demand a fundamental shift in perspective. By challenging existing paradigms and exploring new design spaces, we can create innovations that are not only technologically advanced but also socially responsible and deeply meaningful.

In contrast to incremental innovation that primarily focuses on optimizing existing products and processes, Norman & Verganti (2013) argue for a more radical approach that challenges fundamental assumptions and explores fundamentally new design spaces.

Dell'Era et al. (2020) and Verganti et al. (2020) highlight the transformative potential of design thinking in the age of artificial intelligence (AI). By placing human needs and values at the centre of the innovation process, organisations can leverage the power of AI to create solutions that are not only technologically sophisticated but also socially responsible and human-centred. Building upon this foundation, Norman (2023) proposed the humanity-centric approach, emphasising the importance of embedding sustainability in its widest definition—ensuring that resources are not compromised and that environmental, social, and economic factors are all considered. This approach necessitates a fundamental shift in perspective, moving beyond a purely market-driven model of innovation towards one grounded in a deep understanding of human values and aspirations. Such a paradigm aligns with the Sustainable Development Goals (SDGs), which advocate for practices that promote environmental stewardship, social equity, and economic viability. Furthermore, the humanity-centric approach incorporates ethical considerations, inclusive design, and long-term impact assessment. By doing so, it ensures that technological advancements contribute positively to society and the planet, fostering innovations that are resilient, equitable, and sustainable for future generations.

2 The Case

2.1 Future of Higher Education Institutions

This interconnectedness of future thinking, backcasting and radical innovation of meanings, provides a powerful framework for navigating the complexities of the 21st century. By embracing a future-oriented, value-driven approach, organisations can move beyond incrementalism and embrace transformative change. This involves engaging in critical reflection on existing paradigms, envisioning desirable futures grounded in human values, and leveraging design thinking principles to translate those visions into tangible solutions.

Addressing the future of higher education (HE) demands more than just responding to short-term issues. As noted by Alexander (2020), proactive planning and strategic foresight are crucial. Trend analysis helps to recognise and comprehend the driving forces behind future developments, such as technological innovations, demographic changes, and economic trends. Building on this, scenario planning creates credible narratives about potential future outcomes.

In the HE context, the imperative for future-proof strategies, grounded in frameworks that emphasize backcasting and collaborative innovation, which allow higher education institutions (HEIs) to anticipate and proactively address these challenges, is particularly pronounced. By fostering a culture of continuous adaptation and leveraging interactive, technology-driven teaching methods, HEIs can enhance their resilience and sustain their crucial role in societal development.

2.2 Contextual Background

Historically, HEIs have been regarded as fundamental drivers of social progress. However, in contemporary times, the sector has become increasingly susceptible to global shifts in technology, pedagogy, and policy, requiring a more adaptive and responsive approach to sustain their pivotal role in societal development.

HEIs are undergoing significant transformations due to various factors. Technological advancements have enabled educational platforms to reach a larger number of students at considerably lower costs, or even for free, as demonstrated by the expansion of Massive Open Online Courses (MOOCs). The COVID-19 pandemic further accelerated the adoption of flexible and digital learning models, allowing MOOCs to mature and gain substantial traction within the education sector.

Simultaneously, the term 'university' has seen a dilution of its traditional meaning. Increasingly, companies and organisations are adopting the term to describe internal training and development programmes aimed at upskilling their workforce or educating the public about their platforms.

To add to the complexity, students' brains present unique challenges for educators due to their shorter attention spans and tendencies toward boredom (Schmitt & Lancaster, 2019). These characteristics are attributed to their constant immersion in digital environments. As a result, Gen Z students often struggle with traditional teaching methods, such as lectures and reading, and instead prefer more visual and interactive learning approaches (Schmitt & Lancaster, 2019).

More than 40% of Gen Z adults spend from 1-3 hours per day in social media, while almost 30% spend from 3-5 hours each day (Creatopy, 2022), leading to what Simon (1971, p.40) described as "a wealth of information creates a poverty of attention". Decades later the information has become ubiquitous and produced in mass. Furthermore, the COVID-19 pandemic has exacerbated these challenges by significantly increasing anxiety levels among university students.

At a more regional level, Post-Brexit, the UK has experienced a significant decline in EU student enrolment. This decline has exacerbated financial pressures on universities, particularly with domestic tuition fees frozen since 2012 despite rising inflation (Coughlan, 2022). Hence, UK universities are caught in a "triangle of sadness" between students burdened with debt, a stretched government, and beleaguered staff, noting that these pressures are likely to worsen as the funding crisis deepens.

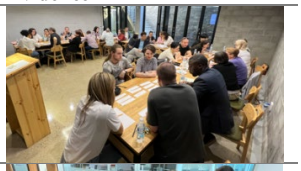
2.3 Research Problem

This methodology was intended to enrich the strategic development process, ensuring that the resulting strategies were robust, adaptable, and reflective of the collective aspirations of the diverse group involved. Additionally, the workshops aimed to contribute to the broader strategic management literature by demonstrating the effectiveness of combining backcasting with tactile collaboration, thereby offering a novel approach for organisations, particularly HEIs, to navigate future challenges and seize emerging opportunities with greater resilience and foresight.

By conducting these participatory workshops in diverse geographical locations—South Korea, the United Kingdom, Slovenia and Chile—and involving participants from twelve institutions and eighteen nationalities, the study aimed to test the framework's applicability across different cultural and institutional contexts. These workshops were designed to facilitate interactive strategic planning sessions focused on the future of HE. Through the hands-on, creative process of building LEGO® models, participants were encouraged to articulate their visions of a desirable future for their institutions and collaboratively work backward to identify the necessary actions to achieve these visions (Table 1). This approach not only

enhanced stakeholder engagement but also promoted creative problem-solving by leveraging the diverse perspectives and expertise of the participants.

Table 1. Visual Summary of the 4 Workshops.

Place	Participants	Evidence
South Korea, 2024	15 Professors, 5 Postgraduate Students, 2 Administrative Staff.	
United Kingdom, 2024	10 Professors, 6 Postgraduate Students, 2 Administrative Staff.	
Slovenia, 2024	8 Professors, 2 Postgraduate Students, 1 Administrative Staff.	
Chile, 2024	7 Professors, 3 Postgraduate Students, 2 Administrative Staff.	

3 Methodology

3.1 Research design and data collection

This study adopts a qualitative research design, utilizing a case study approach to explore and validate the proposed future-proof strategy framework. The qualitative nature allows for an in-depth understanding of participants' perspectives and the nuanced dynamics of strategic planning within HEIs. A case study approach is particularly appropriate for this research as it facilitates a comprehensive examination of the integrated backcasting and tactile collaboration methodologies within real-world settings. By focusing on specific instances of workshop implementation in South Korea, the UK, Slovenia, and Chile, the study captures the contextual variations and cultural nuances that influence strategic development in HE. This approach enables the researchers to gather rich, detailed data and generate insights that are both contextually grounded and transferable to similar organisational environments.

The workshops used primarily the LSP approach which was rooted on Constructivism and Constructionism. Constructivism, as posited by Piaget, focuses on how individuals construct their own understanding through experiences. In the context of LSP, participants built personal models that reflected their individual understandings of what HE could become. This is highly relevant because each participant's background and experience shaped their unique contributions. On the other hand, Constructionism (Papert, 1993) extends this by suggesting that learning happens most effectively when people are actively involved in constructing external artefacts in the world—such as through building models. In this study, LSP allowed participants to physically manifest their ideas about the future, thus facilitating deeper reflection and group discussions. Constructionism in this study means that participants not only construct physical models but co-construct shared meanings of what the future of HE can look like.

Kishita et al. (2024) proposed a comprehensive design framework aimed at guiding researchers and practitioners in the effective planning and operationalisation of backcasting methodologies. This framework systematically addresses the critical questions of When, Which, How, and What to facilitate the successful implementation of backcasting in various contexts. The framework is structured into four primary sections: Usage, Types, Application, and outcomes, each corresponding to the questions and providing a structured approach to backcasting.

South Korea, the UK, Slovenia, and Chile were deliberately chosen for this study due to their complementary profiles and diverse geographical and cultural contexts. South Korea's advanced integration of technology in education, coupled with its highly competitive academic environment, provides valuable insights into innovation and excellence. The United Kingdom, with its longstanding academic tradition and internationalisation, offers a robust framework for research and policy development. Slovenia represents an emerging HE market with a strong emphasis on sustainability and regional collaboration, highlighting progressive approaches to educational reform. Meanwhile, Chile's significant expansion and focus on innovation and entrepreneurship in HE illustrates effective strategies for increasing access and fostering economic relevance. This diverse sample aligns with the United Nations' Sustainable Development Goals, particularly in promoting quality education and sustainable development, thereby enhancing the relevance and impact of the research on an international scale.

Using a purposive sampling approach, the participant selection for this study prioritised individuals with insider knowledge and critical perspectives on HE to ensure targeted data collection, enriching the study's understanding of global educational trends and best practices. Thus, the authors' personal networks were leveraged to obtain the balance of professors of varying seniority, academic staff, and postgraduate students, and the required experience. While workshops took place across different countries, at least half of the participants in each location were from the host country. International participants represented a

diverse range of countries, including Kenya, Chile, Slovenia, Ireland, the UK, South Korea, Poland, North Macedonia, and Estonia.

Data was collected throughout the workshops using multiple qualitative methods, including recordings and transcriptions. A combination of conversations, discussions, and reflections arising from the activities was systematically gathered. Each session was meticulously recorded, and detailed notes were taken to capture both the interactive discussions and the tangible outputs generated through LSP. This comprehensive data collection approach ensured the capture of all relevant information for subsequent analysis. The data included the description of their physical models – created by participants to represent their ideas.

In this qualitative study, the primary data came from participatory workshops using LSP – an innovative facilitation technique that encourages participants to construct physical models representing their visions and ideas, expert interviews, and scenario-building exercises. The focus was on gathering rich, interpretative data that reflects participants' perspectives, imagination, and co-created meanings of the future of HE.

The workshops were structured around an adapted backcasting framework from Kishita et al., (2024), and on average lasted around three and a half hours, where the participants first got to know each other better and then were guided through the steps of the framework. The facilitator remained the same in all workshops to maintain consistency and was always present during the participants' discussions and LEGO® creations to provide support and guidance. Table A1 in the Appendix shows the summary of application of the framework in the project. This strategic planning method was tailored to focus more on conceptual strategy rather than the delegation and implementation of the vision, allowing participants to concentrate on the overarching goals and transformative ideas. The adaptation incorporated elements from Balagatas (2024), ensuring the process was comprehensive and adaptable to the study's specific time constraints and conditions.

In addition to LSP, expert interviews were conducted with educational professionals, policymakers, and thought leaders to supplement the workshop data with authoritative insights and diverse perspectives. Scenario-building exercises were utilised to create narrative pathways from the present state of HE to envisioned future scenarios, enabling participants to explore various potential developments and their implications.

3.2 Data Analysis

The study employed Thematic Analysis (TA) as the primary method for analysing the collected qualitative data. More than 30 hours of recorded conversations and interactions collected throughout the workshops was transcribed and analysed. TA is a flexible and widely used approach that facilitates the identification, analysis, and reporting of patterns (themes) within the data (Braun & Clarke, 2006). This method was chosen for its ability to handle complex and multifaceted data, making it suitable for exploring participants' diverse perspectives on the future of HE.

Following Braun and Clarke's (2006) six-phase framework, the data analysis process began with familiarisation, where researchers transcribed and immersed themselves in the workshop recordings and notes. Initial codes were generated by systematically identifying significant statements and labelling them accordingly. These codes were then collated into potential themes by examining their interconnections and relevance to the research objectives. The themes were reviewed and refined to ensure they accurately represented the underlying data, resulting in a coherent thematic structure. Table 2 shows an extract of the emerging themes from the analysis.

Table 2. Thematic Analysis

Themes	Codes	Extracts from the Transcripts
1. Mind wandering	Unstructured Thought	" When we allow a few moments of unstructured thought, students often make surprising connections that deepen their understanding of the subject matter". S5CH
	Free Exploration	"…By giving both [students and staff] the freedom to explore topics without the constraints of rigid curricula, we naturally spark curiosity and more intentional engagement". S1SLV
Discouraging self-organisation	Group Dynamics	"We're witnessing less and less organic self-organisation, where they [students] form groups around shared interests". S9CH
Coaching	Non-traditional teaching	"The flexibility for both students and staff to design their own sessions has created a natural synergy that truly reflects our collective interests".S6SK
Empowering Originality	Bold ideas	"[as a smaller institution] … It doesn't make sense anymore to focus on chasing trends from bigger and more established universities; we need to have the confidence to make bold decisions that reinforce the importance of making choices that reflect our reality". S11CH
Acknowledging digital disruption	Digital Affectations	"Reducing the use of devices [laptops, tablets and phones] in the classroom must be a priority. Students don't engage deeply when they are overwhelmed by digital distractions." S9SK

	Observed benefits of Digital	"When we implement measures to cut down on unnecessary digital usage, students become more present and reflective during discussions". S8SLV
Prioritizing Staff Wellbeing	Student satisfaction	"For me, staff comes at the centre also […] focusing on our wellbeing isn't just beneficial for us, it has a direct, positive impact on student satisfaction, as we are the human face of the institution and serve as guides on their educational journey". S7NWC "I don't like going above and beyond to treat my students as customers, we are not service providers, we are educators". S2SLV "The university focus in student's experience feels transactional". S2SLV

4 Findings

The study highlights the critical role that fostering mind-wandering and curiosity can play in increasing attention spans within the classroom. By creating environments where both students and professors can self-organise around their interests, free from the constraints of rigid curricula and excessive assessment pressures, there is a natural impetus towards exploring ideas in innovative and unstructured ways. This shift from a prescriptive educational model to one that values organic intellectual exploration supports not only deeper engagement with the subject matter but also encourages the development of intrinsic motivation. As a result, both students and staff are empowered to pursue originality over imitation, thereby cultivating confidence in their visions, decisions, and actions rather than merely chasing trends.

The use of Tactile collaboration transcended mere model-building; it became a language of metaphors. Participants didn't just describe concepts like stakeholder collaboration or pathways to success—they built them, using bridges to represent connections, networks to symbolize interdependence, and even dragons to embody challenges overcome through knowledge. This tactile, metaphorical process demonstrably sparked multifaceted discussions, moving beyond abstract ideas to shared interpretations grounded in physical creations.

The "Geppetto Effect," as discussed in Valencia and Pearce's study (2019), refers to designers' propensity to focus intensively on perfecting a product, often to the detriment of broader business considerations. A similar perspective can be applied to strategy, which should be understood as a dynamic and iterative process requiring craftsmanship, continuous refinement, and ongoing adaptation. Tactile collaboration offers a means for teams to visualise strategic progress, develop prototypes, incorporate feedback, and refine their strategic vision through tangible interactions. By embedding these principles into their strategic frameworks, organisations can cultivate greater flexibility, resilience, and long-term success across diverse sectors.

The results of the interventions reflect six key thematic directions that HEIs could adopt to shape a new, more meaningful future, while outlining potential pathways for further exploration and learning derived from each scenario:

1. Increase of the attention span: Stimulation of mind wandering and curiosity to improve intention and attention in the classroom.
   - Scenario: It is the year 2035, and universities no longer operate under rigid curricula that dictate the learning process. At university, students begin each semester by enrolling in Curiosity Labs, an initiative designed to stimulate exploration through structured mind-wandering. Instead of traditional lectures, professors act as facilitators, presenting weekly exploration themes that invite students to engage in associative thinking and inquiry-driven discussions.
   - Proposed further exploration can be aimed at validating the pedagogical effectiveness of structured mind-wandering, exploring its impact on cognitive flexibility, creativity, and HE problem-solving, and employing longitudinal studies about how self-directed inquiry and associative thinking influence students' ability to engage with complex and interdisciplinary problems.
2. Students and professors self-organise around interest: Structure follows functions and interests and there is freedom to explore ideas without curricula
   - Scenario: In 2035, universities no longer dictate rigid degree pathways. Instead, students and faculty form Knowledge Guilds, dynamic learning communities that self-organise around evolving academic and industry challenges. Rather than passively attending pre-planned courses, students play an active role in shaping their learning journey.
   - Proposed further exploration can employ ethnographic studies of self-organised student groups to examine the emergence of collaborative governance and learning norms and develop comparative case studies of HEIS implementing student-led course design initiatives.
3. Leading to change not to chase: Improve students and staff originality over imitation, stimulating confidence in visions, decisions, and actions.
   - Scenario: In 2035, universities have shifted from being institutions that merely prepare

students for existing job markets to becoming innovation hubs that nurture original thinking and visionary leadership. Students no longer follow prescriptive career pathways but instead engage in "Vision Labs", programmes designed to help them formulate, test, and refine their own unique ideas rather than imitating existing solutions.

- Proposed further exploration can assess the effectiveness of structured ideation and hypothesis-driven thinking in preparing students for future uncertainties and focus on experimental studies evaluating student decision-making abilities.

4. Digital discipline: Reduce screen time to stimulate independent thinking.
   - Scenario: In 2035, students and faculty participate in Digital Detox Days, a weekly initiative designed to counteract the cognitive overload of constant digital engagement. By implementing intentional screen-free moments, universities create an environment where students can reclaim their ability to think independently, free from the constant demands of digital consumption.
   - Proposed further exploration can involve randomised controlled trials measuring cognitive load, attention span, and retention rates in students who participate in Digital Detox Days as well as neuroscientific studies using fMRI or EEG to analyse brain activity differences in screen-based as opposed to screen-free learning.
5. Concentration on staff wellbeing to boost student satisfaction: Highlight professor as the main touch point for student's interactions.
   - Scenario: In 2035, universities have implemented a faculty wellbeing programme that ensures professors have protected time for deep work, research, and professional development. Administrative burdens have been reduced, allowing faculty to focus on meaningful engagement with students. Instead of rushed interactions, they engage in long-form dialogue, exploring ideas in depth.
   - Proposed further exploration can benefit from mixed-methods studies assessing correlations between faculty workload balance, wellbeing metrics, and student satisfaction scores and meta-analyses of faculty retention rates and their relationship to wellbeing interventions.
6. Seeking social impact instead of customer-student impact.
   - Scenario: In 2035, a consortium of European institutions has pioneered a new evaluation model, where universities are judged by their contribution to society rather than customer-driven metrics. Institutions are evaluated based on their engagement with local communities, the launch of social enterprises, and research that benefits the public good.
   - Proposed further exploration can use longitudinal studies tracking university alumni contributions to societal innovation, policy changes, and entrepreneurship and perform comparative institutional analysis of universities that have shifted from customer-based metrics to societal impact indicators.

At a prototype level, these scenarios offer valuable insights into their practical application, potential roadblocks, anticipated benefits, and possible drawbacks. By examining these speculative futures, institutions can better understand the challenges and opportunities associated with their implementation, ultimately informing more effective and sustainable strategic decisions in HE.

Fig. 1 provides an overview of the findings and priorities that stem from these six thematic directions on a matrix of an impact-effort relationship.

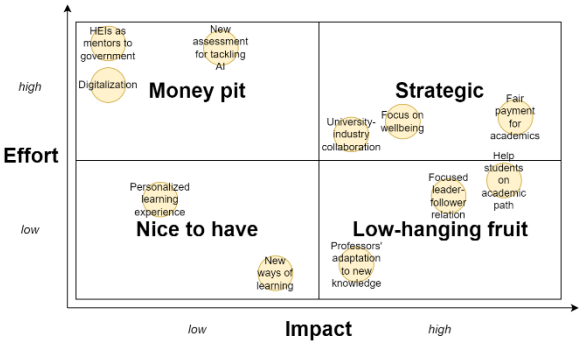


Figure 1. Impact-effort matrix of priorities from identified six thematic directions

5 Discussion

The findings underscore how a participatory, future-oriented framework can enrich HE strategy by bridging foresight and collaborative design. In practice, combining backcasting with LSP enables stakeholders to co-create visionary strategic narratives rather than remain confined to reactive, incremental planning.

Real-world examples illustrate the potential of such frameworks. Kalra et al. (2024) described various participatory backcasting applications, among which in university settings, where lecturers and students engage in envisioning the future of HE. Similarly, LSP has been used in academic contexts for purposes ranging from improving group dynamics to planning curricula (Wheeler, 2023), evidencing its versatility and positive impact on engagement.

Applying this framework in diverse institutional contexts can yield significant benefits for pedagogy, organizational culture, and strategic capacity. Pedagogically, the constructivist underpinnings of the LSP method may inspire a shift toward more experiential and student-cantered learning. When university stakeholders actively "build" future

scenarios, they internalize principles of active learning that could translate into more innovative teaching practices and curricula (Wheeler, 2023). Institutions that embrace such playful, creative exercises might see a cultural shift toward a more open, innovative organizational culture that values collaboration and imaginative thinking (Jerome et al., 2024). Over time, these cultural changes can enhance an institution's strategic capacity. By normalizing foresight activities and creative ideation, universities build internal capability to anticipate and adapt to change. (Ziegler & Porto-de-Oliveira, 2022).

However, we note some research constraints to these implications. First, there may be some potential biases in participant selection and representation. Second, generalizing the results across different cultural and institutional environments is challenging. So, we intentionally spanned workshops across four countries and indeed observed that context matters. The method's playful format, while generally an asset, might be perceived as lacking seriousness by some participants or decision-makers, potentially undermining buy-in. Participants unfamiliar with LEGO® or uncomfortable with creative play may initially struggle, which could limit their initial engagement.

6 Conclusions

Future studies have recently emerged as a valuable source of inspiration for policymaking and strategy development. Depending on the context, future studies adopt different names and scopes; however, the underlying processes and tools remain largely consistent. This study aimed to integrate frameworks drawn from various disciplines to create an amalgamation of tools designed to help organisations learn from what is possible and envision what is preferable.

The combination of tactile collaboration and future thinking benefits from the deliberately slower pace of the intervention, allowing participants time to think carefully and reflect on their responses. In an international setting, anecdotal evidence from this study suggests that this approach can help overcome language barriers by enabling non-native speakers to use tangible metaphors to express complex ideas.

The collaborative story-making process played a central role in shaping future visions during each session. Workshops moved beyond the construction of static models, encouraging participants to weave narratives around their creations. This process of storytelling, prompted by the shared act of building and manipulating LEGO® structures, facilitated the exploration of potential pathways from the present to desired futures. The data generated—particularly the transcribed narratives—appears to contain rich insights into the specific steps, milestones, and potential challenges identified through the sessions.

Future research will expand on these preliminary scenarios, developing them into comprehensive visions and narrative-based scenarios that illustrate possible trajectories for the future of HE. It will also seek to identify the empirical studies and theoretical analyses required to assess the feasibility, desirability, and viability of these envisioned futures.

Acknowledgments

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: HORIZON EUROPE Marie Skłodowska-Curie Actions (MSCA); Grant agreement ID: 101130572.

References

Alexander, B., 2020. 'Connecting the Dots: Metatrends', in Alexander, B. (ed.) *Academia Next: The Futures of Higher Education*. Baltimore: Johns Hopkins University Press, 129-133.

Balagtas, P., 2024. *Making Futures Work: Integrating Futures Thinking for Design, Innovation, and Strategy*. 1st ed. Sebastopol: O'Reilly.

Bibri, S.E., 2018. 'Backcasting in futures studies: A synthesized scholarly and planning approach to strategic smart sustainable city development'. *European Journal of Futures Research*, 6(13).

Braun, V., & Clarke, V., 2006. 'Using thematic analysis in psychology'. *Qualitative Research in Psychology*, 3(2), 77–101.

Coughlan, S., 2022. 'Universities face financial pressures as tuition fees frozen'. *BBC News*. Available at: https://www.bbc.com/news/education-60212345 [Accessed 15 Dec. 2024].

Dell'Era, C., Magistretti, S., Cautela, C., Verganti, R., & Zurlo, F., 2020. 'Four kinds of design thinking: From ideating to making, engaging, and criticizing'. *Creativity and Innovation Management*, 29(4), 637-652.

Garavan, T.N., Darcy, C., & Bierema, L.L., 2024. 'Learning and development in highly dynamic VUCA contexts: A new framework for the L&D function'. *Personnel Review*.

Greenberg, E., Padhi, A., & Smit, S., 2024. '2024 and beyond: Will it be economic stagnation or the advent of productivity-driven abundance? Hoping for a soft landing in 2024 won't mitigate current risks or raise business performance and living standards longer term'. *Strategy & Corporate Finance, McKinsey & Company*. [online].

Hadida, A.G.L., 2013. 'Let your hands do the thinking! Lego bricks, strategic thinking and ideas generation within organizations'. *Strategic Direction*, 29(2), 3-5.

Jerome, L.W., Paterson, S.K., von Stamm, B., & Richert, K., 2024. 'Making transdisciplinarity work for complex systems: A dynamic model for blending diverse knowledges'. *Futures*, 161, 103415.

Kalra, R., Vazirani, K., Kadam, S., & Verma, D., 2024. 'Relevance of Backcasting as a Strategic Tool Towards Organisations' Sustainable Future–A Key to Thrive in the VUCA World'. In *VUCA and Other Analytics in Business Resilience, Part A* (pp. 39-58). Emerald Publishing Limited.

Kishita, Y., Höjer, M., & Quist, J., 2024. 'Consolidating backcasting: A design framework towards a users' guide'. *Technological Forecasting & Social Change*, 202, 123285.

Lee, J., 2018. 'Making hindsight foresight: Strategies and preparedness of failure management'. *Organizational Dynamics*, 47, 165–173.

Magnanini, S., Trabucchi, D., & Verganti, R., 2023. 'Convergence in innovation: The perception of synthesis in articulating a new strategic vision'. *Innovation*, 25(3), 305-327.

Manogue, C.A., Gire, E., & Roundy, D., 2014. 'Tangible Metaphors'. *Conference Paper*.

Norman, D., & Verganti, R., 2013. 'Incremental and Radical Innovation: Design Research vs. Technology and Meaning Change'. *Design Issues*, 30(1), 78-96.

Norman, D.A., 2023. *Design for a Better World: Meaningful, Sustainable, Humanity Centered*. Cambridge, MA: The MIT Press.

Papert, S., 1993. *The Children's Machine: Rethinking School in the Age of the Computer*. Basic Books, New York.

Schmitt, C.A. and Lancaster, R.J., 2019. 'Readiness to practice in Generation Z nursing students'. *Journal of Nursing Education*, 58(10), 604-606.

Thorén, K. and Vendel, M., 2018. 'Backcasting as a strategic management tool for meeting VUCA challenges'. *Journal of Strategy and Management*.

van Vliet, M., & Kok, K., 2013. 'Combining backcasting and exploratory scenarios to develop robust water strategies in face of uncertain futures'. *Mitigation and Adaptation Strategies for Global Change*, 20, 43–74.

Verganti, R., & Öberg, Å., 2013. 'Interpreting and envisioning — A hermeneutic framework to look at radical innovation of meanings'. *Industrial Marketing Management*, 42(1), 86-95.

Verganti, R., 2016. *Overcrowded: Designing Meaningful Products in a World Awash with Ideas*. MIT Press.

Verganti, R., Vendraminelli, L., & Iansiti, M., 2020. 'Innovation and Design in the Age of Artificial Intelligence'. *Journal of Product Innovation Management*, 37(3), 212–227.

Vuorinen, T., Hakala, H., Kohtamäki, M., & Uusitalo, K., 2017. 'Mapping the landscape of strategy tools: A review on strategy tools published in leading journals within the past 25 years'. *Long Range Planning*.

Wheeler, A., 2023. 'Lego® Serious Play® and higher education: encouraging creative learning in the academic library'. *Insights*, *36*(1).

Ziegler, R., & Porto-de-Oliveira, L. C., 2022. 'Backcasting for sustainability–An approach to education for sustainable development in management'. *The International Journal of Management Education*, *20*(3), 100701.

Appendix

Table A1. Application of the Design Framework for Backcasting (Kishita et al., 2024) to this study.

Section	Item	Description	This study
A. When is backcasting used?	**Goals**	Define long-term and highly significant goals.	Define long-term strategic planning for HEIs within the global network belonging to the GETM4 project to address future challenges by 2034. To create the vision of the future that later on it could be translated to goals and objectives, a couple of questions were explored – What is the most meaningful gift that Universities give to students? And What is the most meaningful gift that Students give to society?
	Objectives	Define project objectives, including shared visions, managing conflicts, and developing models.	Develop shared visions for the future of HEIs, gain clarity to combat ambiguity and enhance resilience through human value prioritisations.
	Time Horizon	Specify the timeframe for backcasting projects: short-term (up to 10 years), mid-term (10-30 years), or long-term (over 30 years).	Mid-term 10 years.
	Scale	Determine the scale of the project.	Global scale, involving multiple international HEIs from countries like South Korea, Slovenia, Poland, Ireland, North Macedonia, Estonia, Chile, and Kenya.
	Domain/ Topic	Identify the sectors or topics covered.	HE Strategy, Sustainability, Technological Integration, and Pedagogical Innovations.
	Core Partners and Target Groups	Identify key stakeholders and partners, including researchers, policymakers, and corporate strategists, depending on the interdisciplinary nature of the project.	Researchers, administrative staff, professors and students belonging to the research conglomerate GETM4 representing 14 nationalities.
B. Which type of backcasting is chosen?	**Goal-Oriented or Path-Oriented**	Choose between goal-oriented backcasting (focusing on achieving predefined goals) and path-oriented backcasting (exploring how future scenarios might develop from current trends).	Goal-Oriented Backcasting to establish clear changes necessary to adapt to emerging circumstances in HE. Create a shared vision among the participants.
	Degree of Participation	Determine the level of stakeholder involvement, ranging from non-participatory to highly participatory approaches involving various stakeholders such as citizens, policymakers, and industry experts.	Highly Participatory, involving stakeholders through three international workshops using LSP and other participatory methods.
	Qualitative or Quantitative	Decide whether the backcasting will use qualitative data, quantitative data, or a combination of both based on the project's focus.	Qualitative data from workshops participants.
C. How is backcasting applied?	**Process**	Outline the sequence of steps and activities, including problem framing, scenario development, and pathway exploration, aligned with the project's objectives.	Use of LSP to create shared visions within the participants. Followed by a Backcasting exercise to create a roadmap of actions to build a desirable future.
	Methods and Tools	Select appropriate methods and tools such as design techniques, analytical models, participatory workshops, interviews, questionnaires, logic trees, road mapping, and simulation models to support the backcasting process.	Problem identification, Problem Framing, Interactive Problem Solving and Visualisation using LSP, activity Prioritisation Matrix and Human Value prioritisation.
D. What are the results and outcomes?	**Content Results - Conclusions**	Generate design elements like visions, pathways, proposals, and interventions; analytical findings; knowledge regarding policy and sustainability; models and simulations; and refined methods and tools.	The development of six thematic directions that synthesise the topics and reflections gathered from the four workshops.

Digital Twin-Enhanced Fire Response Through Unmanned Aerial Reconnaissance System

Domagoj Tuličić, Snježana Križanić
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{domtulicic, skrizanic}@foi.unizg.hr

Abstract. *This paper first examines the digital twin as a conceptual framework and then as an implementation method. Several existing definitions are reviewed, and a new one is proposed, emphasising the time dimension where the digital object represents a future state of the physical object. A key element is the inclusion of an observer or decision-maker (man in the loop). An AI-based automated decision-making system (AI agent) is also proposed and discussed within this framework. The concept is applied to a real-world case: the coastal reconnaissance segment of the Croatian fire-fighting system.*

Keywords. Digital twin, unmanned aerial vehicles, fire-fighting system, human in the loop, AI Agent.

1 Introduction

In recent years, particularly during the summer months, wildfires have become a frequent and severe threat in the Republic of Croatia and across other Mediterranean countries. Despite Croatia's long-standing firefighting tradition and a relatively well-developed fire protection system, wildfires still occur with potentially devastating consequences, including significant material damage and loss of human life. According to the Croatian Firefighting Association, in 2024, the ten largest fires affected a total of 13,378 hectares, which corresponds to 133.78 million square meters (Croatian Firefighting Association, 2025 (a)).

The primary causes of such large-scale fires include mild winters, more frequent and intense heatwaves compared to previous years, and human-related factors. To protect coastal and near-coastal areas, where the most destructive fires typically occur, Croatian firefighting forces carry out systematic observation and reconnaissance. This is done using cameras installed primarily on antenna towers operated by *Odašiljači i veze Ltd*. (OiV, the national provider of broadcasting and communication infrastructure in Croatia) as well as on infrastructure belonging to mobile network operators. In certain cases, manned aircraft and unmanned aerial vehicles (UAVs) are deployed. By 2024, 220 cameras had been installed at 110 locations, providing 360-degree visual coverage (Zaštita.info, 2025). In addition, systems such as the Pilatus PC-9 and the Orbiter 3 UAV are used for early fire detection, reconnaissance in hard-to-reach areas, and real-time fire monitoring (Croatian Firefighting Association, 2025 (b)).

These facts clearly show that technology plays a critical role in fire prevention and early warning, leaving room for further enhancements. One promising direction is the development of a digital twin of the early fire detection system. Such a model could include spatial mapping of camera coverage, detection of under-monitored areas, and the application of optimization algorithms to support UAV-based reconnaissance. In uncovered regions, UAV routes can be optimized using bio-inspired algorithms such as Max-Min Ant Colony Optimization (MMAS), which dynamically adapt to spatial and temporal variables, including the fire weather index.

In addition, the proposed framework incorporates an AI-based automated decision-making system (AI agent) within the digital twin. This agent enhances the model's ability to analyse incoming data, optimise UAV reconnaissance routes, and support timely decision-making in dynamic wildfire conditions. The aim of this research is to design a digital twin of the fire detection subsystem.

2 Theoretical background of digital twins

The digital twin is one of the key concepts emerging within the paradigm known as the Fourth Industrial Revolution (Industry 4.0). This revolution is based on the integration of digital technologies, such as the Internet of Things (IoT), artificial intelligence, automation, and simulation, into industrial processes. The concept of the digital twin was first applied during research space missions conducted by NASA in the 1960s (NASA, 2025) (Allen, 2021). During the Apollo 13 mission, an oxygen tank explosion endangered the crew, but thanks to NASA engineers testing real-time scenarios on physical and digital simulators, the astronauts returned safely, an experience now seen as

a precursor to today's concept of the digital twin: a continuously updated simulation that mirrors a real system.

The idea of simulating a real-world system that receives data and responds dynamically, much like the physical system itself, was also proposed by computer scientist David Gelernter in his 1991 book Mirror Worlds (Gelernter, 1991). The term digital twin was first used in this context by Santiago Hernandez & Luis A. Hernandez-Ibanes (1997), authors of the paper Application of Digital 3D Models on Urban Planning and Highway Design. The paper highlights the advantages of creating three-dimensional digital models for construction projects and demonstrates how such models overcome the limitations of conventional design through three practical case studies.

The foundations for applying the digital twin concept in industry were laid by American scientist Michael Grieves in 2002, when he first comprehensively formulated the concept within the context of Product Lifecycle Management (PLM). In his work Origins of the Digital Twin Concept (Grieves, 2016), Grieves described the key elements of a digital twin as the physical space, the virtual space, and the connection between the two throughout the entire product lifecycle.

Križanić & Vrček (2025) used simulation experiment as a method for developing digital twin of a real production business process. The authors also proposed a new methodology for developing production digital twins, building upon existing business process management frameworks.

In 2012, NASA formalized its own definition of the digital twin, emphasizing it as the integration of multiple models, sensor data, and analytical tools into a cohesive virtual representation of a physical entity (Glaessgen & Stargel, 2012).

Various researchers have presented similar definitions of the digital twin. For example, the digital twin is a computer model of a physical device or system that represents all its functional features and is connected to its operational components (Chen, 2017). A digital twin is essentially a living model of a physical asset or system that continuously adapts to operational changes based on collected network data and information, and is capable of predicting the future state of the corresponding physical system (Liu, 2025). It is a set of virtual information that fully describes a potential or real physical product, ranging from the micro (atomic) level to the macro (geometric) level (Zheng et al., 2019). A digital twin is a digital representation of a physical item or assembly that uses integrated simulations and service data. This digital representation incorporates information from multiple sources throughout the entire product lifecycle (Vrabič, 2018). A digital twin is a virtual instance of a physical system that is continuously updated with data on its operation, maintenance, and condition throughout the system's entire lifecycle (Madni et al., 2019). It can represent both living and non-living entities, such as manufacturing processes, medical devices, or even people, by integrating sensor data, simulation, and analytics to provide insight into current and future operational states (Interagency Modeling and Analysis Group, 2025). Designed to accurately reflect its physical counterpart, a digital twin incorporates real-time data, simulation, machine learning, and reasoning to support decision-making across the lifecycle of the object or system (IBM, 2025).

According to IBM (2025), digital twins can be categorized based on their application domain and defined at various levels, from components to entire processes. Different types of digital twins may coexist within the same system or process. The main types include:

1. Component or part twins, which represent the smallest functional elements. While similar, part twins typically refer to less critical components.
2. Asset twins, which consist of two or more interacting components. They enable the analysis of component interactions and performance, generating valuable insights.
3. System or unit twins provide a broader view of how assets work together to form a complete, functional system. They help identify potential performance improvements.
4. Process twins operate at the highest (macro) level, representing how multiple systems collaborate across a production environment. They help determine whether systems are synchronized for maximum efficiency and identify delays that affect overall performance.

The concept of the digital twin has evolved from early space exploration applications to a foundational element of modern industrial systems, supported by advancements in data integration, simulation, and real-time analytics. Its layered structure, ranging from component-level twins to process-level representations, enables comprehensive modeling and optimization of complex systems throughout their entire lifecycle. This evolution and the diversity of interpretations were comprehensively reviewed by Dalibor et al. (2022), who analysed 1,471 publications, of which 356 were examined in detail, identifying conceptual properties, engineering practices, and evaluation methods. Building on the considerations presented above, this paper proposes the following definition of a digital twin: *A digital twin is an extension of a sociotechnical system composed of a physical element, integrated through automated information channels with its digital representation in both its current state and near future, and a human actor as a corrective factor within the system*.

3 Methodology

The methodology applied in this study follows a multi-layered approach to the development of a digital twin of a firefighting system, aimed at early wildfire

detection and prevention. The digital twin architecture is structured into three core layers: the physical entity, the digital model, and the AI-supported decision-making layer.

The physical layer includes georeferenced surveillance infrastructure such as PTZ cameras distributed across coastal and near-coastal regions of Croatia, unmanned aerial vehicles (UAVs) equipped with high-resolution and thermal sensors, and a centralized data center. The digital model layer integrates multiple spatial and environmental data sources, including camera coverage zones, uncovered terrain points, UAV flyover coordinates, and fire risk indicators such as the Fire Weather Index (FWI), calculated using meteorological data. The UAVs' operational status, flight range, and current positioning are also modeled digitally.

The decision-making layer is supported by an AI agent capable of integrating heterogeneous inputs to autonomously generate UAV reconnaissance routes, dynamically prioritize zones based on FWI values and weather changes, and initiate missions in high-risk areas without human intervention. However, the system maintains a human-in-the-loop configuration, where the operator performs data verification, overrides automated decisions in critical scenarios, and provides feedback to improve system adaptability.

For geospatial processing and visualization, all maps used in the study were created using QGIS, an open-source geographic information system. The maps were generated in the EPSG:3765 projection (HTRS96/TM), the official coordinate reference system for the Republic of Croatia. QGIS was used to define camera coverage zones, visualize UAV flight paths, and overlay FWI values to identify critical surveillance gaps.

To communicate system structure and component interactions, a UML component diagram was employed in Visual Paradigm, capturing the high-level architecture and illustrating interfaces between modules such as the data center, AI agent, digital model, and physical subsystems.

4 Digital twin of the fire-fighting system for fire prevention and early detection

This section describes a digital twin of a firefighting system designed for fire prevention and early detection. The digital twin belongs to the category known as system or unit twins. In the following, the architecture of the digital twin, its key functional components, and the way it supports the optimization of unmanned aerial vehicle (UAV) routing within the firefighting system are presented.

4.1 Physical object

In this digital twin, the physical object is the firefighting system, more precisely, its subsystem that is, in this context, considered an independent system focused on fire prevention and early detection. The physical object includes the following elements: the monitored terrain, surveillance panoramic cameras, unmanned aerial vehicles (UAVs), a data center.

The monitored terrain represents a system element which, although not a device like cameras, data centers, or UAVs, is nevertheless a key part of the physical reality that must be captured for the digital representation to be meaningful. Since the terrain is observed, measured, modeled, and used for decision-making, it constitutes the central physical entity. Without its digital representation, fire simulations and predictions would lack practical value.

Surveillance panoramic cameras are deployed across the coastal and near-coastal areas of the Republic of Croatia. A total of 230 cameras have been installed at 115 locations, with each location equipped with two PTZ (Pan-Tilt-Zoom) cameras. This type of camera enables horizontal rotation up to 360°, vertical tilt up to 90°, and optical and/or digital zoom, allowing for detailed monitoring of objects at long distances. Fig. 1 shows the locations of the panoramic cameras in Croatia. Each dot on the figure represents a single location, and since each location contains two cameras, it is assumed that the system is designed with redundancy to ensure operational reliability.

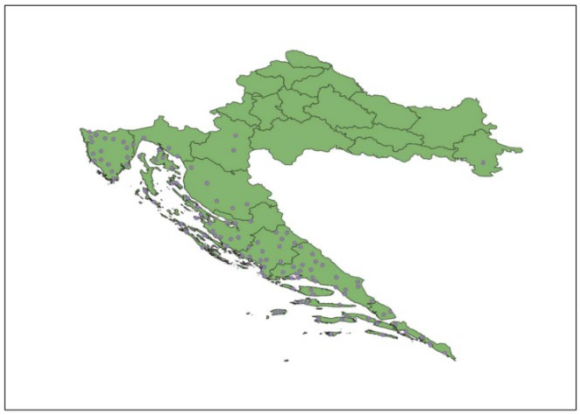


Figure 1. The spatial distribution of PTZ cameras for fire detection in the EPSG:3765 projection

Unmanned aerial vehicles (UAVs) serve as mobile sensor platforms used for reconnaissance of remote or hard-to-reach areas that are either not covered by cameras or where existing cameras are non-operational for various reasons. These UAVs are equipped with high-resolution cameras, thermal sensors, and/or air quality sensors, enabling rapid data collection over large terrain areas. The UAVs are connected to a communication system and transmit the collected data in real time to a central server, where it is further processed and integrated into the digital twin

representation. Their use enables early detection of fire indicators such as elevated temperatures, smoke, or changes in vegetation.

The data center represents the physical component of the system where data collected from the field are stored, processed, and distributed. This includes images and videos from surveillance cameras, data from unmanned aerial vehicles (UAVs), as well as information on topography, vegetation, and microclimatic conditions. Depending on the system architecture, the data center may be located locally (i.e., edge computing) or in the cloud. In either case, it plays a crucial role in creating and maintaining an up-to-date digital representation of the real world. Its computational power enables real-time processing of large volumes of data, which is essential for timely detection and response to potential threats such as wildfires. In addition, the data center hosts algorithms for optimizing terrain reconnaissance, including one that calculates the optimal UAV flight route based on uncovered areas, weather conditions, and available resources.

4.2 Digital representation

The digital model represents a computer-based representation of the system for fire prevention and early detection. It is based on georeferenced data regarding the locations of surveillance cameras and the coordinates of areas not covered by the cameras, which need to be monitored by unmanned aerial vehicles (UAVs). The model also includes indicators of fire risk, such as the Fire Weather Index (FWI), which is used to assess the likelihood of fire ignition and spread based on meteorological data, as well as a digital model of the UAV itself, which in its basic version indicates the current position and flight path of the UAV in space. Thus, the model consists of multiple interconnected layers.

The first layer, as previously mentioned, contains spatial data on the positions of the surveillance cameras and points that are not within their fields of view. Around each camera, a zone with a radius of 10 kilometres is defined, and the points outside these zones represent areas that require additional UAV reconnaissance. Fig. 2 shows the zones covered by the surveillance cameras. However, camera visibility is limited by terrain relief Geographic areas not covered by cameras are indicated by red points, which also represent UAV flyover points. These points are either manually entered by an operator or automatically generated by an artificial intelligence-based algorithm.

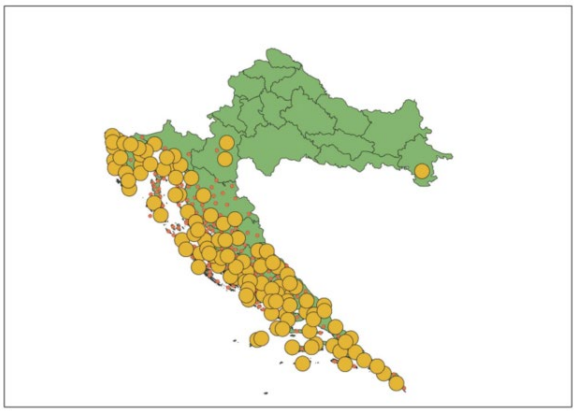


Figure 2. Map showing UAV flyover points (EPSG:3765 projection)

The second layer of the model is the temporal-dynamic component, which includes values of the Fire Weather Index (FWI). The Fire Weather Index is a meteorological indicator widely used around the world to assess wildfire danger. It consists of various components that take into account the effects of fuel moisture and wind on fire behavior and spread. The higher the FWI value, the more favorable the meteorological conditions are for fire ignition and propagation (Copernicus, 2025). Although FWI is typically calculated once per day, within this digital model it can be updated more frequently, depending on the availability of high temporal resolution meteorological data.

Based on the FWI values, reconnaissance priority is assigned to specific areas. In cases where the FWI is low, the UAV may omit such locations from its route and instead focus on zones with increased risk. In this way, the flight path is optimized, and the efficiency of territory surveillance across the Republic of Croatia is improved.

Fig. 3 shows a map of the Republic of Croatia with calculated FWI values over an extended time period, while Table 1 presents the classification values used to interpret wildfire risk levels. If the FWI values for a specific area fall between 0 and 21 (see Table 1), the UAV will not pass-through points within that area, as the fire risk is considered low to moderate.

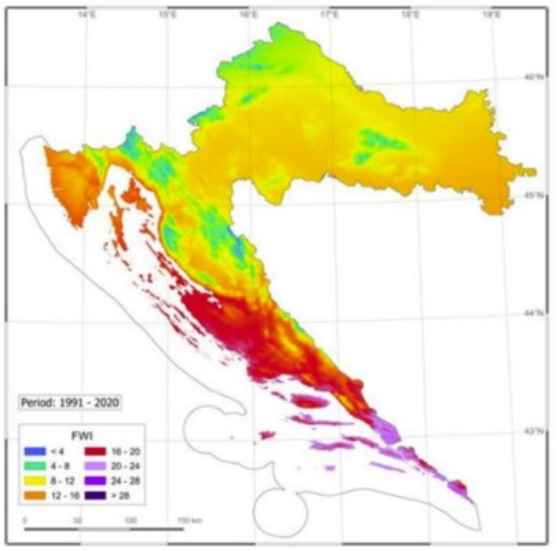


Figure 3. Fire Weather Index over an extended time period (Source: https://www.sumins.hr/en/projekti/modflux/)

Table 1. Interpretation of FWI values

FWI Range	Fire Danger Rating	Comment
0 – 5	Very Low	Moist conditions, low chances of fire ignition or spread.
5 – 12	Low	Moderately dry conditions, fire may ignite but spread is limited.
12 – 21	Moderate	Conditions allow ignition and limited fire spread.
21 – 30	High	Dry and windy conditions, rapid fire spread is possible.
30 – 50	Very High	Very favorable conditions for uncontrolled fire spread.
> 50	Extreme	Extremely hazardous conditions; fire spreads quickly and is hard to control.

The third layer of the model refers to the operational elements of the system, specifically the digital representation of unmanned aerial vehicles (UAVs), their sensor capabilities, limitations in range and flight duration, and their current status (position, battery level, activity). UAVs in the model can be deployed based on predefined rules or route optimization algorithms, and their movement is simulated in real time.

The digital model also includes a logical component that connects input data with decision-making algorithms. Based on defined rules and thresholds, the model can automatically recommend changes in reconnaissance schedules, issue alerts to operators, or generate reports on uncovered high-risk zones.

This integrated approach enables not only real-time monitoring and planning but also short-term forecasting and the execution of simulations for training purposes, system efficiency assessment, and strategic decision-making in the context of fire prevention.

4.3 Human in the Loop

In the context of the firefighting system digital twin for early fire detection, the human (operator or analyst) remains a key element in the decision-making loop, especially in situations where the automated system is unable to make an optimal decision or when system verification is required. This human role, known as Human in the Loop (HITL), serves as a corrective and supervisory-decisional factor within an otherwise highly automated system.

In the implemented digital twin, the human fulfils multiple roles:

1. Data verification: Validates the accuracy of data automatically collected and processed by system components (e.g., cameras, sensors on UAVs).
2. Decision-making: Based on the current and projected system state (e.g., fire prediction using the FWI index), the human decides on activating UAV reconnaissance.
3. Intervention: In cases of unexpected system behavior (e.g., UAV malfunction, sudden weather changes), the human can modify algorithm-generated flight plans.
4. System learning and adaptation: Feedback provided by the human can be used to train and improve the performance of predictive models within the digital twin.

The role of the human is not only reactive but also proactive, as it involves interpreting complex scenarios and making decisions that go beyond the current capabilities of automated systems. Therefore, the digital twin does not replace the human in the system, but rather provides tools to support better situational assessment and more effective responses aimed at fire prevention and environmental protection.

5 Concept of developing an automated decision-making system (AI agent)

The role of the human as a corrective factor within the fire surveillance and reconnaissance system remains indisputable, particularly in decision-making under complex and unpredictable conditions. However, with the advancement of technological infrastructure and the increasing availability of real-time data, there is a

growing need for a higher degree of automation. In this context, the concept of an automated decision-making system based on artificial intelligence becomes particularly prominent.

Such an AI agent, implemented within the digital twin of the firefighting system, is responsible not only for analyzing and interpreting data but also for making decisions within predefined boundaries of autonomy.

The level of autonomy of the AI agent may vary depending on the context and complexity of the situation. At the lowest level, the agent functions as a recommendation system, providing suggestions to the operator, while the human still retains final decision-making authority. In more functionally advanced scenarios, the agent autonomously handles routine tasks, such as identifying reconnaissance areas based on fire danger index values, while seeking human confirmation for more complex decisions. At the highest level of autonomy, the agent is capable of making and executing decisions in real time, relying on previously learned patterns, safety rules, and system operational constraints.

Within the digital twin, the AI agent integrates various input data, such as geospatial information on terrain coverage by surveillance cameras, Fire Weather Index values, real-time weather forecasts, operational status and availability of unmanned aerial vehicles (UAVs), historical wildfire data, and records of surveillance equipment failures. Based on this information, the agent is capable of autonomously generating optimal reconnaissance routes, dynamically adjusting priorities in response to changing weather conditions and assessed risks, estimating the required number of UAVs for a given operation, and, in emergency situations, initiating reconnaissance without the need for human intervention.

The goal of implementing such a system is not to eliminate the human from the decision-making process, but rather to increase efficiency and reduce system response time in situations that require urgent action. This creates the foundation for advanced early warning systems, in which the digital twin functions not merely as a passive reflection of reality, but as an active, predictive, and adaptive entity capable of acting in real time.

6 UML component diagram

The system architecture is modelled using a UML component diagram to illustrate the main modules and their interactions within the proposed framework. The digital twin is central to this architecture, integrating real-time surveillance data, predictive simulation models, and AI-based decision-making to support early wildfire detection and reconnaissance.

Fig. 4 shows the main components, including the unmanned aerial vehicle (UAV) fleet, data centre, digital twin model, AI agent, optimisation algorithms (such as MMAS), and camera network, as well as the interfaces and dependencies between them. The diagram highlights how data flows from surveillance cameras and UAVs to the data centre, where it is processed and fed into the digital twin. The AI agent analyses this data, optimises UAV reconnaissance routes, and can initiate missions in urgent situations.

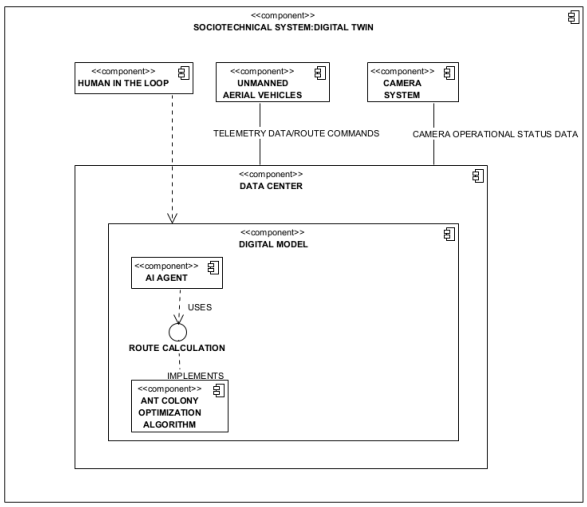


Figure 4. UML component diagram

In addition to the component diagram, other UML diagrams could also have been used to represent the system from different perspectives, such as use case diagrams, activity diagrams, sequence diagrams, class diagrams, and state diagrams. The component diagram was selected because it provides a clear and structured representation of the high-level architecture of the digital twin system. It effectively illustrates the key components, along with their interactions and dependencies. This makes it particularly suitable for communicating the modular structure and functional relationships within a complex socio-technical system.

System components:

- **Sociotechnical system** – The integrated framework combining human operators, technical infrastructure, and organisational processes for wildfire detection and reconnaissance.
- **Digital twin** – A virtual representation of the physical reconnaissance system that continuously integrates real-time data, enabling simulation, prediction, and decision-support.
- **Unmanned aerial vehicles (UAVs)** – Remotely piloted aircraft equipped with sensors and cameras for aerial reconnaissance of under-monitored or high-risk areas.
- **Camera system** – A network of fixed surveillance cameras providing continuous visual monitoring of wildfire-prone zones.
- **Telemetry data / route commands** – Data sent from UAVs to the data centre, including location, status, and sensor readings, and route instructions sent back from the control system.

- **Camera operational status data** – Information on the functional state of each camera, used for fault detection and maintenance planning.
- **Data centre** – The central hub for data collection, processing, and storage, hosting the digital twin and AI modules.
- **Digital model** – The core simulation environment within the digital twin that integrates real-time and historical data to predict system behaviour.
- **AI agent** – An AI-based automated decision-making module that analyses data, optimises UAV routes, and can autonomously initiate actions.
- **Route computation** – The process of calculating optimal UAV reconnaissance paths based on fire weather index, camera coverage, and UAV availability.
- **Algorithm** – Computational methods for optimisation, such as Max-Min Ant Colony Optimisation (MMAS), which dynamically adapts to changing conditions.

The component diagram was chosen because it provides a clear, high-level view of the architecture of the digital twin system and effectively conveys the modular structure and functional relationships within a complex socio-technical system.

7 Conclusion and future work

This paper presented a conceptual and architectural framework for the development of a digital twin of a firefighting system designed for early fire detection and prevention. By integrating physical components such as surveillance cameras, unmanned aerial vehicles (UAVs), and data centers with a multilayered digital model and an AI-based decision-making agent, the system enables real-time monitoring, adaptive response, and strategic planning. Special emphasis was placed on the interoperability between human operators and automated processes, highlighting the role of the human-in-the-loop in ensuring system reliability and informed decision-making.

The proposed approach demonstrates the potential of digital twins not only as passive representations of physical systems but as active, predictive, and autonomous tools capable of supporting complex operations in dynamic environments. The incorporation of real-time data streams, AI-driven route optimization, and scenario-based simulations lays the foundation for advanced early warning systems in wildfire management.

Future research will focus on enhancing the system's adaptability and autonomy by incorporating additional environmental and operational data into the optimization process. In particular, UAV routes in uncovered regions can be optimized using bio-inspired algorithms such as Max-Min Ant Colony Optimization (MMAS), which dynamically adapt to spatial and temporal variables, including the fire weather index.

References

Allen, B. D. (2021). Digital Twins and Living Models at NASA. *Work of the US Gov. Public Use Permitted*.

Chen, Y. (2017). Integrated and Intelligent Manufacturing: Perspectives and Enablers. *Engineering, vol. 3, no. 5*, pp. 588–595, doi: https://doi.org/10.1016/J.ENG.2017.04.009.

Copernicus (2025). Fire Weather Index. Retrieved from https://climate.copernicus.eu/fire-weatherindex (accessed May 18, 2025).

Croatian Firefighting Association (2025) (a). "More than 35,000 firefighting interventions recorded in 2024." (Original title: "U 2024. godini zabilježeno više od 35.000 vatrogasnih intervencija") Retrieved from https://hvz.gov.hr/vijesti/u-2024-godini-zabiljezenovise-od-35-000-vatrogasnih-intervencija/4954 (accessed Apr. 13, 2025).

Croatian Firefighting Association (2025) (b). "Croatian Firefighting Association – Fire Season." (Original title: "Hrvatska vatrogasna zajednica - Požarna sezona"). Retrieved from https://hvz.gov.hr/izdvojeno-11/pozarna-sezona/117 (accessed Apr. 13, 2025).

Dalibor, M., Jansen, N., Rumpe, B., Schmalzing, D., Wachtmeister, L., Wimmer, M., & Wortmann, A. (2022). A cross-domain systematic mapping study on software engineering for digital twins. *Journal of Systems and Software*, 193, 111361. https://doi.org/10.1016/j.jss.2022.111361

Gelernter, H. D. (1991). Mirror Worlds, Or, The Day Software Puts the Universe in a Shoebox - How it Will Happen and what it Will Mean. *Oxford University Press*.

Glaessgen, E. H. and Stargel, D. S. (2012). The digital twin paradigm for future NASA and U.S. air force vehicles, *53rd AIAA/ASME/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf.* 2012, pp. 1–14.

Grieves, M. (2016). Origins of the Digital Twin Concept. DOI: 10.13140/RG.2.2.26367.61609.

Hernandez-Ibañez, L. and Hernandez, S. (1997). Application of Digital 3D Models on Urban Planing and Highway Desing.

IBM (2025). What Is a Digital Twin? Retrieved from https://www.ibm.com/think/topics/what-is-adigital-twin (accessed Apr. 23, 2025).

Interagency Modeling and Analysis Group (2025). Digital Twin Overview. Retrieved from

https://www.imagwiki.nibib.nih.gov/content/digital-twin-overview (accessed Apr. 23, 2025).

Križanić, S. and Vrček, N. (2025). Methodology of the Development of Production Digital Twins Using a Simulation Experiment. *Journal of Computer Science, 21(6)*, pp. 1379-1390. https://doi.org/10.3844/jcssp.2025.1379.1390.

Liu, Z. et al. (2025). The Role of Data Fusion in Predictive Maintenance Using Digital Twins BT – Handbook of Nondestructive Evaluation 4.0. N. Meyendorf, N. Ida, R. Singh, and J.33 Vrana, Eds. Cham: *Springer Nature Switzerland*, pp. 1–23.

Madni, A. M., Madni, C. C. and Lucero, S. D. (2019). Leveraging Digital Twin Technology in Model-Based Systems Engineering. *Systems, vol. 7, no. 1.*, doi: 10.3390/systems7010007.

NASA (2025). Digital Twins and Living Models at NASA - NASA Technical Reports Server (NTRS). Retrieved from https://ntrs.nasa.gov/citations/20210023699 (accessed Apr. 17, 2025).

Vrabič, R. et al. (2018). Digital twins: Understanding the added value of integrated models for through-life engineering services. *Procedia Manuf., vol. 16*, pp. 139–146. Retrieved from https://doi.org/10.1016/j.promfg.2018.10.167.

Zaštita.info (2025). "Transmitters and Communications Expand the Intelligent System for Early Fire Detection in the Republic of Croatia." (Original title: "Odašiljači i veze proširuju inteligentni sustav za ranu detekciju požara u RH."). Retrieved from https://zastita.info/hr/novosti/odasiljaci-i-veze-prosirujuinteligentni-sustav-za-ranu-detekciju-pozara. 34848.html (accessed Apr. 13, 2025).

Zheng, Y., Yang, S. and Cheng, H. (2019). An application framework of digital twin and its case study. *J. Ambient Intell. Humaniz. Comput., vol. 10, no. 3*, pp. 1141– 1153, doi: 10.1007/s12652-018-0911-3.

Digital Transformation and Corporate Culture

Anamarija Zron
Impact of Technostress on the Adoption of Artificial Intelligence - A Bibliometric Analysis

Krešimir Kavran, Maja Pušnik
Planning and Development Processes for the Information System Based on the Command Query Responsibility Segregation Design Pattern

Lebohang Nkhabu, Machdel Matthee
Barriers to Cyber Hygiene in Hybrid Work Environments: A Case Study of a Lesotho Development Parastatal

Daniel Fernando Olmedo Mejía, Violeta Cvetkoska
From Human-Centered to Hybrid: Mapping the Emergence of GenAI in Organizational Actor-Networks

Zvonimir Vukić
Digitalization of Accounting in the 21st Century – A Bibliometric Analysis

Anamarija Horvat, Renata Mekovec
Protecting privacy on social media platforms: a case study of Croatian users

Krešimir Kavran, Boštjan Šumak
Indirect User Experience Evaluation: Insights from Employees Working on a Public Sector Information System

Igor Pihir, Ivan Kardašić
Digital Transformation in the Gaming Industry

Impact of Technostress on the Adoption of Artificial Intelligence - A Bibliometric Analysis

Anamarija Zron
University of Rijeka
Faculty of Economics & Business
Department of Computer Sciences
Ivana Filipovića 4, 51000 Rijeka
anamarija.zron@efri.uniri.hr

Abstract. *This study presents a bibliometric analysis of research on technostress and its impact on the adoption of Artificial Intelligence (AI) from 2022 to 2025, a period marked by the rapid rise of generative AI tools such as ChatGPT. Based on 19 carefully selected publications from the Web of Science database, including the top 10 most cited papers, the paper identifies key research themes, disciplinary intersections, and industry-specific insights. Grounded in the Technostress Model by Tarafdar et al. (2011), findings show that technostressors, particularly techno-insecurity, techno-complexity, and techno-uncertainty, are critical barriers to effective AI adoption across sectors such as IT, healthcare, finance, and tourism. The results highlight how technostress varies by industry, reflecting variations in digital maturity and employee readiness. Practical implications are outlined, with emphasis on strategies to reduce employee stress and resistance. Overall, the paper contributes to a deeper understanding of human factors in AI implementation and offers a foundation for future research on managing tech-driven organizational change.*

Keywords. Technostress, Artificial Intelligence, AI Adoption, Bibliometric Analysis, VOSviewer

1 Introduction

Artificial Intelligence (AI) represents a major technological revolution across various sectors, particularly in the IT industry. However, its integration into organizational settings also brings a significant rise in technostress among employees, which may, in turn, hinder their willingness to adopt these advanced technologies (Chang et al., 2024). Technostress studies have evolved over the past thirty years, from their roots in biology and psychology to being an interdisciplinary field with significant organizational, behavioral, and human resources implications (Grummeck-Braamt et al., 2021).

With the adoption of AI technologies, workers have to learn new competencies, implement new work techniques, and integrate AI systems within their regular work activities (Pereira et al., 2023; Zirar et al., 2023).

Kumar et al. (2023) and Malik et al. (2021) identify technological complexity as a primary stressor, referring to it as 'techno-complexity,' while also addressing the equally important theme of 'techno-insecurity,' which underscores the potential risk of job displacement due to automation and AI implementation. Kim & Lee (2024) found that the adoption of AI leads to increased job stress, which eventually results in burnout. On the contrary, they state that high self-efficacy in learning diverse functionalities of AI reduces this effect. These insights highlight the necessity of organizations to manage stress and support employee well-being. Moreover, numerous studies have explored the effect of AI adoption on employees' mental health, with a specific focus on workplace exhaustion, stress, and burnout syndrome (Bankins et al., 2023; Budhwar et al., 2022; Pereira et al., 2023).

Despite everything stated above, the sudden adoption of AI technology has enabled companies to simplify complex processes, automate conventional tasks, enhance decision-making, and optimize resource utilization (Chen et al., 2023; Makridis & Mishra, 2022; Uren & Edwards, 2023).

This paper builds on the Technostress Model developed by Tarafdar et al. (2007, 2011), which identifies key stressors such as techno-overload, techno-invasion, techno-complexity, techno-insecurity, and techno-uncertainty. These constructs help explain how the rapid integration of AI technologies can lead to employee stress, resistance, and decreased productivity. By analyzing the literature through this lens, the study contributes to a deeper understanding of how technostress influences AI adoption in different organizational contexts.

To the best of the author’s knowledge, no comprehensive bibliometric review on the impact of technostress on the adoption of artificial intelligence exists to date. As stated by Zupic & Čate (2015), bibliometric analysis helps authors to understand the composition of the research field, key concepts, significant authors, and associations.

The research addresses a clear and timely gap by analyzing how, and in which industries, technostress

most significantly influences AI adoption - focusing on the rapid rise of AI technologies since 2022. This timeframe (2022–2025) was deliberately chosen, as 2022 marks the emergence of generative AI tools and widespread usage. The study aims to capture how academic research has evolved in response to this technological shift.

Scientifically, this study maps recent research trends and gaps in the field, helping scholars understand where the field stands and what remains unexplored. Importantly, all findings are analyzed through the lens of the Technostress Model proposed by Tarafdar et al. (2007, 2011), which distinguishes between key stressors (such as techno-overload, techno-insecurity, techno-complexity, techno-invasion, and techno-uncertainty) relevant to AI adoption. By systematically linking these theoretical constructs with the most common keywords and thematic clusters identified via VOSviewer among 19 papers, the study provides an overview of how the core dimensions of technostress specifically manifest in the context of AI across industries. Practically, the study provides organizations with evidence-based recommendations for reducing technostress and facilitating a smoother transition to AI adoption by identifying key psychological stressors such as anxiety, job insecurity, and resistance to change. Ultimately, the goal is to answer the following research questions:

RQ1: How has the research landscape of technostress and its impact on AI adoption evolved, particularly in the last few years, with the rapid advancement of AI technologies?

RQ2: What are the dominant fields of research and emerging trends on technostress and AI adoption, as revealed by the analysis in the bibliometric study?

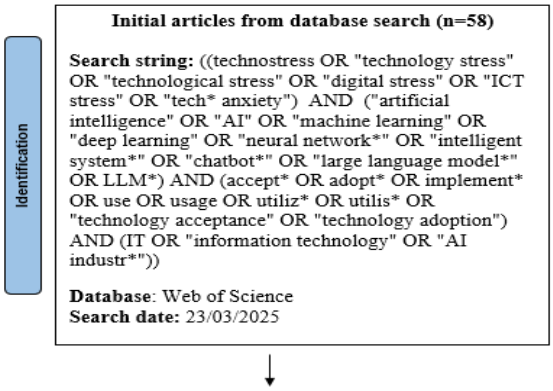
2 Methodology

The author followed a comprehensive, systematic three-step approach to identify and select relevant articles for the study. The entire methodology is explained in Fig. 1 as a PRISMA flowchart, which is widely used for bibliometric analysis (Puspitasari et al., 2025; Susitha et al., 2024). The first step of the process involved conducting a thorough search of the Web of Science (WoS) database using a specific search string tailored to the research topic (refer to Box 1, Fig. 1 for the exact search terms). This initial search identified 58 articles that were considered as potential papers for inclusion in the review.

The second step in the methodology involved filtering and refining the search results based on a set of predefined inclusion and exclusion criteria. The review period was strictly defined between 2022 and 2025, focusing specifically on articles and proceeding papers published within this timeframe. To ensure a relevant selection, the articles were later screened according to their categorization in various academic disciplines. The categories included Applied Psychology, Computer Science, and Business (see Box 2, Fig. 1 for a detailed breakdown of the categories). While disciplines like Religion and Linguistics could be relevant, this analysis focuses more on the different business-oriented sectors and psychological aspects of AI adoption through the lens of technostress. Other domains were either less directly related or would have broadened the study scope. This category-based filtering narrowed down the initial 58 papers to a more manageable set of 27 articles that passed the initial review and were deemed relevant for further analysis.

In the third and final step, the author manually reviewed the titles and abstracts of the remaining 27 articles to ensure that they closely align with the core focus of the study. The articles were assessed for their relevance to the specific aspects of technostress and AI adoption. Through this careful screening process, the number of papers was further reduced to approximately 19. Each of these 19 publications was rigorously evaluated to ensure it met the empirical criteria required for inclusion in the review, with particular attention paid to the practical and real-world insights provided by these studies. The final selection of papers formed the basis of the bibliometric analysis, contributing to a deeper understanding of the connection between technostress, AI adoption, and how its effects are influencing individuals across different industries.

While all 19 selected publications were included in the overall bibliometric analysis (e.g., keyword co-occurrence, research areas, and trends), the decision to highlight the top 10 most cited papers was driven by their influence and visibility within the academic discourse. These papers received the highest number of citations and were therefore considered representative of the most impactful contributions in the field. In such emerging domains, where the total body of literature is still limited, narrowing the sample to the most influential works allows for the identification of dominant themes, authorship patterns, and intellectual foundations (Rodríguez-Navarro & Brito, 2024; Zupic & Čater, 2015). Similar methodology has been used in other studies focused on early-stage or fragmented fields (Gutiérrez-Alvarez et al., 2025), where the aim was not to generalize widely but to provide a focused analytical lens on the most impactful scholarly contributions.



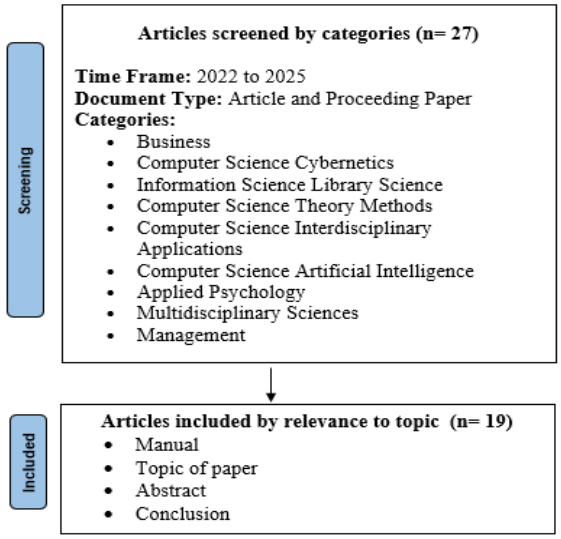


Figure 1. Flow diagram of search process (Source: Author's research according to Web of Science Core Collection database, as of March 23rd, 2025)

3 Results

This paper attempts to explore how technostress influences the uptake of AI across various industries. The years between 2022 and 2025 were explored since this was a pivotal time when AI, particularly generative AI, became widely accessible and transformed the way people engage with technologies (Kocoń et al., 2023).

A bibliometric analysis was conducted using data from the Web of Science Core Collection database. The following section of this chapter will analyze the outcomes of this research. According to the study's findings, 19 publications in the Web of Science core collection database were relevant to the bibliometric analysis. These publications serve as the foundation for assessing major research trends and influential authors.

Table 1 lists the top 10 most referenced papers from the 19 selected publications, emphasizing the most important research in this discipline. These studies offer significant insights into the relationship between technostress and AI adoption across different industries.

The analysis presented in Table 1 reveals that technostress related to AI adoption is a widespread phenomenon, observed across a range of sectors including information technology, tourism, financial services, healthcare, manufacturing, and remote work environments.

In technology-oriented industries such as IT, financial services, and e-commerce, technostress is primarily driven by the rapid pace of technological change and the constant pressure to adapt to complex AI systems, which may undermine employee satisfaction and productivity (Siitonen et al., 2025; Evenstad, 2024). In contrast, the tourism sector experiences technostress as staff must quickly master new AI-driven customer service tools, often amidst uncertainty regarding the effectiveness and reliability of these technologies. Kang et al. (2024) found that AI adoption increases job stress and turnover intentions among hotel employees in South Korea, especially towards other sectors. The healthcare industry faces unique stressors associated with integrating AI into sensitive areas, such as diagnostics and patient care, which raises concerns about system reliability, ethical responsibility, and patient safety (Issa et al., 2024). Manufacturing and service industries are facing workflow disruptions and job uncertainty due to AI-driven automation; however, involving employees in the design of ethical and transparent AI improves their well-being and performance, thereby reducing technostress (Verma et al., 2024). Furthermore, the rise of remote work introduces additional sources of technostress, such as techno-overload, blurred boundaries between work and private life (techno-invasion), and sustained pressure to remain productive through virtual means (Di Dalmazi et al., 2022).

The paper titled "Adoption of Artificial Intelligence (AI) based Employee Experience (EEX) Chatbots" by Pillai et al. (2024), published in Information Technology & People, is the most cited in this research, with 59 citations. It makes it the most influential study on the researched topic, highlighting the growing interest among both researchers and industry professionals in AI-driven solutions.

Additionally, the author was interested in identifying the most common journals publishing research on this topic. The analysis revealed that the journal Information Technology & People is the most frequently represented, with three of the top ten most cited papers published in this journal.

Table 1. The ten most cited papers in the WoS database on Technostress and the adoption of Artificial Intelligence

No.	PAPER	AUTHOR(S)	JOURNAL	YEAR	FIELD/INDUSTRY	WOS CITATIONS
1	Adoption of artificial intelligence (AI) based employee experience (EEX) chatbots	Pillai, R; Ghanghorkar, Y; Sivathanu, B; Algharabat, R; Rana, NP	Information Technology & People	2024	IT/ ITeS / E-commerce industries	59

2	What drives tourists' continuance intention to use ChatGPT for travel services?	Pham, HC; Duong, CD; Nguyen, GKH	Journal of Retailing and Consumer Services	2024	Tourism industry	47
3	Does AI-Driven Technostress Promote or Hinder Employees' Artificial Intelligence Adoption Intention?	Chang, PC; Zhang, WH; Cai, QH; Guo, HC	Psychology Research and Behavior Management	2024	General Workplace AI Adoption (information technology industry, financial services industry, manufacturing industry, etc.)	17
4	Unraveling the dark side of ChatGPT: a moderated mediation model of technology anxiety and technostress	Duong, CD; Ngo, TVN; Khuc, TA; Tran, NM; Nguyen, TPT	Information Technology & People	2024	Online survey (different fields and industries)	9
5	Enhancing value in customer journey by considering the (ad)option of artificial intelligence tools	Dhiman, N; Jamwal, M; Kumar, A	Journal of Business Research	2023	Marketing & Customer Experience	9
6	Employees' Adaptation to Technology Uncertainty in the Digital Era	Liu, NC; Wang, YC; Lin, YT	IEEE Transactions on Engineering Management	2023	Financial services industry	7
7	Navigating AI unpredictability: Exploring technostress in AI-powered healthcare systems	Issa, H; Jaber, J; Lakkis, H	Technological Forecasting and Social Change	2024	Healthcare industry	4
8	Acceptance of artificial intelligence devices in banking services	Cintamür, IG	International Journal of Bank Marketing	2024	Banking industry	2
9	Influence of Technostress on Work Engagement and Job Performance During Remote Working	Di Dalmazi, M; Mandolfo, M; Stringhini, C; Bettiga, D	Engineering Psychology and Cognitive Ergonomics	2022	Remote Work (IT, Sales/Business development, Banking, E-commerce)	2
10	Elevating employees' psychological responses and task performance through responsible artificial intelligence	Verma, S.; Singh, V.; Tudoran, A. A.; Bhattacharyya, S. S.	Information technology & people	2024	Service/ Manufacturing industries	1

Source: Author's research according to Clarivate Web of Science, as of March 23rd, 2025

Furthermore, all papers were published in English and authored by 62 researchers from 15 different countries, with China (4) and India (4) leading in terms of the number of publications. In total, there were 140 citations across these papers. The most significant increase in research on this phenomenon happened in 2024, as the first signs of stress caused by AI implementation started to appear (Kim & Lee, 2024).

Building on previously stated, AI (particularly generative AI) gained traction about two years after its initial widespread introduction (the launch of ChatGPT), so the sudden increase in published papers during 2024 that deal with technostress and AI was not unexpected. This increase in the number of publications and citations mentioned above is visually represented in Fig. 2 on the next page.

Moreover, based on the research areas shown in Figure 3 and considering the applied search query, the results indicate that the analyzed scientific papers are predominantly related to Business Economics (7 papers, 36.84%), Computer Science (6 papers, 31.58%), and Engineering (6 papers, 31.58%).

Additionally, a notable number of studies fall under Information Science and Library Science (4 papers, 21.05%). Other research areas with fewer publications include Psychology (2 papers, 10.53%), Psychiatry (1 paper, 5.26%), and Public Administration (1 paper, 5.26%). It is important to note that some articles are classified under multiple research domains, so the total number in Fig. 3 exceeds 19.

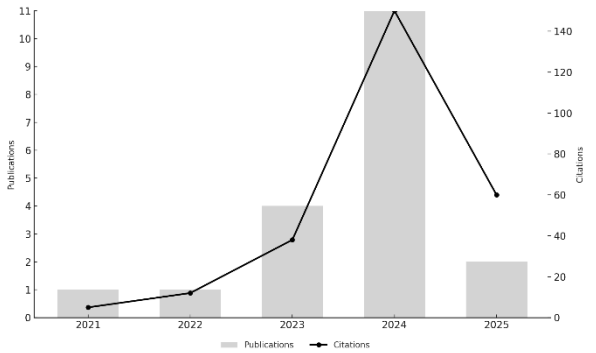


Figure 2. Distribution of the number of published papers and citations by year
(Source: Author's research according to Clarivate Web of Science, as of March 23rd, 2025)

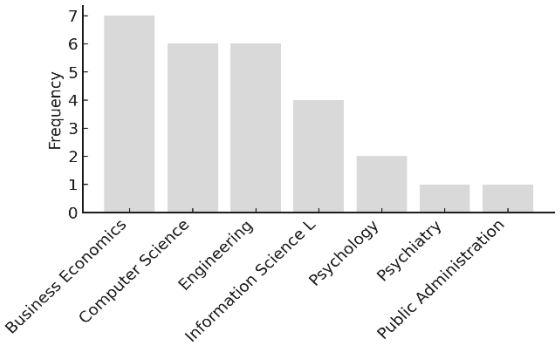


Figure 3. Number of published papers according to research area
(Source: Author's research according to Clarivate Web of Science, as of March 23rd, 2025)

An in-depth analysis of these 19 carefully selected papers related to the main topic is visually represented using the VOSviewer software. A total of 37 items (keywords) from the bibliometric analysis on the topic of technostress and its relationship with the adoption of AI were identified, each appearing more than two times in the selected papers from the WoS Core Collection database. These items were distributed into five distinct clusters, visually represented by different colors, forming a network of 240 links with an overall total link strength of 308 (see Fig. 4). The first cluster (red) includes keywords associated with technology anxiety, emotions, satisfaction, artificial intelligence, and loyalty, reflecting psychological and behavioral aspects of technostress and AI. The second cluster (green) is centered around technology, users, the internet, and anthropomorphism, highlighting the interaction between users and technological systems. The third cluster (blue) captures concepts related to performance, validation, psychology, and work, emphasizing the impact of technostress on workplace productivity.

The findings point out the importance of addressing these challenges to ensure successful AI integration in workplaces. The fourth cluster (yellow) groups keywords such as impact, technostress, stress, and organizations, illustrating factors that influence the adoption of new technologies. The fifth cluster (purple) focuses on engagement, acceptance, and computer anxiety, addressing occupational outcomes linked to technostress. Among these, the most frequently occurring keyword and the central concept in this analysis is "technostress", appearing 9 times, with 27 links and a total link strength of 51. This is followed by "impact" with 7 occurrences, 21 links, and a total link strength of 38. The keywords "satisfaction" and "technology" appear 6 times, while "performance" and "stress" each occur 5 times, forming critical thematic connections. Other frequently appearing keywords include "psychology," "organizations," "technology anxiety," "AI anxiety," and "user acceptance," all contributing to the broader discourse on the consequences of technostress.

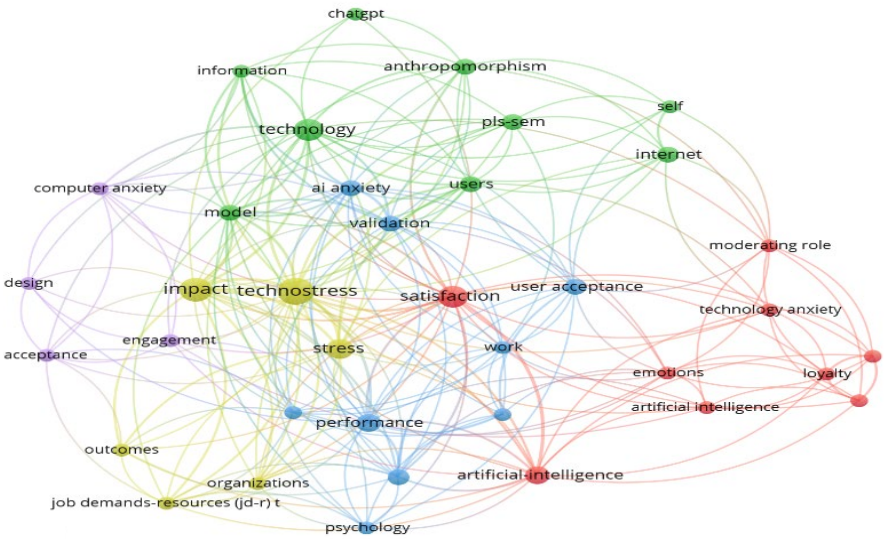


Figure 4. Visualization map of keywords co-occurrence (Source: Author's research according to Clarivate Web of Science, as of March 23rd, 2025)

4 Discussion

Based on the proposed RQ1, the research landscape on technostress and AI adoption has undergone a significant transformation between 2022 and 2025, as evidenced by a sharp increase in the number of relevant publications, particularly in 2024. This rise has occurred alongside the widespread adoption of AI tools (more specifically, generative AI tools), which introduced complex AI functionalities to the general workforce. The bibliometric analysis revealed that the research in this domain is still emerging, but gaining significant attention. The evolving narrative has transitioned from general concerns about digital transformation to more specific examinations of emotional and cognitive responses to AI-driven work environments. Notably, psychological constructs such as technological anxiety, burnout, and AI-related insecurity are now central themes in the literature (Bankins et al., 2023; Budhwar et al., 2022). The VOSviewer keyword mapping confirms this trend, with terms such as technostress, technology anxiety, job performance, and AI adoption forming dense clusters of research activity.

Continuing to the RQ2, the bibliometric findings show that research on technostress and AI adoption is highly interdisciplinary, with significant contributions from Business & Economics (36.84%), Computer Science (31.58%), and Engineering (31.58%). Emerging trends indicate a growing interest in context-specific studies across diverse sectors - ranging from IT and e-commerce to healthcare, tourism, and banking. The top-cited studies reflect this diversity and underscore the global scope of the problem. For example, the IT sector focuses on AI-based employee experience platforms (Pillai et al., 2024), while healthcare studies emphasize unpredictability and mental strain caused by AI systems (Issa et al., 2024). Remote work is another recurring theme, with studies addressing how AI adoption during the pandemic further intensified employee stress (Di Dalmazi et al., 2022).

Moreover, the keyword cluster analysis revealed five conceptual groupings, each pointing to a distinct thematic direction in the field. The clusters range from emotional and psychological responses to AI to user interaction patterns and organizational outcomes such as productivity, satisfaction, and acceptance. These clusters reflect the fragmented but expanding nature of technostress research and highlight how AI affects not only technological infrastructure but also human behavior and well-being.

The findings of the bibliometric analysis align closely with the dimensions of the Technostress Model proposed by Tarafdar et al. (2011). For example, the prevalence of keywords such as "complexity", "anxiety", and "AI unpredictability" across the top-cited studies reflects the stressors of techno-complexity and techno-uncertainty (Chang et al., 2024; Issa et al., 2024). Similarly, concerns over job displacement and the pressure to quickly adapt to AI tools correspond to techno-insecurity and techno-overload (Kim & Lee, 2024; Malik et al., 2021). These associations suggest that technostress is not only an emerging theme in AI literature but also a theoretical lens through which the challenges of AI adoption can be better understood.

5 Conclusion

Bibliometric analysis reveals that technostress manifests differently across sectors, shaped by the nature of work, digital maturity, and AI use-cases. For instance, in the IT and e-commerce sectors, technostress frequently arises from rapid digital transformation, constant technology updates, and increased pressure for digital performance. In healthcare, technostress is linked to cognitive overload and unpredictability introduced by complex AI systems, often raising unique concerns about safety, ethics, and decision-making. The financial and banking sectors report technostress mainly in the form of monitoring-related stress caused by automation and algorithmic management. In marketing, tourism, and customer service, technostress often manifests through increased customer expectations and the pressure to quickly adapt to AI-driven tools.

Technostress is not a uniform phenomenon but is shaped by industry-specific factors such as regulatory demands, job design, and the level of digital maturity. Organizations should therefore adopt tailored strategies to manage AI-related stress, aligned with their sector-specific conditions.

For instance, for HR professionals, this involves providing structured AI upskilling and digital literacy training, particularly in heavily regulated industries such as healthcare and finance, where technostress will more likely be in the form of cognitive overload and risk-avoidance-oriented. In fast-paced environments such as IT and services, well-being programs, burnout prevention measures, and accessible tech-support channels should be prioritized to minimize techno-overload and techno-invasion.

From the point of view of change management, employees need to be engaged early in AI deployment through participatory design, pilots, and feedback loops to minimize uncertainty and resistance. This is especially important in industries with low autonomy or those in which AI affects core decision-making (e.g., customer service, manufacturing). Leaders in digital transformation need to clearly explain the purpose, scope, and impact of AI initiatives, applying open messaging to diffuse techno-insecurity.

Finally, developers and system designers should integrate stress-reducing features into AI tools, such as user-friendly interfaces or explainable AI outputs, to preserve perceived control and autonomy. This is especially relevant for high-stakes applications (e.g., healthcare diagnostics). Aligning AI tool design with diverse user capabilities can reduce techno-complexity and increase acceptance.

6 Limitations and future work

While this study offers relevant insights into the link between technostress, AI adoption, and different industries, certain limitations should be noted. First, the sample was restricted to a relatively small body of literature, which may not reflect the full scope and diversity of research in the field. However, as AI technologies (particularly generative AI tools) continue to be increasingly integrated into business operations, research on technostress related to AI is expected to expand substantially. Second, contextual and moderating variables, including industry, organizational culture, employee characteristics, and regional differences, were not addressed. Empirical research exploring these factors, preferably using surveys or mixed methods, is needed for a more comprehensive understanding. Third and final, with the rapid progression of AI technologies, longitudinal studies are crucial to track how technostress and adaptive strategies evolve over time. Such research would directly inform the development of targeted interventions, training initiatives, and responsible AI governance in organizations. Addressing these gaps will provide a firmer theoretical and practical foundation for understanding the impact of AI-driven technostress on employees and organizations.

References

Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2023). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior*. https://doi.org/10.1002/job.2735

Budhwar, P., Malik, A., De Silva, M. T. T., & Thevisuthan, P. (2022). Artificial intelligence – challenges and opportunities for international HRM: a review and research agenda. *The International Journal of Human Resource Management*, *33*(6), 1065–1097. https://doi.org/10.1080/09585192.2022.2035161

Chang, P.-C., Zhang, W., Cai, Q., & Guo, H. (2024). Does AI-driven technostress promote or hinder employees' artificial intelligence adoption intention? A moderated mediation model of affective reactions and technical self-efficacy. *Psychological Research and Behavioral Management, 17*, 413–427. https://doi.org/10.2147/PRBM.S441444

Chen, Y., Hu, Y., Zhou, S., & Yang, S. (2023). Investigating the determinants of performance of artificial intelligence adoption in the hospitality industry during COVID-19. *International Journal of Contemporary Hospitality Management, 35*(8), 2868–2889. https://doi.org/10.1108/IJCHM-04-2022-0433

Cintamür, I. G. (2024). Acceptance of artificial intelligence devices in banking services: Moderation role of technology anxiety and risk aversion. *International Journal of Bank Marketing, 42*(7), 2143–2176. https://doi.org/10.1108/IJBM-10-2023-0563

Dhiman, N., Jamwal, M., & Kumar, A. (2023). Enhancing value in customer journey by considering the (ad)option of artificial intelligence tools. *Journal of Business Research, 167,* 114142. *https://doi.org/10.1016/j.jbusres.2023.114142*

Di Dalmazi, M., Mandolfo, M., Stringhini, C., & Bettiga, D. (2022). Influence of technostress on work engagement and job performance during remote working. *Engineering Psychology and Cognitive Ergonomics, EPCE 2022, 13307*, 149–163. https://doi.org/10.1007/978-3-031-06086-1_12

Duong, C. D., Ngo, T. V. N., Khuc, T. A., Tran, N. M., & Nguyen, T. P. T. (2024). Unraveling the dark side of ChatGPT: A moderated mediation model of technology anxiety and technostress. *Information Technology & People, 38*(4), 2015–2040. https://doi.org/10.1108/ITP-11-2023-1151

Evenstad, D. (2024, June). *Technostress in the Age of AI, Algorithmic Management, and Digital Surveillance*. LinkedIn. https://www.linkedin.com/pulse/technostress-modern-era-why-its-more-serious-than-ever-evenstad-dp1af/

Grummeck-Braamt, J.-V., Nastjuk, I., Najmaei, A., & Adam, M. (2021). A bibliometric review of technostress: Historical roots, evolution, and central publications of a growing research field. *Proceedings of the Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2021.796

Gutiérrez-Alvarez, O. A., Pantoja-Villa, L. A., & Calderón-Rojas, B. M. (2025). Bibliometric analysis of the publications that list the most-cited articles in endodontics. *Endodontology, 37*(2), 128–135. https://doi.org/10.4103/endo.endo_185_24

Issa, H., Jaber, J., & Lakkis, H. (2024). Navigating AI unpredictability: Exploring technostress in AI-powered healthcare systems. *Technological Forecasting and Social Change, 202,* 123311. *https://doi.org/10.1016/j.techfore.2024.123311*

Kang, J., Shin, H., & Kang, C. (2024). Hospitality labor leakage and dynamic turnover behaviors in the age of artificial intelligence and robotics. *Journal of Hospitality and Tourism Technology*, *15*(5), 916–933. https://doi.org/10.1108/JHTT-12-2023-0411

Kim, B. J., & Lee, J. (2024). The mental health implications of artificial intelligence adoption: The crucial role of self-efficacy. *Humanities and Social Sciences Communications, 11*, 1561. https://doi.org/10.1057/s41599-024-04018-w

Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., Bielaniewicz, J., Gruza, M., Janz, A., Kanclerz, K., Kocoń, A., Koptyra, B., Mieleszczenko-Kowszewicz, W., Miłkowski, P., Oleksy, M., Piasecki, M., Radliński, Ł., Wojtasik, K., Woźniak, S., & Kazienko, P. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion, 99*, 101861. https://doi.org/10.1016/j.inffus.2023.101861

Kumar, A., Krishnamoorthy, B., & Bhattacharyya, S. (2023). Machine learning and artificial intelligence-induced technostress in organizations: A study on automation-augmentation paradox with socio-technical systems as coping mechanisms. *The International Journal of Organizational Analysis, 32*(4), 681–701. https://doi.org/10.1108/IJOA-01-2023-3581

Liu, N. C., Wang, Y. C., & Lin, Y. T. (2023). Employees' adaptation to technology uncertainty in the digital era: An exploration through the lens of job demands-resources theory. *IEEE Transactions on Engineering Management, 71,* 7286–7297. *https://doi.org/10.1109/TEM.2023.3264293*

Makridis, C. A., & Mishra, S. (2022). Artificial intelligence as a service, economic growth, and well-being. *Journal of Service Research, 25*(4), 505–520. https://doi.org/10.1177/10946705221120218

Malik, N., Tripathi, S. N., Kar, A. K., & Gupta, S. (2021). Impact of artificial intelligence on employees working in industry 4.0 led organizations. *International Journal of Manpower*, *43*(2), 334-354. https://doi.org/10.1108/IJM-03-2021-0173

Pereira, V., Hadjielias, E., Christofi, M., & Vrontis, D. (2023). A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective. *Human Resource Management Review, 33*(1), 100857. https://doi.org/10.1016/j.hrmr.2021.100857

Pham, H. C., Duong, C. D., & Nguyen, G. K. H. (2024). What drives tourists' continuance intention to use ChatGPT for travel services? A stimulus-organism-response perspective. *Journal of Retailing and Consumer Services, 78,* 103758. https://doi.org/10.1016/j.jretconser.2024.103758

Pillai, R., Ghanghorkar, Y., Sivathanu, B., Algharabat, R., & Rana, N. P. (2024). Adoption of artificial intelligence (AI)-based employee experience (EEX) chatbots. *Information Technology & People, 37*(1), 449–478. https://doi.org/10.1108/ITP-04-2022-0287

Puspitasari, E., Yuen, C. W., & Ibrahim, M. R. (2025). The research landscape of first- and last-mile in public transport systems: A bibliometric analysis. *KSCE Journal of Civil Engineering, 29*(8), 100159. https://doi.org/10.1016/j.kscej.2025.100159

Rodríguez-Navarro, A., & Brito, R. (2024). Rank analysis of most cited publications: A new approach for research assessments. *Journal of Informetrics, 18*(2), 101503. https://doi.org/10.1016/j.joi.2024.101503

Siitonen, V., Ritonummi, S., Salo, M., Pirkkalainen, H., & Mauno, S. (2025). Coping with technostress in the software industry: Coping strategies and factors underlying their selection. *Journal of Systems and Software, 225*, 112341. https://doi.org/10.1016/j.jss.2025.112341

Susitha, E., Jayarathna, A., & Herath, H. M. R. P. (2024). Supply chain competitiveness through agility and digital technology: A bibliometric analysis. *Supply Chain Analytics, 7*, 100073. https://doi.org/10.1016/j.sca.2024.100073

Tarafdar, M., Tu, Q., Ragu-Nathan, B. S., & Ragu-Nathan, T. S. (2007). The impact of technostress on role stress and productivity. *Journal of Management Information Systems, 24*(1), 301–328. https://doi.org/10.2753/MIS0742-1222240109

Tarafdar, M., Tu, Q., Ragu-Nathan, T. S., & Ragu-Nathan, B. S. (2011). Crossing to the dark side: Examining creators, outcomes, and inhibitors of technostress. *Communications of the ACM, 54*(9), 113–120. https://doi.org/10.1145/1995376.1995403

Uren, V., & Edwards, J. S. (2023). Technology readiness and the organizational journey towards AI adoption: An empirical study. *International Journal of Information Management, 68*, 102588. https://doi.org/10.1016/j.ijinfomgt.2022.102588

Verma, S., Singh, V., & Bhattacharyya, S. S. (2024). Elevating employees' psychological responses and task performance through responsible artificial intelligence. *Information Technology & People, 37*(7), 2551–2567. https://doi.org/10.1108/ITP-05-2023-0431

Zirar, A., Ali, S. I., & Islam, N. (2023). Worker and workplace artificial intelligence (AI) coexistence: Emerging themes and research agenda. *Technovation, 124*, 102747. https://doi.org/10.1016/j.technovation.2023.102747

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429–472. https://doi.org/10.1177/1094428114562629

Planning and Development Processes for the Information System Based on the Command Query Responsibility Segregation Design Pattern

Krešimir Kavran
InfoDom Ltd.
Andrije Žaje 61, 10000 Zagreb, Croatia
kresimir.kavran@infodom.hr

Maja Pušnik
University of Maribor
Faculty of Electrical Engineering and Computer Science
Koroška cesta 46, 2000 Maribor, Slovenia
maja.pusnik@um.si

Abstract. *The development of a complex information system requires planning and the definition of all business requirements. Business requirements can change over time. Agile methodology enables development teams to adapt to real-time changes and the development of major features according to customer requests. One such complex information system is a platform for the electronic operations of state administration bodies. It contains multiple modules based on the Command Query Responsibility Segregation design pattern. A Croatian software development company created it for the Croatian state administration bodies. This paper describes the development process of creating this information system and explains the maintenance process.*

Keywords. Command Query Responsibility Segregation, Agile, Scrum, Project planning

1 Introduction

When creating a new product or improving an existing one, developers and designers must consider the end-users, their needs, and how they want to use the product. End users care about completing their tasks and achieving their goals, not about how the product was created or the company that created it (Saffer, 2010).

This also affects internal processes in software development companies. Employees must adjust to customer demands and requirements. To address this obstacle, software development companies use different methodologies to plan further development and realize and maintain such plans.

Agile methods, such as Scrum, are currently the most commonly used methodologies in software development (Omonije, 2024). Scrum, in particular, is a framework for developing complex projects and organizing work based on values, principles, and practices that provide a foundation for all members. The Scrum framework offers excellent flexibility and versatility when facing changing constraints (Andrei et al., 2019).

This paper aims to describe the planning and development processes in a Croatian Software development company whose product is an information system based on the Command Query Responsibility Segregation (CQRS) design pattern.

This paper is structured as follows. Section 2 provides background information and related work about Scrum and Command Query Responsibility Segregation design pattern. Section 3 provides information about the information system and its components. Section 4 contains information about the company's planning and development processes. Section 5 includes a discussion. Section 6 concludes the paper.

2 Background and Related Work

This section presents general information about the Command Query Responsibility Segregation design pattern and Scrum. It consists of two parts. The first part describes the Command Query Responsibility Segregation (CQRS) design pattern, including its advantages and disadvantages, while the second provides information about Scrum.

2.1 Command Query Responsibility Segregation

The primary objective of Command Query Responsibility Separation is to distinguish between actions that modify data in the database (referred to as commands) and requests that retrieve data from the database without altering it (referred to as queries) (Overeem et al., 2021).

The benefits of this approach are best observed in distributed environments. Because commands and

queries are separated, it is possible to use different models or define different execution paths for each type. The separation of commands and queries makes each component easier to develop because components do not depend on each other (Kufner & Marik, 2019).

The separation of read and write operations allows the choice of different databases for read and write operations. Developers can select the most performant alternative for queries without sacrificing the benefits of the original (relational) database for state mutation operations. Furthermore, each query can be optimized separately by simultaneously maintaining several different read models. These benefits come with a cost associated with the synchronization of multiple data models and the underlying storage (Debski et al., 2017).

This separation makes each component easier to develop and maintain. While the overall architecture may appear more complex, the individual components become simpler and more manageable, resulting in easier application development (Kufner & Marik, 2019).

Research papers on CQRS are limited in number. The reason for that is the area of application. The CQRS design pattern is intended for use in high-load distributed environments, and as such, it is interesting only to a limited number of solution architects and developers. All published research papers include descriptions of CQRS and additional techniques used in combination with it.

All research papers on CQRS-based information systems include a basic description of the CQRS design pattern and descriptions of other techniques used.

Published research papers describe various elements, implementations, advantages, and disadvantages of CQRS.

The authors (Diakov et al., 2019) compare the CQRS design pattern with traditional CRUD systems, and (Weerakoon & Kumara, 2018) compare CQRS with monolithic architecture, focusing on eventual consistency and its effects.

Research papers (Debski et al., 2017) and (Overeem et al., 2017) focus on event sourcing and describe the benefits and challenges of using event sourcing in CQRS applications. While (Lima et al., 2021) propose improvements to event sourcing in CQRS-based applications.

The authors (Erb & Kargl, 2014) describe how to use the CQRS pattern in event-driven architectures, while (Kufner & Marik, 2019) explain how to implement the CQRS design pattern in REST APIs.

The research paper (Kabbedijk et al., 2012) introduces techniques to enhance the CQRS design pattern, such as aggregating roots and snapshotting to increase variability.

Finally, (Rajković et al., 2013) present improvements in the performance of existing information systems by using two databases, each for read and write operations, along with a method for data synchronization between them.

2.2 Scrum

In 2001, seventeen software development professionals published the Agile Manifesto. The Agile Manifesto focuses development on four core values. The Agile Manifesto also contains twelve additional principles to ensure user satisfaction by continuously delivering valuable, simple, high-quality software (Itzik & Roy, 2023).

In Agile, teams are self-organized, allowing teams to adjust to any problem and choose the best approach for solving problems. The definition of roles in agile software development depends on the development method and the organization's size (Bomström et al., 2023).

Scrum is a management and control process that cuts through complexity to focus on building software that meets business needs (Schwaber & Beedle, 2002).

The word "Scrum" originated from the sport of rugby, where a Scrum is a formation that enables quick adoption of strategies by team members, with every player playing a specific role (Zayat & Senvar, 2020).

Key elements of Scrum are:

- *Product Backlog* is a prioritized list of all product requirements. It describes everything the system should include regarding functionality, features, and technology.
- *Product Owner* is responsible for prioritization of the backlog. The Product Owner decides the order in which things are built (Schwaber & Beedle, 2002).
- *Scrum Teams* are small, cross-functional teams that perform all development. They are also referred to as development teams.
- *Sprints* are development iterations lasting between one and four weeks. Every Sprint must finish with an increment of the system and new product functionality.
- *Sprint Backlog* is a collection of items taken at the beginning of a Sprint by the development team that can be completed during the Sprint and turned into an increment of the system.
- *Scrum Master* is a management representative who enforces Scrum practices and helps the development team to make decisions or acquire resources as needed (Schwaber & Beedle, 2002).
- *Daily Scrum* is a daily meeting held by the Scrum Team to review progress and identify problems.
- *Sprint Review Meeting* is a meeting between the Scrum Team and the management held at the end of every Sprint. It is used to inspect the product increment that has been built.

3 The Information System

The chosen information system is used for the electronic operations of state administration bodies. It is intended for state administration bodies, local and regional self-government units, public authority bodies, and all enterprises that provide public services to citizens and businesses. Croatian software company created and maintains it, and Croatian state administration bodies use it.

The information system is a modular microservice-based information system, with its core components built around the CQRS design pattern. It is composed of many modules. The most used modules of the information system are:

- primary web interface,
- secondary web interface,
- modules for integration with government services,
- plugin for e-mail clients,
- integration service.

Regardless of the modules used, all modules employ the same core components, ensuring their basic functioning is consistent, and they all utilize the CQRS design pattern.

The primary user interface is intended for employees of state administration bodies. This web interface is what most end users use in everyday business. It is the most complex module and offers access to all functionalities of the information system. It utilizes role-based access control, allowing employees to access specific functionalities based on their assigned roles.

The secondary web user interface is intended for citizens and businesses to use. Its goal is to enable more accessible communication with state administration bodies, simplify communication, facilitate the submission of requests to state administration bodies, and allow citizens to monitor the progress of their inquiries. A secondary user interface is not obligatory for the correct work of the information system.

Modules for integration with government services enable the retrieval of data, such as information about citizens or companies, from government services. Using these services, the information system can always contain up-to-date information without requiring employees to manually update the records. Each government service has a separate module in the information system. These modules are not standalone solutions and must be incorporated with other modules.

An elective option for the information system is a plugin for e-mail clients, which enables inputting data directly from the e-mail client without using the primary web interface. A plugin for e-mail clients was created to allow employees to input citizens' e-mails, contact information, and data into the information system and automatically generate and send confirmation e-mails.

Depending on the employee's work position, a primary web interface may not be needed for the employee to complete tasks. Some users are tasked with communicating with citizens, and by using this plugin, they can complete their tasks without needing to use the primary web interface, thereby reducing the load on the web interface.

Integration service is a module that enables the exchange of data with other information systems. If a customer has another application or information system from a different manufacturer, they can utilize this integration service to establish a connection between the two solutions, thereby eliminating potential data duplication across multiple information systems. Integration service does not offer all the functionalities of the primary web interface. It provides most read-and-write operations. Operations that are not supported include those that could affect the entire information system's functionality, such as configuration properties and limitations imposed by legal regulations.

The information system was developed as a modular microservice system. Fig. 1 shows the simplified architecture of the information system for end users (the central part of the information system).

All end users' requests, regardless of how they are made, are sent to the Representational State Transfer Application Programming Interface (REST API) for data manipulation (data retrieval, inserting new data, or updating existing data).

Web user interfaces use separate controllers from other modules. Most controllers in the REST API are generated at startup, depending on the available queries and commands. This simplifies adding new queries and commands because no configuration is needed to use them. Some specific controllers are not generated but are written by developers to enable the execution of particular tasks. Due to this level of customization, only web user interfaces created by the company can utilize them.

E-mail client plugin and other modules use separate controllers. These controllers are written by developers and are not generated at the startup. These controllers form an integration service that enables other information systems to exchange data with this information system.

This means that only actions permitted by the company are allowed for use with other information systems. Because other information system modules perform specific tasks, unlike web user interfaces, separate controllers were made. These modules do not require all available operations, such as web user interfaces.

Controllers are responsible for accepting user requests and determining whether the request should be forwarded to other components or if the controller can process the request itself without needing to query the database and send the response to the web interface.

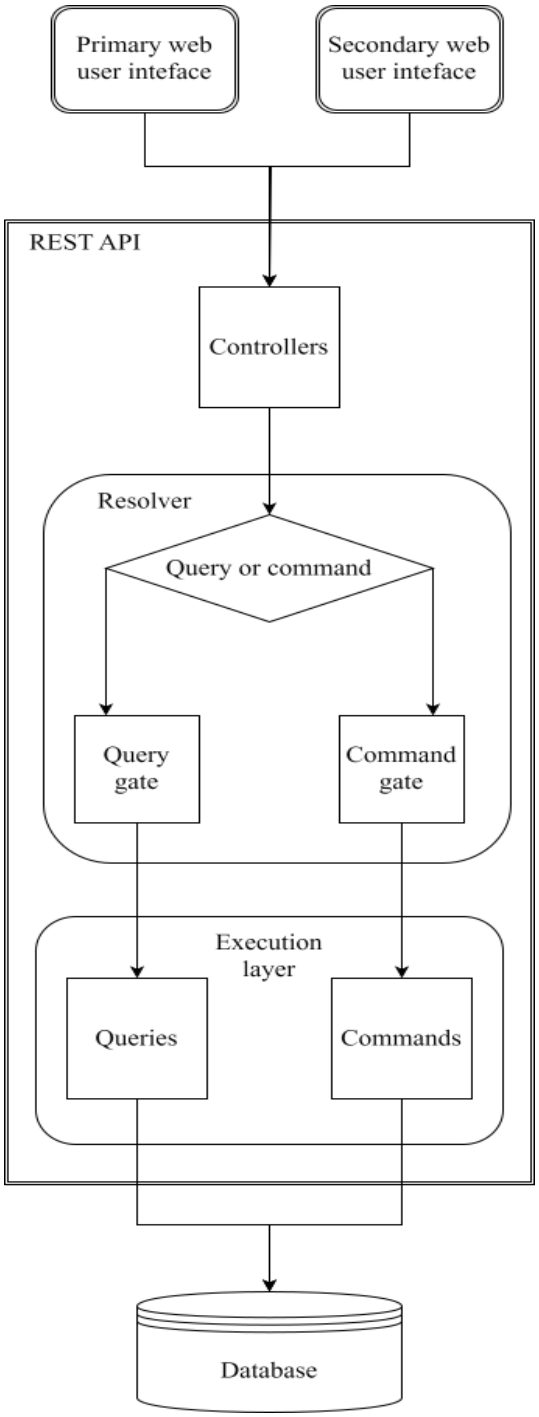


Figure 1. Information system architecture overview

If access to the database is needed, whether for a query or command, the request is forwarded to the resolver. The resolver determines whether a query or command needs to be executed to process the request. The request is sent to the query gate if it is a query. Otherwise, it is sent to the command gate.

Each gate executes a query or command in the execution layer. Depending on what component was called, the result can be a data set (the result of a query) or information on whether modifications to the data were successful (the result of a command). The list of queries and commands is not static. It is automatically generated at the start time, and no configuration is needed to add a new query or command. This enables a quick and easy way to add new queries and commands or modify existing ones.

The execution layer executes queries and commands on the given database. It can be configured to use one or more databases. The default configuration utilizes a single database, but it also supports multiple databases. For example, it can be configured to use two databases. In that case, one database can be used for queries (read) and a second for commands (write). In the case of multiple databases, it is necessary to ensure data synchronization between databases.

4 Planning and Development Processes

The company uses the Agile method Scrum for planning and development. To simplify the entire process, the company uses the tool Azure DevOps. This section contains descriptions of Azure DevOps and planning and development processes.

4.1 DevOps

The acronym DevOps comes from the concept of development and operations. These are related to the creation of software products. Its challenges include the adequacy of applying techniques and software development processes that align with all necessary operations to deliver a quality product. In this sense, a relationship between quality, costs, operations, time, and activities is essential and must be related to the participation of all software development team members (Bento et al., 2023).

DevOps is a set of practices that integrate software development and IT operation teams by working together across the entire software development life cycle (Dileepkumar & Mathew, 2021).

Azure DevOps is a professional tool used by companies specializing in software development (Bento et al., 2023). It is widespread worldwide and has a large amount of documentation and working methods. Azure DevOps enables organizations to manage their end-to-end software development and deployment processes effectively and efficiently (Dileepkumar & Mathew, 2021).

Azure DevOps is composed of several key components that form its backbone. These include

- Azure Repos, which manages source code repositories, Azure Pipelines for continuous integration and delivery,
- Azure Boards for agile project management,
- Azure Artifacts for package management.

These components work together to streamline the software development process, thereby enhancing team collaboration and productivity (Borra, 2024).

Azure Boards is a versatile and adaptable tool for organizing and overseeing work across teams. It facilitates agile project management functions, including backlog management, sprint planning, task tracking, and visual boards. With customizable features such as work item configuration, burndown charts, and seamless integration with other Azure DevOps services, Azure Boards ensures that teams maintain visibility and transparency, which are crucial for alignment and productivity (Borra, 2024).

Azure Repos provides Git repositories designed for efficient code version control, enabling seamless collaboration among teams. Regardless of project scope, teams benefit from features such as branching, pull requests, code reviews, and branch policies, ensuring code quality, consistency, and collaboration (Borra, 2024).

Azure Pipelines emerges as a robust continuous integration and continuous delivery (CI/CD) service, automating the build, test, and deployment processes. It accommodates various programming languages, platforms, and deployment targets, seamlessly empowering teams to deploy applications to any cloud or on-premises environment. With features such as multi-stage pipelines, parallel execution, and integration with third-party tools, Azure Pipelines enables teams to deliver software updates confidently and efficiently (Borra, 2024).

Azure DevOps is used in the selected company for planning, development, building new versions, testing, and deployment. The project is based on .NET technology, and the logical choice was to use the most compatible tool.

4.2 Planning process

The chosen information system is designed to support state administration bodies in their operations. For this reason, multiple sources influence the planning process and future development. These sources offer ideas or requests for improving the information system and defining new functionalities. The source can be an employee, a customer, a legal regulation, or a third party (for integration services). This means that the need for new functionalities rises during the year, and because the company uses Scrum, that is not a problem.

Azure Board enables the definition of new functionalities and agile ways of work management. The company uses Azure Board to log incoming requests or ideas as work items regardless of the source. Also, it enables tracking completed work.

For this reason and planning, it is necessary to group all related functionalities. As mentioned before, the company uses Azure. The Azure Board enforces grouping all work items into groups called Epics and Features.

The company uses group Epic to group work items per project or, in the case of complex projects, into multiple Epic groups. In the case of the chosen information system, the company uses the following Epic groups:

- one epic group for main functionalities and minor modules of the information system,
- one epic group for architectural improvements,
- three epic groups for the three most significant modules and improvements to the information system,
- one epic group for deployment and implementation of solutions in the customer's environments for core components and all modules except one module,
- one epic group for deployment and implementation of the most complex module.

These groups were created to simplify the process and enable easier tracking of completed work, rather than placing everything in one group. Creating a new epic group is allowed only for senior employees and only for legitimate business reasons.

Some of the properties for epic groups are group name, description, start date, and target (end) date. Because this is a long-term project, these epic groups do not have a defined target date. The information system is constantly updated with new functionalities and improvements.

Feature groups are part of an epic group, representing specific information system features. A feature group is a core component, a module component, an improvement requested by the customer, and other information system features. When this paper was written, there were 141 feature groups distributed in seven epic groups.

Feature groups have the same properties as epic groups, like group name, description, start date, and target (end) date. All feature groups have a target date except those used for core components.

Azure Boards (a part of Azure DevOps) offers a graphical representation called a feature timeline of all feature groups for easier planning and prioritizing work. It uses a Gantt chart for a graphical representation of feature groups.

Feature groups contain User Stories and Bugs. A User Story is a work item that includes a description of a new feature or improvement of an existing feature of the information system. A Bug describes a found or reported bug in the information system. User Stories and Bugs must include all relevant information for developers to create new features, improve existing features, or fix bugs in the information system. They represent assignments for the developers.

Each user story and bug must contain a Task. A task represents specific actions that a developer must do to fulfill the assignment. A user story and a bug must contain at least one task.

User stories and bugs can contain issues besides the tasks. An issue is added later in the development if an error is found in the solution during the testing, or if

something is not done as described in the user story or bug. The issue can be added to the user story or bug only if the user story or bug is still in the sprint backlog. A new Bug must be created if the original user story or bug is not in the sprint backlog.

In the company, only senior employees can create epic and feature items (groups). Everyone can create user stories, bugs, tasks, and issues, regardless of their position or role.

All employees are divided into two groups. These groups are:

- business group (business analysts, project managers, product owner, implementers, testers),
- technical group (developers, solution architects, project technical managers, scrum master).

Business group members are responsible for communicating with the customers, configuring the features of the information system, and testing new versions. The business group does not have permission to modify the source code.

Technical group members are responsible for developing new features or improving the existing features of the information system. The technical group has permissions to access and modify the source code.

The information system is not available to everyone. It can only be obtained by directly contacting the company and is available only for specific clients. The information system is designed for use by state administration bodies and state agencies. When buying the information system, a contract is made between the customer and the company. The contract lists all details about the installation, maintenance, and deployment of new versions. A dedicated contract is made if a customer wants a specific new functionality.

The company assigns a project manager and a project technical manager to every customer. A project manager is tasked with communicating with the customer's representative about the functionalities of the information system and the deployment of new versions (if deployment and maintenance are listed in the contract).

The project technical manager is responsible for assisting the project manager with technical issues and, if necessary, communicating with the customer's representative about maintenance and technical issues (if that is listed in the contract).

Business group members are responsible for creating new work items for the developers. Depending on the source of the work item, a work item can have a target date or not. If the source for the work item is a customer, there is often a target date.

For new functionalities requested by the customer, a new dedicated contract with the customer must be created and signed. If the customer insists on a specific deployment date for the new version with the new functionality, and a target date is specified in the contract, new work items will also have a target date. Before the target date, a new version must be deployed. On the target date, a new version must be given to the customer for use.

There is no need for a separate contract if the customer reports a bug. If it is a critical bug, a solution will be deployed as soon as possible. If it is a minor bug, the solution will be deployed on the scheduled date (if continuous deployment of the new version is specified in the contract).

The deployment of new versions varies per customer. Some customers prefer monthly deployments, while others opt for a new version deployment once a year.

Business group members must decide if existing feature groups can be used for new work items when planning future work. If the existing feature group fits the purpose for new work items, the employee creates new user stories and bugs inside that feature group.

If existing feature groups are not suited, senior members must create a new feature group and decide on the group name and its properties. One of the important properties is the target date because this property influences the feature timeline and deployments. Also, it influences the prioritization of work items.

Depending on the complexity of the new functionality, an employee creates one or more user stories. A user story must be completed within a sprint. If that requirement can not be fulfilled, multiple user stories are made.

Feature groups that do not have a target date are dedicated to user stories for functionalities that are nice to have but are not requested by the customer, or for functionalities for which a release date has not been decided. Also, feature groups in the epic item from architectural improvements do not have a target date. These groups are dedicated to technical group members and user stories with tasks for optimizing or improving core components that affect the entire information system.

Business group members have meetings every two weeks to prioritize work items (user stories and bugs). These meetings occur before sprint planning for the upcoming sprint. The prioritization is done according to the feature timeline (current contracts) and by the significance to the information system.

When a work item (user story or bug) enters a sprint, responsible business group members must be available to the developers for additional information regarding the work item (if necessary).

After a new version of the information system is built and deployed on the company servers, testers conduct tests on the information system. If an issue is found and a user story or bug is still present in the sprint, the tester creates an issue and notifies the responsible developer, whose task is to resolve the issue as soon as possible. If a tester finds an issue whose related user story or bug is not in the sprint, he/she must create a new work item (bug).

If a new version does not have critical bugs, it becomes a release candidate. The project manager's task is to decide if the new version will be deployed on

the customer's servers. This also depends on the signed contract and deployment plan listed in the contract.

Project managers must contact customer representatives if a new version needs to be deployed. After agreeing on the time and date of the deployment, project managers must communicate the date and time to the project technical manager. The project technical manager is responsible for the deployment of the new version. If the project technical manager is unavailable to perform the deployment, senior members of the technical group must decide who will perform the deployment instead.

Deployments can be automated or manual. Deployments on the company test servers are automated. The deployment method varies per customer due to infrastructure, business, and security restrictions. An example of manual deployment is the use of an information system within the customer's local network, without internet access.

4.3 Development process

The technical group is responsible for development. The group consists of developers and solution architects, some of whom hold the role of project technical manager. The group is also referred to as the development team.

At the time of writing this paper, the technical group consisted of

- three junior .NET developers,
- three mid .NET developers,
- two senior .NET developers,
- three SQL developers,
- two solution architects,
- scrum master/solution architect.

Developers are responsible for developing new features, improving existing features, and fixing bugs. Solution architects are responsible for the architecture of the entire information system and deciding on the best approach for implementing new features or improving core components. The development team has three solution architects, but not all three are in charge simultaneously. One solution architect, who also serves as a Scrum Master, is the lead solution architect. The remaining two solution architects support the lead solution architect, and each is responsible for a different area. One is dedicated to databases, while the other is dedicated to programming. In the absence of a lead solution architect, one of the remaining two solution architects assumes the role of lead solution architect.

Technical group members are shielded from outside influence and are a self-organizing team. The exceptions are project technical managers who only communicate with the customer representative if necessary.

The development teams' work is organized into sprints. One sprint lasts two weeks. The exception is the month of August, when the sprint lasts the entire month because most developers are on vacation. In August, the development team works at half capacity due to vacations and the shift of developers returning from vacation and those just going on vacation.

Before every sprint, the development team organizes a meeting called sprint planning. That meeting is held a day or a couple of days before the start of a new sprint, depending on the available time of the development team members. All development team members are required to participate in the meeting. Also, the product owner is obligated to be present at the meeting. Sprint planning is led by the lead solution architect or, in the case of absence, one of the two remaining solution architects.

At the beginning of sprint planning, all unfinished items from the previous sprint are transferred to the new sprint. After that, new items from the product backlog are added to the new sprint according to the priority established by the business group members. New items are added until 70% of sprint capacity is achieved.

In Azure DevOps, every developer has an assigned daily capacity from which a total capacity for the sprint is calculated. Developers' daily capacity depends on their knowledge and experience. It varies from three hours (for junior developers) to six hours (for senior developers). The remaining working hours are allocated for breaks, code reviews, learning, and knowledge transfer.

The 70% limit was chosen based on previous experiences. During the sprint, unforeseen circumstances could result in the absence of one or more developers or solution architects. The reason for absence could be illness, family problems, or specific emergency problems. Also, a critical bug could be found during the sprint, which requires immediate action. These bugs are added later during the sprint in the sprint backlog and also affect the team's capacity.

The installation of the information system on the servers of the new customer or the installation of new modules on the servers of existing customers are user stories (work items) that can be added to the sprint backlog during sprint planning. Reasons for this are contracts that specify fixed dates for installation and delivery, and because the initial installation or installation of additional modules requires the installation of support software on customer servers.

Deployments of new versions are not user stories (work items) because the deployment process is simplified, and the date for deployment is agreed upon when a new stable version is created (unless the date is also listed in the contract or if a version contains a feature request by the customer). For deployment of upgrades, it is unnecessary to install any additional support software (it is already installed on the servers). This is also one of the reasons for the limit of 70% capacity.

Work items are not immediately assigned to each developer during sprint planning. Instead, each

developer takes over a new work item once the previous one is completed. Junior developers can choose which work item they want to work on next. This is permitted due to their limited experience and knowledge. Junior developers take tasks they already know how to complete or tasks for learning purposes. All other developers and solution architects must take work items according to priority. The exceptions are sensitive tasks (modification of core components or task complexity). These work items are immediately assigned to the senior developer or solution architect.

A daily meeting is held during the sprint at the beginning of each work day. Members of both groups are present at the daily meetings. At these meetings, technical group members must report on their activities from the previous working day and announce their tasks for the current day. At the end of the meeting, business group members are allowed to announce any new information they have obtained during the previous day (if necessary for the development). Daily meetings can last between five and twenty minutes, depending on the number of present technical group members.

The development team holds weekly meetings called development backlog meetings, during which technical topics are discussed. The topics of these meetings include the presentation of changes to core components, discussion on the best way to implement a specific new feature, assisting another developer in solving a particular problem they encountered during development, and other related topics. The duration of this meeting is approximately one hour. Development backlog meetings are not always held every week. It depends on the number of available developer team members and if there is a suitable topic for the meeting.

A sprint review and retrospective are held at the end of the sprint. Members of both groups are present at these meetings. The purpose of a sprint review and retrospective is to present new features or significant improvements to existing features (if any have been introduced) to all employees working on the project, and to discuss problems encountered during the sprint and possible solutions to improve future sprints.

Additionally, during these meetings, a discussion on metrics is held. The most used metrics are velocity, sprint progress, burndown chart, pipeline pass rate, and deployment pass rate. Metric results are discussed at the meeting to determine possible problems and bottlenecks. Especially if a negative trend is identified, this discussion enables every member to express their opinion and fosters open communication. It also creates an opportunity for every member to propose a solution for improving planning and development processes.

5 Discussion

The previous section describes planning and development processes separately, but it is necessary to view these processes as a whole. Members of both groups must work together to improve and maintain the information system. Fig. 2 illustrates a simplified flow diagram for both processes, along with the corresponding links between them.

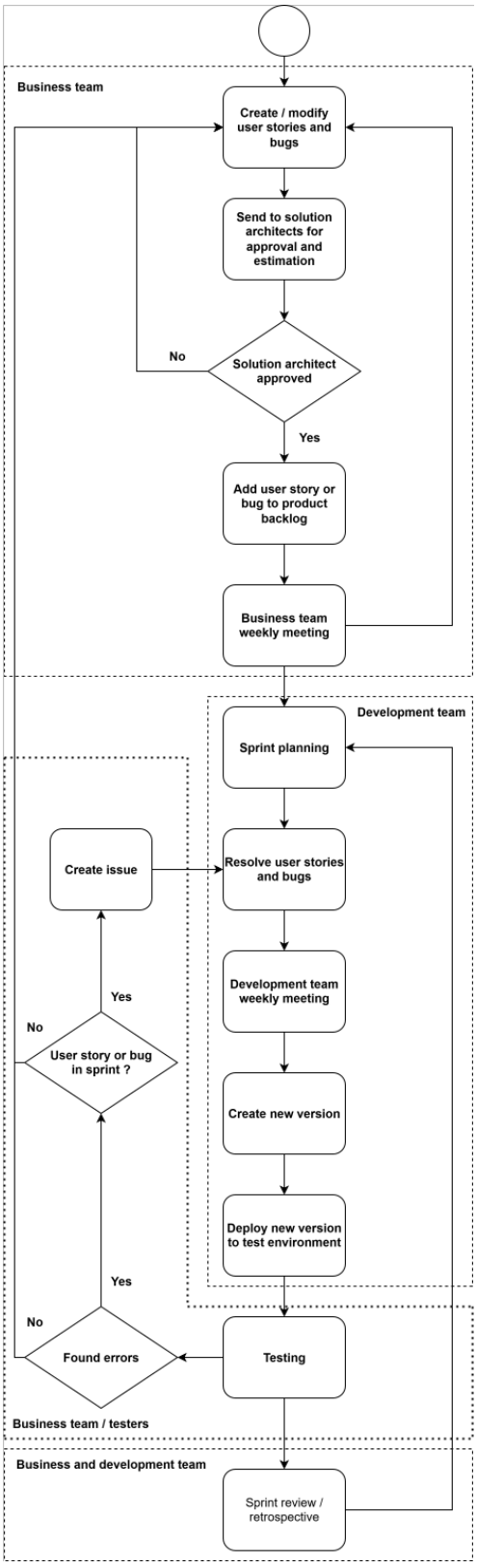


Figure 2. Planning and development processes

The company's solution architects, project managers, and product owner cooperate when defining new work items (user stories and bugs). After the business team member creates a work item, the solution architect reads what was written and leaves a comment. These comments are helpful because they enable the business team member to correct any mistakes or provide additional information before a work item is sent to the developer. It also helps developers to better understand what is asked of them.

The entire process begins with a discussion between business team members about new functionalities or improvements, regardless of whether the customer has requested it or the idea originated from an employee. Anyone can propose ideas for improving the information system. Sometimes, changes to the information system are necessary due to updates in legal regulations.

After creating a work item, if the employee is satisfied with the content, it is sent to the solution architects for review. The original author of the work item becomes the owner of the user story or bug. If the work item is well-written and contains all the necessary information, a solution architect assesses the effort required for the developer to complete all tasks associated with that work item.

If the solution architect is not satisfied with the description in the work item or if relevant information is missing, the solution architect leaves a comment and asks for an amendment or makes modifications to the work item. Regardless of what a solution architect does, a work item is returned to the owner.

If changes were made to the work item, the owner's task is to assess the changes and send a review to the solution architect. If a comment (review of the solution architect) was left, the owner must decide whether and what changes will be made to the work item and send the modified work item to the solution architect. This process repeats until the owner and solution architects are satisfied with the contents of a work item.

This has been shown to be a best practice for creating new work items, as it helps both employees understand each other's positions better and shortens development time when necessary information is provided in the work item. The disadvantage is that it takes longer to write work items and creates additional assignments for solution architects. However, compared to previous practices where only business team members wrote work items, it demonstrates better communication between employees and increased user satisfaction. It is easier for a developer to program something when the description and all resources are in the work item. This means fewer bugs in the newly added feature, which results in a better user experience.

After an estimate is given, the work item enters the product backlog at the end (the last item in the product backlog). It stays at the end of the product backlog until the prioritization meeting. At this meeting, all members of the business teams discuss all new work items and decide on a priority. The priority of work items is determined based on the feature timeline (Gantt chart), the agreed-upon deployment dates, and members' options. At that time, work items can change position in the product backlog.

At that point, business team members are done with their work on the discussed work item, and the technical team members (development team) consider these work items during the next sprint planning. At sprint planning, new work items are added to the new sprint until the capacity reaches the 70% limit of the maximum capacity. New items are added according to the previously determined priority.

Most work items are discussed during sprint planning before being added to a new sprint backlog. Some work items are not addressed because they are simple and have straightforward solutions. This discussion is the final point at which work items can be discussed before proceeding to development.

In practice, this discussion has proven beneficial because it allows other development team members to point out possible mistakes or business rule violations. Additionally, it is beneficial for junior developers to observe the thought process of evaluating work items and develop an understanding of what is essential for effective work items. It enables junior developers to advance their knowledge and learn how to estimate the effort needed to resolve work items.

During sprint planning, the development team can reject a work item after discussion if they determine possible mistakes, limitations, or conflicts with other work items. The rejected work items are returned to the business team and are not added to the sprint backlog. The rejected work items can be returned to the development team for the next sprint planning after modification of the description, depending resources or additional clarification.

After sprint planning, when a new sprint begins, the development team starts working on work items in the sprint backlog. During this time, business team members are available to answer any questions that may arise during development.

During the sprint, the business team works on work items for the next sprint and tests and evaluates work items that have already been completed from the previous or current sprint (if a new version is built during the sprint). Additionally, the business team must maintain contact with customers to understand their needs and assist them in resolving any issues that arise during the use of the information system. The task of the business team is to shield developers from outside influence and enable them to focus on solving assigned work items.

The business team also holds daily meetings to discuss its plans. The development team does not participate in these meetings. At these meetings, members discuss feature groups, deadlines, and future deployment dates per customer.

The development team holds daily meetings where every member is required to report on their previous workday and outline their plan for the current workday.

It is beneficial for senior developers and solution architects to know what the junior developers are working on because it enables them to predict if or when they would need assistance. It helps senior developers and solution architects plan their daily obligations.

For junior developers, daily meetings are important to see what could be achieved if they choose to stay in the IT business domain and invest time in learning and code analysis of solutions made by the senior developers.

Every week, the development team has a development backlog meeting. These meetings are used to present significant changes to the core components, address questions and problems for junior developers, and discuss future changes to the information system's architecture. As mentioned before, there is one epic group for architectural improvements. This group is reserved for the development team for features and work items related to the information system's architecture. In this group, developers create work items for themself. These work items address problems that are not known to the business team. The business team lacks insight into the information system's background process and core components.

Developers are responsible for creating work items and further improving core components. Senior developers and solution architects propose ideas at development backlog meetings, and the entire team discusses them. If the proposed changes to the information system would provide benefits, a selected senior developer or solution architect creates a work item. Also, the development team proposes a priority for the work items.

Work items created by developers follow the same procedure as all other work items. Work items are sent to the business team. The business team discusses the work items and sends feedback to the development team. The business team assigns priority to these work items based on the development team's proposition, and most of the time, it is the same as proposed.

At sprint planning, at least one work item created by the development team is added to the sprint backlog, and a senior developer is immediately assigned. The assigned developer is usually the same developer who created the work item. When a new version is built, testing completed work is done by other senior developers or solution architects. Testers do not test these solutions.

During the sprint, one or more new versions of the information system are built. At least one version of the information system must be built at the end of the sprint.

When a new version is created, it is immediately deployed on company servers, and the business team is notified. Testers and project managers are responsible for testing the latest version and analyzing completed work. Completed work items can be returned to the development team for necessary changes. These requests for changes can be made in the form of an issue, a completely new user story, or a bug.

New versions should be presented at the end of the sprint at the sprint review and retrospective meeting, but in practice, that is not the case. Depending on the impact of the changes, that part of the entire process is sometimes skipped.

The described process works best for these teams and their project. This is evident in the company's overall income and customer satisfaction.

6 Conclusion

This paper presented the planning and development process for one of the projects in the Croatian software development company. The paper also describes the chosen information system to provide full context.

The aim of this paper is to describe how the Scrum framework was adopted for the selected project, the impact on the team, and the best practices that emerged, highlighting how this implementation enabled agility in response to quick changes. This implementation demonstrates how the Scrum framework can be applied in practice and customized for each company, project, and team.

The presented processes show the current state of the project. This project is an ongoing project and will continue to be until a new iteration of the information system is started (the new information system started from scratch in the newer technologies) or the customers find a better product.

The processes have possibilities for improvement. However, whether the improvement will be accepted or not will be shown if further analysis and follow-up research are conducted.

References

Andrei, B.-A., Casu-Pop, A.-C., Gheorghe, S.-C., & Boianggiu, C.-A. (2019). A study on using waterfall and agile methods in software project management. *Journal of Information Systems & Operations Management*, 13(1), 125-135

Bento, A. C., Delgado D. A. C., & Camacho-Leon, S. (2023). Experimental Survey Results on Azure.DevOps Application for Management Student's Projects. *2023 Future of Educational Innovation-Workshop Series Data in Action*, 1-5

Bomström, H., Kelanti, M., Annanperä, E., Liukkunen, K., Kilamo, T., Sievi-Korte, O., & Systä, K. (2023). Information needs and presentation in agile software development. *Information and Software Technology*, 162, 107265

Borra, P. (2024). Maximizing Efficiency and Collaboration with Microsoft Azure DevOps. *International Journal of Advanced Research in Science, Communication and Technology*, 4(2), 556-562

Debski, A., Szczepanik B., Malawski, M., Spahr S., & Muthig, D. (2017). A Scalable, Reactive Architecture for Cloud Applications. *IEEE Software*, 35(2), 62-71

Diakov, S. O., Zubrei, T. E., & Samoidiuk, A. S. (2019). Application of Event Sourcing and CQRS Patterns in Distributed Systems. *Adaptive Systems of Automatic Control*, 1(34), 16-22

Dileepkumar, S. R., & Mathew, J. (2021). Optimize Continuous Integration and Continuous Deployment in Azure DevOps for a controlled Microsoft .NET environment using different techniques and practices. *IOP Conference Series: Materials Science and Engineering*, 1085(1), 012027

Erb, B., & Kargl, F. (2014). Combining Discrete Event Simulations and Event Sourcing. *Seventh International Conference on Simulation Tools and Techniques*, 51-55

Itzik, D., & Roy, G. (2023). Does agile methodology fit all characteristics of software projects? Review and analysis. *Empirical Software Engineering*, 28, 105

Kabbedijk, J., Jansen, S., & Brinkkemper, S. (2012). A Case Study of the Variability Consequences of the CQRS Pattern in Online Business Software. *17th European Conference on Pattern Languages of Programs (EuroPLoP '12)*, 2, 1-10

Kufner, J., & Marik, R. (2019). Restful State Machines and SQL Database. *IEEE Access*, 7, 144603-144617

Lima, S., Correia, J., Araujo, F., & Cardoso, J. (2021). Improving observability in Event Sourcing Systems. *Journal of Systems and Software*, 181, 111015

Omonije, A. (2024). Agile Methodology: A Comprehensive Impact on Modern Business Operations. *International Journal of Science and Research (IJSR)*, 13(2), 132-138

Overeem, M., Spoor, M., & Jansen, S. (2017). The Dark Side of Event Sourcing: Managing Data Conversion. *IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 193-204

Overeem, M., Spoor, M., Jansen, S., & Brinkkemper, S. (2021). An empirical characterization of event sourced systems and their schema evolution - Lessons from industry. *Journal of Systems and Software*, 178, 110970

Rajković, P., Janković, D., & Milenković, A. (2013). Using CQRS Pattern for Improving Performances in Medical Information Systems. *6th Balkan Conference in Informatics (BCI 2013)*, 86-91

Saffer, D. (2010). *Designing for Interaction: Creating Innovative Applications and Devices, Second Edition*. New Riders.

Schwaber, K., & Beedle, M. (2002). *Agile Software Development with Scrum*. Prentice Hall

Weerakoon, W., & Kumara, B. T. G. S. (2018). Data Handling and Maintaining Data Consistency in Scalable Replicated Micro-Services. *11th International Research Conference General Sir John Kotelawala Defence University*, 281-286

Zayat, W., & Senvar, O. (2020). Framework Study for Agile Software Development Via Scrum and Kanban. *International Journal of Innovation and Technology Management*, 17(04), 2030002

Barriers to Cyber Hygiene in Hybrid Work Environments: A Case Study of a Lesotho Development Parastatal

Lebohang Nkhabu, Machdel Matthee
Department of Informatics
University of Pretoria
Lynnwood road, Pretoria, South Africa
le.nkhabu@gmail.com, machdel.matthee@up.ac.za

Abstract. *This study's objective is to investigate the barriers to maintaining effective cyber hygiene in hybrid work environments, focusing on a case study within a development parastatal in Lesotho. The shift to hybrid work models, accelerated by the COVID-19 pandemic, has introduced new vulnerabilities in cybersecurity, particularly regarding remote work. Utilizing qualitative, research through interviews with IT specialists, managers, and end-users, the study identifies key barriers confirming themes found in literature affecting cyber hygiene in companies namely, user characteristics (such as awareness and attitude),environmental support (awareness and training, policies, leadership, ICT support) and environmental characteristics (organisational culture, poor infrastructure). It is shown that the developing country context of the parastatal exacerbate these barriers.*

Keywords: Cyber hygiene, hybrid work, cybersecurity, remote work, barriers, Lesotho parastatal

1 Introduction

With the rise of digital transformation, organisations worldwide have shifted towards hybrid work environments, blending in-office and remote work setups (Beno, 2021). The COVID-19 pandemic has significantly accelerated this trend bringing forth both opportunities and challenges in maintaining secure information systems (Huang & Yen, 2021). Among these challenges is the increasing importance of cyber hygiene, which refers to the practices and behaviours individuals and organisations adopt to protect themselves from cybersecurity threats (Vishwanath, et al., 2020). Cyber hygiene is also referred to as cybersecurity behaviour in this study.

Hybrid work environments, while offering flexibility, introduce various cybersecurity vulnerabilities, particularly in remote work contexts where organisational oversight and secure infrastructure are limited (Morris & Still, 2023). Emerging economies are progressively allocating resources towards leveraging web-enabled platforms and services (Lallie, et al., 2021). For developing countries such as Lesotho, where infrastructure and cybersecurity literacy may already be constrained, this shift presents significant risks (Mosola, Moeketsi, Sehobai, & Pule, 2019).

This research investigates the barriers to cyber hygiene in a hybrid work model within a Lesotho development parastatal. Using an inductive, qualitative approach, the study aims to explore how the factors impeding cyber hygiene in this specific context.

2 Background

The rapid transition to hybrid work arrangements during the COVID-19 pandemic has fundamentally altered the cybersecurity landscape for many organisations (Beno, 2021). As employees increasingly operate outside traditional office environments, new and more sophisticated cyber threats have emerged (Al-Mohannadi, et al., 2016). This shift has made cyber hygiene, a set of practices aimed at safeguarding information systems more crucial than ever. However, the hybrid work model brings with it a range of challenges, particularly for employees working remotely without the protective measures provided by corporate IT infrastructure. This necessitates a renewed focus on understanding and mitigating barriers to effective cybersecurity practices.

2.1 Cyber Hygiene and Its Role in Cybersecurity

As the pandemic spurred remote work, the need for robust cyber hygiene became clear. Research identifies cyber hygiene as essential for preventing breaches that compromise system confidentiality, integrity, and availability (Kalhoro, Rehman, Ponnusamy, & Shaikh, 2021). Cyber hygiene within remote settings emphasises secure configurations, device usage protocols, and safe internet practices, which are complicated by unsupervised environments and potentially insecure home networks (Droppa & Harakal, 2021).

2.2 Remote Work Challenges and Cyber Threats

Remote work environments inherently elevate cyber risk by relying on public and personal networks, which lack organisational controls. The dependence on personal devices and unmonitored networks has increased the attack surface, making devices susceptible to malware, phishing, and other cyber threats (Borkovic & Skovira, 2020). Moreover, as users perform sensitive activities like online banking and corporate transactions on potentially insecure networks, the likelihood of exposure to cyber-attacks grows (Kovacevic, Putnik, & Toskovic, 2020) .

2.3 Human and Organizational Factors in Cyber Hygiene

Human behaviour is a significant determinant in cyber hygiene practices, influenced by demographic factors, social norms, and cybersecurity awareness levels. Organisational structures and policies also play crucial roles in shaping employee adherence to cybersecurity protocols. Notably, organisations that actively involve management in cybersecurity initiatives experience better compliance and improved cyber hygiene behaviours among employees (Li, et al., 2019).

2.4 Cyber Hygiene in Hybrid Work

The beginning of the COVID-19 pandemic has heightened the necessity to explore and analyse digital transformation (Kabanda & Chingoriwo, 2021). This exploration has exposed many organisations to new cybersecurity risks (Georgiadou, Mouzakitis, & Askounis, 2022). Studies indicate that cyberattacks, such as phishing, malware, and ransomware, have surged during the pandemic due to increased internet use by remote workers (Lallie, et al., 2021). Phishing attacks, for instance, increased significantly as attackers exploited users' reliance on email communications and their lack of awareness of cybersecurity threats when working from home (Georgiadou, Mouzakitis, & Askounis, 2022). Similarly, malware attacks such as ransomware affected critical sectors, including healthcare and education, with remote workers often lacking adequate security measures to protect sensitive data (Pranggono & Arabo, 2020).

Cyber hygiene is central to mitigating these threats, involving practices such as strong passwords, software updates, and multi-factor authentication (Cain, Edwards, & Still, 2018). However, hybrid work environments, particularly in developing regions, present additional challenges. Employees working from home or public spaces frequently connect to unsecured networks, increasing the risk of cyberattacks (Li, Xin, & Siponen, 2022). Without the protection of corporate IT infrastructures, these individuals become the first line of defence, requiring both technical skills and cybersecurity awareness to mitigate risks (Vishwanath, et al., 2020).

2.5 Barriers to Cyber Hygiene

Studies show that it has become increasingly essential to address the human aspects of cybersecurity, and this study aims to address those aspects as barriers in cybersecurity behaviour or simply, cyber hygiene (Li, et al., 2019). Barriers to cyber hygiene in hybrid work environments can be categorised into personal, environmental, and organizational factors.

2.5.1 Demographics

Research indicates that age, gender, and educational background significantly influence cybersecurity behaviours. Younger employees, while more comfortable with technology, may be less cautious about cybersecurity, often neglecting best practices such as password management and software updates (Whitty, Doodson, Creese, & Hodges, 2015). On the other hand, older employees, though generally more security-conscious, may lack the technical expertise to implement effective cybersecurity measures. Gender also plays a role, with studies suggesting that women may exhibit greater caution when sharing information online but also tend to lack confidence in their cybersecurity skills (Anwar, He, Ash, Yuan, & Li, 2017).

2.5.2 Cybersecurity Awareness

One of the primary barriers to cyber hygiene is the lack of adequate training and awareness programs, particularly in developing countries. Many organisations provide limited cybersecurity training, leaving employees unaware of the risks associated with remote work (Hadlington, 2017). Cybersecurity awareness needs to be tailored to specific roles and responsibilities, with ongoing training that addresses emerging threats (Ani, He, & Tiwari, 2019). However, in Lesotho and similar contexts, budgetary

constraints and limited access to professional development opportunities hinder the implementation of comprehensive cybersecurity programs.

2.5.3 Organisational Culture

The role of organisational culture in promoting cyber hygiene cannot be overstated. Organisations that actively promote cybersecurity through policies, regular audits, and leadership involvement tend to have employees with better cyber hygiene practices (Li, Xin, & Siponen, 2022). In contrast, organisations where cybersecurity is treated as the sole responsibility of the IT department often see lower engagement from employees in maintaining secure practices. Furthermore, inconsistent policy enforcement across departments exacerbates this issue, creating gaps in organisational security (Desolda, Ferro, Marrella, Catarci, & Costabile, 2022).

2.5.4 Environmental Factors

Remote workers often use public or home networks that lack the same level of security as office environments, making them more vulnerable to cyberattacks. Additionally, the absence of direct supervision leads to lax cybersecurity practices, such as failing to update software or using weak passwords. Family dynamics, such as shared use of devices or networks, further complicate cybersecurity efforts at home, especially when devices used for work are also accessed by other household members (Kalhoro, Rehman, Ponnusamy, & Shaikh, 2021).

2.6 Framework for understanding citizen's cybersecurity behavior

Li et al. (2022) places the factors discussed in section 2.5 in a framework to depict its interrelatedness in influencing citizens' cybersecurity behaviour. Li's Framework is shown in Fig. 1 below.

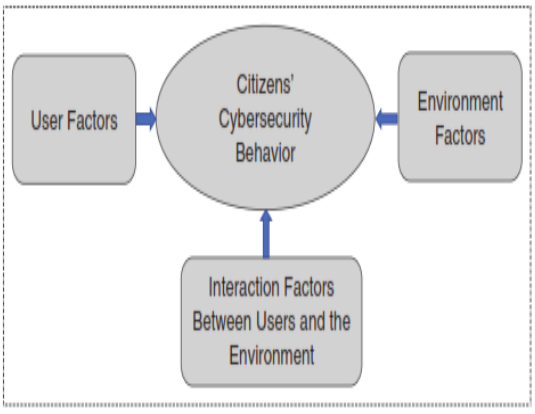


Figure 1. Li's Framework for understanding citizens' cybersecurity behaviour (Li et al., 2022)

Li's Framework identifies user factors like demographics, knowledge and cognitive skills, and environmental factors like activities, support, security, safety, policies, awareness training, sharing and connections, to investigate how they can impact cybersecurity behaviour of people.

Li et al. (2022) also highlights the environmental factors including the characteristics of the home environment, connected devices (IOT) and the shared usage of devices. The interaction between the environmental factors and user factors are guided by the following:

Safety Climate: In organisational contexts, a strong safety climate, shaped by management and peer attitudes, encourages compliance with security protocols. In contrast, home environments rely on the collective awareness and security practices of all family members, making it challenging to establish a cohesive safety climate.

Support Systems: Organisations typically provide robust IT support, training and clear policies to enhance cybersecurity. In contrast, home users often lack access to professional support and rely on informal networks or third-party services, which may not provide adequate protection.

Policies and Training: The presence of formal cybersecurity policies and awareness training in organisations contrasts sharply with the informal and often inconsistent approaches taken by home users. This disparity highlights the need for tailored educational initiatives aimed at improving cybersecurity behaviours amongst citizens.

The framework underscores several important implications for enhancing citizens' cybersecurity behaviour. Firstly, there is a need for comprehensive cybersecurity education that caters to diverse user demographics, addressing specific knowledge gaps and promoting awareness of potential threats. Secondly initiatives aimed at improving the security of home and public networks and devices are essential. This includes providing resources for users to understand how to secure their home environments effectively. Thirdly, developing community-based support systems that facilitate knowledge sharing and provide access to professional cybersecurity resources can help to mitigate risks for remote work users. Lastly policymakers should consider creating frameworks that promote cybersecurity awareness and support for citizens, recognising the unique challenges faced by home users compared to organisational employees.

Note that the focus of Li et. al is also on people working from home, but extends to people working remotely from the field and from public environments. Li et al. go further to break down the environmental factors that may affect behaviour into the characteristics associated with the home environment and the support provided by the work

environment. Fig. 2 below shows an illustration of this:

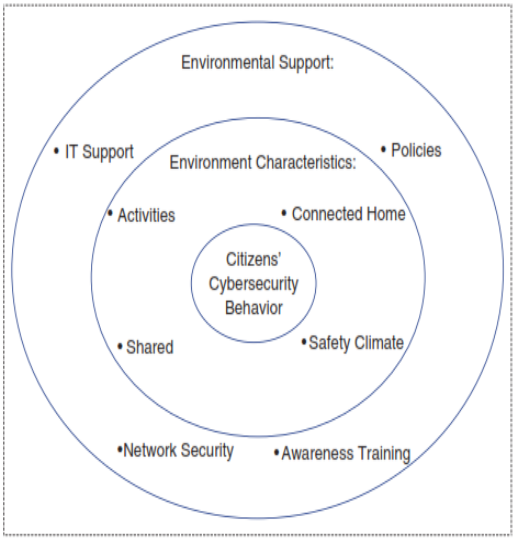


Figure 2. The Environmental Influences on Citizen's Cybersecurity behaviour (Li et al., 2022)

According to Li et al (2022), the environment is characterised by a diverse range of computer and internet-centric activities, such as online shopping, banking, communication, entertainment and education. This contrasts with the more work-focused activities in an organisational context.

Individuals may underestimate the security risks associated with the highly connected home environment. Remote work environments often have a shared environment where family members share internet connections and devices. While convenient, this can lead to privacy concerns and challenges in managing cybersecurity responsibilities.

Organisations can implement comprehensive security plans with significant investments in resources. However, such professional IT support is generally absent in remote work environments, with users relying on third-party cybersecurity services or informal support from social relationships.

Organisational networks are typically more secure compared to home networks, which may have inadequate security measures and configurations. Organisations provide formal cybersecurity awareness training to employees, while home users may only have access to informal training from various sources, such as mass media or social relationships. Organisational environments have relatively complete cybersecurity policies, while home environments often lack clear and comprehensive policies.

The framework emphasises that the characteristics of the home environment, such as diverse activities, connected devices, shared usage and lack of a strong safety climate, combined with limited environmental support in terms of IT assistance, network security, awareness training and policies, create a more challenging landscape for citizens to maintain effective cybersecurity behaviours compared to organisational settings (Li et al, 2022)

3 Methodology

The study employs a qualitative research design within an interpretivist approach, focusing on a single case study of a development parastatal in Lesotho. This is a bi-national organisation established to implement and manage a major water resources development project between Lesotho and South Africa. It has an employment number of around 600 employees. Its primary responsibilities include harnessing water resources for mutual benefit and generating hydroelectric power for domestic use. The project involves the construction and maintenance of dams, tunnels, and associated infrastructure to transfer water from the highlands to South Africa, while also ensuring energy security for Lesotho. It also plays a significant role in environmental management, resettlement planning, and community development in areas affected by the project. In addition, it ensures compliance with legal and environmental obligations, promotes sustainable resource use, and collaborates with various stakeholders to ensure the project's long-term success and minimal ecological impact (LHDA, 2024).

The first author collected data through semi-structured interviews with 13 participants from the organisation, including IT specialists, managers, and end-users from various departments (see Table 1). Purposeful sampling was used to ensure the inclusion of a variety of roles. Job descriptions relate to water quality operation and management, reservoir operation and management, environment management to name but a few. The interviews were conducted between September 2023 and July 2024 using MS Teams. Recordings were downloaded and transcribed. The interviews were conducted in English but the interviewer and interviewee sometimes reverted to Sesotho, the native language to improve correct interpretation.

Table 1. Participant Personal Information

Level	Male	Female	Age bracket
Junior	3	0	33-38
Senior	6	0	38-50
Manager	1	3	48-55

The interviews were designed to gather insights into the participants' experiences with cybersecurity practices during remote work. The interview questions were grouped and designed according to

the framework of Li et al (2022). Organizational documents, such as information security policies, were also analysed to triangulate the data and ensure reliability (Cain, Edwards, & Still, 2018).

Thematic analysis was used to process the interview transcripts, guided by the themes from the framework of Li et al (2022). Environmental (organizational) support, organizational and environmental characteristics, and personal characteristics such as attitudes toward cyber hygiene (Parsons, McCormac, & Butavicius, 2013) were used as broad themes. The analysis aimed to identify barriers and challenges that impede effective cybersecurity practices in hybrid working environments in a developing context. The identified themes are discussed in Section 4.

4 Findings

The findings are presented according to the themes and sub-themes discussed below.

4.1 Environmental support

4.1.1 User awareness and training

This theme emerged from discussions about whether the organisation provided cybersecurity awareness, training, or knowledge to its employees. The awareness initiatives appeared to focus on general cybersecurity practices rather than practical guidance for remote or hybrid work scenarios. Participants generally lacked familiarity with the term cyber hygiene, although many could infer its meaning. A manager noted:

"If you're saying cyber hygiene probably you are saying you are doing the good practices with your computers or with your personal devices."

A junior participant replied:

"I can maybe take it from the term hygiene as in cleanliness. Maybe how we can protect ourselves from someone hacking into the system through us"

As a country with evolving ICT infrastructure and varied digital literacy levels across sectors, formal training on nuanced cybersecurity topics is still emerging.

The participants awareness of cybercrime in general was also explored. Few participants mentioned attacks on local companies such as the Central Bank of Lesotho, the Lesotho National High Court. Others mentioned personal cyber-attacks such as phishing and hacking of WhatsApp. One participant mentioned an attack launched against the case study organisational which fortunately failed.

It appeared that formal knowledge and training were inconsistent across the organisation. For example, a junior participant could not recall receiving any dedicated cyber hygiene training, while others described existing training as overly technical and inaccessible. Another junior participant suggested that the organisation should have a weekly bulletin that provides information on cyber threats and new trends on the cyber threat landscape, to inform every one of them and suggest how they can protect themselves. The interview process itself prompted reflection, as a senior participant noted:

"I realized that there might be practices I can improve after the type of questions I had to answer."

In Lesotho's parastatal sector, which combines technical project delivery with public service obligations, such gaps highlight how cybersecurity awareness must be tailored to a non-uniform workforce. This includes technical field officers, administrative staff, and community-facing personnel, each requiring context-sensitive guidance rather than generic instruction.

4.1.2 Policies

In many Lesotho parastatals, policy development and resource allocation are often reactive, driven by broader national mandates and constrained by budgetary limitations. The organization has an information security policy and most of the participants are aware that it exists, but don't seem to have read it. Some participants did not know that the policy exists, because they said that they normally do not bother to read policies placed on SharePoint. While some staff acknowledged receiving VPNs or work-issued modems, policy clarity was lacking. One manager highlighted the urgency of the pandemic-driven shift to remote work:

"There was no time for anyone to write policies."

This reflects a common reality in Lesotho's government-linked institutions, where digital transformation initiatives are often fast-tracked without accompanying governance frameworks. A senior participant also indicated uncertainty around policy applicability to remote contexts.

It appears that policies exist but staff are generally not aware of it. The following organizational documents were considered: the Information Security Policy, the ICT Acceptable Use Policy, the ICT Asset Management Policy and a Disaster Recovery Plan. The documents touch on password management, device sharing, home and public assets, and the use of personal mobile devices. Clear guidelines are given.

4.1.3 Leadership

Leadership influence emerged as a key enabler. A manager stated:

"Leadership by example and role modelling helped shape responsible behaviour."

This illustrates how in Lesotho's parastatal structure, where hierarchical leadership is strong and often culturally respected, behavioral change can be achieved more effectively through visible senior example rather than formal enforcement mechanisms. Trust and informal mentoring thus play critical roles in shaping secure behaviors.

4.1.4 Infrastructure and IT support

In Lesotho, where many homes lack enterprise-grade internet and power fluctuations are common, technical limitations are significant.

Participants cited issues such as weak Wi-Fi security, outdated equipment, and unsecured smart devices. A concerned was expressed:

"The TV connects to the Internet… they can access it."

This reflects a growing issue in Lesotho's urban households, where the adoption of smart technology often outpaces user awareness about cybersecurity risks. Furthermore, a manager said:

"I am not even sure if we are allowed to use our personal emails for work, I just do it when necessary."

In the parastatal setting, where work often involves cross-border communication, public sector documentation, and sensitive infrastructure data, this ambiguity can expose the organization to significant risk. This also points to a systemic challenge: the coexistence of modern digital demands with legacy policies and limited ICT support capacity in developing nations like Lesotho.

4.2 Environmental characteristics

4.2.1 Connected home and shared public spaces

This theme covers employees' experiences working remotely versus in the office, and how their environment shapes their cybersecurity behaviours. In Lesotho, where infrastructure and connectivity differ drastically between urban and rural areas, remote work often occurs in settings not originally designed for professional tasks.

Participants described common challenges such as domestic distractions, device sharing, and use of insecure networks: "*You are provided with a USB modem for connection which you share with your family"*. A junior participant, who had a more controlled home setup, shared:

"I currently live alone… no one is allowed to come to my home while I work."

Conversely, a senior participant admitted to leaving devices unattended, while another demonstrated awareness by saying:

"I never forget the shortcut Windows logo L to lock my computer."

Such variability is typical in Lesotho's hybrid workforce, where staff may alternate between working from central headquarters, rural project sites, or home environments lacking stable power or connectivity. The disparity in conditions underscores a key issue that parastatal employees must manage cybersecurity responsibilities without consistent infrastructure or supervisory support, making personal habits a critical line of defence.

4.3 User characteristics

4.3.1 Perceptions and Attitudes Towards Cyber Hygiene

This theme investigates how employees perceive cybersecurity risks and their attitudes towards mitigating them

Many employees showed greater trust in the protections offered by office networks:

"There are firewalls, there are proxies... at home, you are not okay."

This belief is grounded in the reality that most homes in Lesotho lack even basic cybersecurity measures, such as network segmentation or up-to-date firewalls. Similarly, a junior participant stated:

"I believe when the organisation gives us the modems, there's some kind of security features."

Such attitudes reflect the broader dependency on institutional protection, common in countries like Lesotho, where individual cyber responsibility is underdeveloped, and access to training or IT support at home is rare. This reliance reinforces a reactive mindset, and employees trust the system when inside the office but feel vulnerable and unsupported outside of it.

Some participants even expressed resignation towards cyber threats, suggesting a lack of empowerment. This indicates the need for a cultural shift within Lesotho's public sector institutions from viewing cybersecurity as a technical issue to seeing it as a shared responsibility requiring both system-level support and individual agency.

5 Discussion

The findings highlight the complex interplay between individual behaviours, organisational structures, and environmental factors that contribute to poor cyber hygiene in hybrid work environments (Droppa & Harakal, 2021).

The findings of this study align with literature and constructs defined by the framework suggested by Li et al (2022): This study confirmed that younger employees were generally more proficient in adopting secure digital practices. In addition, although training exists, participants complained that it is overly technical and inaccessible. Policy implementation is inconsistent and users are

concerned about cyber safety when working from public places or home.

The hybrid work model presents unique cybersecurity challenges, particularly in remote settings where employees are less likely to follow organisational policies. This is consistent with the study by Manzil & Naik (2022). The use of unsecured Wi-Fi networks, shared devices, and weak authentication methods were frequently cited as vulnerabilities as is the case in the investigation by Cross & Gillett (2020).

This study confirmed that organisational culture plays a significant role in shaping cybersecurity behaviours. In environments where management actively promoted and enforced cybersecurity protocols, employees were more likely to adopt good cyber hygiene practices similar to the study by Mosola et al. (2019). Conversely, organisations with less emphasis on cybersecurity training and support saw higher incidences of poor security practices.

The uniqueness of this study lies in the role the context plays in exacerbating the barriers to cyber hygiene. Lesotho is a developing country with unstable electricity supply and poor infrastructure in general. This leads to workers working form different locations depending on the availability of infrastructure, access and electricity. The abdication of responsibility of cyber hygiene to leaders can be ascribed to cultural respect for leaders. In addition, cyber hygiene training in this context is complicated by the parastatal organisation context of non-uniform workers. This context is underexplored in existing literature, which often focuses on developed nations with advanced technological infrastructures.

This study highlights unique cybersecurity vulnerabilities outside controlled office environments. These include unsecured networks, environmental privacy concerns, and socio-cultural barriers.

6 Conclusion

This study provides important insights into the barriers that impede effective cyber hygiene in hybrid work environments, particularly within the context of a Lesotho development parastatal. The findings underscore the need for organisations to adopt a holistic approach to cybersecurity, addressing not only technological factors but also organisational culture and individual behaviours.

To improve cyber hygiene in hybrid work, organizations in developing countries must invest in comprehensive cybersecurity training programs, not only create but enforce strict security policies, and provide adequate technological support for remote workers. From the interviews it appears that good cyber hygiene starts with good leadership and mentorship. By fostering a culture of cybersecurity awareness through effective leadership, organizations can better protect themselves from the growing threat of cyber-attacks in the digital age.

References

Al-Mohannadi, H., Mirza, Q., Namanya, A., Awan, I., Cullen, A., & Disso, J. (2016). Cyber-Attack Modeling Analysis Techniques: An Overview. *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)* (pp. 69-76). Vienna: IEEE. doi:10.1109/W-FiCloud.2016.29

Ani, U. P., He, H., & Tiwari, A. (2019). Human factor security: evaluating the cybersecurity capacity of the industrial workforce. *Journal of Systems and Information Technology, 21*(1), 2-35. doi:10.1108/JSIT-02-2018-0028

Anwar, M., He, W., Ash, I., Yuan, X., & Li, L. (2017). Gender difference and employees' cybersecurity behaviors. *Computers in Human Behavior, 69*, 37-43. doi:10.1016/j.chb.2016.12.040

Bandura, A. (1986). *Social foundations of thought and action : a social cognitive theory.* New Jersey: Prentice Hall.

Beno, M. (2021, May). Analysis of Three Potential Savings in E-Working Expenditure. *Frontiers in Sociology, 6*. doi:10.3389/fsoc.2021.675530

Borkovic, D. J., & Skovira, R. J. (2020). Working from home: Cybersecurity in the age of Covid-19. *Issues in Information Systems, 21*(4), 234-246. doi:10.48009/4_iis_2020_234-246

Cain, A. A., Edwards, M. E., & Still, J. D. (2018). An exploratory study of cyber hygiene behaviors and knowledge. (ScienceDirect, Ed.) *Journal of Information Security and Applications*. doi:10.1016/j.jisa.2018.08.002

Cross, C., & Gillett, R. (2020). Exploiting trust for financial gain: an overview of business email compromise (BEC) fraud. *Journal of Financial Crime, 27*(3), 871-884. doi:10.1108/JFC-02-2020-0026

Desolda, G., Ferro, L. S., Marrella, A., Catarci, T., & Costabile, M. F. (2022). Human Factors in Phishing Attacks: A Systematic Literature Review. *ACM Computing Surveys, 54*(8), 1-35. doi:10.1145/3469886

Droppa, M., & Harakal, M. (2021). Analysis of Cybersecurity in the Real Environment. *2021 Communication and Information Technologies (KIT)* (pp. 1-7). Vysoke Tatry, Slovakia: IEEE. doi:10.1109/KIT52904.2021.9583748

Georgiadou, A., Mouzakitis, S., & Askounis, D. (2022). Working from home during COVID-19 crisis: a cyber security culture assessment survey. *Security Journal, 35*(2), 486-505. doi:10.1057/s41284-021-00286-2

Hadlington, L. (2017). Human factors in cybersecurity; examining the link between Internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours. *Heliyon, 3*(7). doi:10.1016/j.heliyon.2017.e00346

Huang, M., & Yen, B. P. (2021). Driving Forces for Digital Transformation – Case Studies of Q-Commerce. *International Conference on Electronic Business (ICEB). 21*, pp. 117-128. Nanjing, China: Association for Information Systems.

Kabanda, G., & Chingoriwo, T. (2021). A Cybersecurity Culture Framework for Grassroots Levels in Zimbabwe. *Oriental Journal of Computer Science and Technology, 14*(1-2-3), 17-34.

Kalhoro, S., Rehman, M., Ponnusamy, V., & Shaikh, F. B. (2021). Extracting Key Factors of Cyber Hygiene Behaviour Among Software Engineers: A Systematic Literature Review. *IEEE Access, 9*, 99339-99363. doi:10.1109/ACCESS.2021.3097144

Kovacevic, A., Putnik, N., & Toskovic, O. (2020). Factors Related to Cyber Security Behavior. *IEEE Access, 8*, 125140-125148. doi:10.1109/ACCESS.2020.3007867

Lallie, H. S., Shepherd, L. A., Nurse, J. R., Erola, A., Epiphaniou, G., Maple, C., & Bellekens, X. (2021). Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. *Computers & Security, 105*, 192248.

LHDA. (2024). *LinkedIn LHDA Page*. Retrieved from LinkedIn: https://www.linkedin.com/company/lesotho-highlands-development-authority-lhda/mycompany/

Li, L., He, W., Xu, L., Ash, I., Anwar, M., & Yuan, X. (2019). Investigating the impact of cybersecurity policy awareness on employees' cybersecurity behavior. *International Journal of Information Management, 45*, 13.24. doi:10.1016/j.ijinfomgt.2018.10.017

Li, Y., & Siponen, M. (2011). A Call For Research On Home Users' Information. *Pacific Asia Conference on Information Systems. 112*. AIS Electronic Library. Retrieved from http://aisel.aisnet.org/pacis2011/112

Li, Y., Xin, T., & Siponen, M. (2022). Citizens' cybersecurity behavior: Some major challenges. *IEEE Security & Privacy, 20*(1), 54-61. doi:10.1109/msec.2021.3117371

Manzil, H., & Naik, M. (2022). COVID-Themed Android Malware Analysis and Detection Framework Based on Permissions. *2022 International Conference for Advancement in Technology, ICONAT 2022*. Goa, India: IEEE Xplore. doi:10.1109/ICONAT53423.2022.9726024

Morris, T. W., & Still, J. D. (2023). Cybersecurity Hygiene: Blending Home and Work Computing. In W. Patterson, *New Perspectives in Behavioural Cybersecurity: Human Behavior and Decision-Making Models* (pp. 107-119). Boca Raton: CRC Press. doi:10.1201/9781003415060

Mosola, N. N., Moeketsi, K. F., Sehobai, R., & Pule, N. (2019). Cybersecurity Protection Structures: The Case of Lesotho. *International Journal of Computer and Information Engineering*, 158-163. Retrieved from https://publications.waset.org/10010176/pdf

Parsons, K., McCormac, A., & Butavicius, M. (2013). The Development of the Human Aspects of Information Security Questionnaire (HAIS-Q). *Proceedings of the 24th Australasian Conference on Information Systems* (pp. 1-11). Melbourne Australia: ACIS 2013: Information systems: transforming the future.

Pranggono, B., & Arabo, A. (2020). COVID-19 pandemic cybersecurity issues. *Internet Technology Letters, 4*(2). doi:10.1002/itl2.247

Vishwanath, A., Neo, L. S., Goh, P., Lee, S., Khader, M., Ong, G., & Chin, J. (2020). Cyber hygiene: The concept, its measure, and its initial tests. *Decision Support Systems*, 113160. doi:10.1016/j.dss.2019.113160

Whitty, M. T., Doodson, J., Creese, S., & Hodges, D. (2015). Individual Differences in Cyber Security Behaviors: An Examination of Who Is Sharing Passwords. *Cyberpsychology, Behavior, and Social Networking, 18*(1), 1-7.

From Human-Centered to Hybrid: Mapping the Emergence of GenAI in Organizational Actor-Networks

Daniel Fernando Olmedo Mejía
Faculty of Economics and Media
Hochschule Fresenius University of Applied Sciences
Jägerstraße 32, 10117, Berlin, Germany
olmedo.dan@gmail.com

Violeta Cvetkoska
Faculty of Economics-Skopje
Ss. Cyril and Methodius University in Skopje
Blvd. Goce Delchev 9V, 1000 Skopje, North Macedonia
vcvetkoska@eccf.ukim.edu.mk

Abstract*. Generative Artificial Intelligence (GenAI) is reshaping organizations by acting as a non-human actor that both disrupts and catalyzes innovation in socio-technical networks. In this study, we adopt a dual theoretical lens combining Actor-Network Theory (ANT) and the Business Model Canvas (BMC) to analyze how GenAI's integration reconfigures human machine relations and alters core business-model elements. We introduce a novel interactive methodology using GPT-based journaling to elicit insights from professionals, treating GenAI as both the subject and instrument of inquiry. Preliminary findings reveal that GenAI automates routine tasks, sparks cognitive shifts from rote memorization to deeper comprehension, and motivates new learning behaviors, effectively serving as a "reflective partner" in organizational processes. Through ANT, we observe how GenAI enrolls into networks by defining new roles and workflows; through BMC, we see changed value propositions, resources, and cost structures. These results underscore GenAI's potential as a productive disruptor enabling organizations to design innovative processes and value creation pathways. Our contributions are theoretical, by extending ANT to contemporary AI contexts, and practical, by demonstrating how GenAI can be strategically integrated to drive innovation impact.*

Keywords: Generative AI, Actor-Network Theory, Business Model Canvas, Organizational Innovation, Productive Disruption, GPT-based Journaling, Hybrid Agency

1 Introduction

Generative Artificial Intelligence (GenAI) is no longer confined to the realm of experimental technology; it has become an active participant in redefining how modern organizations operate. The rapid proliferation of GenAI tools, particularly large language models such as ChatGPT, DALL-E, and others, marks a shift in how non-human agents interact within organizational environments. These systems do not merely assist humans in executing tasks; they co-construct workflows, influence decisions, and, increasingly, redefine the meaning of knowledge work itself. This phenomenon presents both opportunities and challenges. While GenAI promises automation, efficiency, and creative augmentation, it also introduces uncertainties related to trust, governance, and the erosion of traditional roles.

Why is this problem relevant? Organizations across sectors are already integrating GenAI into diverse processes from customer service and marketing to research and strategic planning. A global study by LeanIX (2024) revealed that 80% of IT professionals report GenAI adoption in their companies, yet only 14% claim full visibility into its use. This imbalance underscores a critical gap in strategic integration and governance. As organizations rush to adopt GenAI, they often do so without fully understanding its impact on business models, internal structures, and the sociotechnical networks that sustain daily operations. Previous research has explored GenAI from multiple angles. Operational studies confirm its productivity benefits, while psychological and design research investigate user perceptions and usability. However, these studies often treat GenAI as a tool rather than as an actor capable of altering organizational dynamics. Simultaneously, Actor-Network Theory (ANT) has provided powerful insights into the distributed agency of human and non-human entities, especially in technology-rich environments. Yet its application to GenAI remains limited. The Business Model Canvas (BMC) has also proven effective in tracking changes in business structures but rarely intersects with sociotechnical theories like ANT.

To address this gap, our research combines ANT and BMC to examine GenAI as a productive actor in organizational networks. We argue that GenAI not only supports innovation but actively disrupts and reconstructs value creation processes. Our methodology is equally innovative: we use GPT-based interactive journaling to capture professionals lived experiences with GenAI. Participants engage with an AI-powered journaling tool that prompts reflective

thinking about GenAI's role in their work. This method transforms GenAI from object of analysis into co-participant, generating insights that are both experiential and theoretically grounded.

This study extends ANT by demonstrating how GenAI acts as a non-human agent in organizational translation processes. It also shows how GenAI reconfigures business model components, especially key activities, resources, and value propositions. Furthermore, it introduces GPT-based journaling as a novel tool for qualitative research, aligning with participatory and reflexive methodologies. Strategically, the paper positions GenAI as a designed disruptor that can be intentionally embedded to drive innovation.

The rest of the paper is organized as follows: Section 2 reviews literature on ANT, BMC, and GenAI in organizations. Section 3 describes our qualitative methodology and data collection procedures. Section 4 presents and analyzes the results. Section 5 discusses the results. Section 6 concludes by evaluating whether we met our aim, highlighting limitations, and outlining future research directions.

2 Literature Review

2.1 Theoretical Framework: ANT and BMC

Actor-Network Theory provides a sociotechnical lens for this study, treating both people and artifacts as agents in networks that co-create social reality. Foundational works by Callon and Latour introduced the idea of translation processes: *problematization*, *interessement*, *enrolment*, *mobilization* through which new actors (human or non-human) redefine relations and stabilize networks. In digital contexts, scholars note that algorithms and AI systems exhibit agency, reshaping power dynamics (Montoriu Gutiérrez, 2023) and giving rise to "hybrid collectives" of humans and machines. Recent ANT applications (e.g., Jensen & Sandström, 2019; Bertella & Tomassini, 2024; Ryan et al., 2024) demonstrate how combining ANT with discourse or case analysis can uncover how technological and human interests align or conflict in complex projects. However, few studies have explicitly treated AI tools as actors on equal footing with people. This gap suggests the need to extend ANT into the era of generative AI, where non-human agency can be more visible and transformative.

Complementing ANT, the Business Model Canvas offers a practical structure to capture organizational innovation. Osterwalder and Pigneur (2010) developed the BMC to help practitioners articulate key elements: value proposition, customer segments, channels, revenue streams, etc. of a business. Recent work has further adapted the BMC for digital and sustainable contexts, emphasizing ecosystems, co-creation, and platform governance. For example, Bachmann et al. (2025) expand the BMC to 18 elements for digital sustainability challenges, and Schlimbach and Asghari (2020) propose a "Digital Canvas" for designing digital business models. These frameworks underline that business model innovation is critical when new technologies emerge. Yet, how generative AI specifically alters these canvas components for instance, by enabling new value propositions or automating previously human activities remains underexplored. Integrating ANT with the BMC, therefore, allows us to theoretically capture GenAI's multi-dimensional impact: as it enrolls into networks (ANT), it simultaneously reshapes business logic and strategies (BMC).

2.2 Generative AI in Organizational Practice

Generative AI refers to systems (like GPT, DALL-E, etc.) that produce content or insights, and it has begun to significantly affect workplaces. Empirical studies reveal both promise and caution. Productivity gains are often reported: Noy and Zhang (2023) found that deploying ChatGPT reduced task completion time by 40% while also improving the quality of professional writing. Cui (2024) shows that Chinese manufacturers using GenAI saw higher productivity, leading to stronger e-commerce brand-building. Startups are rapidly embracing GenAI for *growth hacking*, e.g., for product-led development or automated sales content though leaders also worry about AI bias and loss of human touch. Knowledge workers in science labs similarly use GenAI primarily for writing assistance and workflow automation, alongside concerns about data reliability and overdependence. Users often regard AI outputs as helpful advice: Böhm et al. (2023) report that people still trust AI-generated solutions even when rating AI authors as slightly less competent than humans.

However, challenges abound. Many companies lack governance frameworks: a global survey (LeanIX, 2024) reports that only 14% have visibility or governance in place. Potential downsides include unregulated "shadow AI" usage and biases in AI outputs. Researchers note that public discourse on GenAI tends to emphasize economic and business themes, whereas academic discussions focus on technical possibilities (Zhang et al., 2024), indicating a gap between hype and scholarly insight. Overall, this literature establishes that GenAI is a disruptive force in organizations (boosting performance, changing job content), and thus warrants analysis as an actor that can both destabilize and innovate organizational arrangements.

2.3 Actor-Network Theory in Technological Integration

Applying ANT in practice often reveals how success or failure emerges from actor interactions rather than

isolated factors. For instance, studies of IT implementation (e.g., Cresswell et al., 2010) found that outcomes depended on how human users and technical systems enrolled each other through translation processes. In healthcare, Ryan et al. (2024) systematically review over 100 ANT studies and note that although ANT is widely used, many applications only partially employ its core elements. This suggests caution: we should strive to fully trace the stages of translation and network formation when studying GenAI. Other ANT cases like the Kiruna mine study (Jensen & Sandström, 2019) highlight that infrastructures and local communities co-enact "spaces" of work. These examples encourage us to treat GenAI not just as a background tool, but as a "quasi-actor" that participates in negotiation, co-creation, and potentially dissent within organizational networks. Recent ANT scholarship on AI (e.g., Montoriu Gutiérrez, 2023) also emphasizes new power dynamics of algorithms, reinforcing our stance that GenAI's agency deserves explicit theorization.

2.4 Business Model Canvas and Digital Innovation

The BMC literature confirms that business models must evolve when confronted with digital disruptions. Studies using the BMC show how incumbents and startups reconfigure their offerings. For example, Rahman et al. (2025) analyze U.S. gas stations using the BMC, revealing how traditional businesses integrate EV charging and retail services to diversify. Knowledge of the BMC also correlates with entrepreneurial intent: Hendrasto et al. (2024) find that awareness of the BMC significantly boosts young people's interest in starting businesses. Specialized canvases such as the Platform Business Model Canvas (Eisape, 2019) further illustrate how complex value networks can be captured. The consensus is that static business models struggle with emerging tech, so organizations benefit from iteratively adapting their BMC components (partners, activities, revenue streams, etc.) in response to innovation. Our study situates GenAI within this tradition: we use the classic BMC to trace how GenAI's presence adds new value streams and alters cost structures, complementing recent digital extensions of the canvas.

This review highlights that while GenAI's operational effects are documented (productivity boosts, usability concerns, shifts in task perception), there is still a theoretical and empirical gap in viewing GenAI as an actor that *re-designs* networks and business models. We address this by synthesizing ANT and BMC perspectives in the following research.

3 Methodology and Data

3.1 Research Approach

We follow an exploratory qualitative design, well-suited for unpacking a nascent phenomenon like organizational GenAI usage. Rather than testing hypotheses, we aim to trace patterns and meanings as experienced by practitioners in context. This inductive approach acknowledges that GenAI's role is rapidly evolving, so open-ended data collection can surface unexpected themes. To align with ANT, we focus on interactions: how does GenAI get enrolled, who are the allies/adversaries, what controversies emerge? Concurrently, the BMC guides us to map any noted changes onto business-model dimensions (e.g., "new customer relationships" or "altered key activities"). Thus, our method is both theoretical and empirical, blending sociological interpretation with business analytics.

A novel element of our method is GPT-based interactive journaling. Inspired by Pretorius and Pretorius (2025), we use a custom GPT chat interface as an "AI diary". Participants answer open-ended questions about their GenAI experiences, and the GPT system adaptively asks follow-ups to probe themes like task automation, collaboration, productivity, and business. Crucially, the GPT journal does *not* give advice; it prompts reflection. This technique has three key advantages: (1) it standardizes questioning while allowing rich, free-form responses, (2) it implicitly demonstrates GenAI's conversational role

(participants see the AI as a partner in reflection), and (3) it generates data that already aligns with our theoretical lenses (the journal can incorporate BMC concepts in prompts).

In parallel, we collected WhatsApp voice notes (1–2-minute recordings) on guided themes, broadening participation to those who might find voice more natural. These audio reflections covered emotional responses to AI, observed workflow shifts, and impact on value generation. Together, the text journals and voice notes provide a triangulated dataset, enhancing depth. All participants gave informed consent and data were anonymized, with ethical protocols (GDPR compliance, member checking) strictly followed.

3.2 Sampling and Data Collection

We targeted professionals across roles and industries who have direct experience using GenAI at work. Initial contacts were recruited via a LinkedIn call-out, supplemented by snowball referrals. Selection was purposive, not random: we sought diversity in job function (e.g., developers, marketers, analysts) rather than representativeness. Each participant engaged in one GPT journaling session (five adaptive questions, lasting about 10-15 minutes) and submitted one voice note. The journaling interface was multilingual (English, Spanish and German) to accommodate

backgrounds, and each session concluded by sharing the transcript for analysis. While the broader study includes six participants, this paper focuses exclusively on the data collected from one participant, selected for its richness and relevance in illustrating the conceptual integration between Actor-Network Theory and the Business Model Canvas.

No quantitative data was collected or analyzed, as the study follows a purely qualitative and exploratory design.

3.3 Data Analysis

All qualitative data were analyzed using inductive content coding. Transcripts were first coded openly to identify emergent themes. Then we applied the ANT/BMC frameworks deductively: for each code, we asked, "Does this indicate a new actor-role or network effect?" and "Which canvas component does this touch?". For instance, when a participant described GenAI helping them learn faster, we coded it under "learning transformation" (an emergent theme) and linked it to the BMC's "key resources" and "value proposition" (as it became a cognitive resource and added learning value). We used ANT concepts (from Callon's translation stages) to trace how GenAI was introduced and stabilized in the workflow

(problematization: defining a need for summarization; interessement: convincing workers to try GenAI; enrolment: integrating it into processes).

To ensure rigor, we practiced reflexivity and triangulation. The researcher kept an analytic journal to note personal biases, and key insights were reviewed with participants (member checking) where possible. By intersecting multiple data sources (GPT journal, voice notes) and theories, we built a rich understanding of GenAI's organizational role that goes beyond any single viewpoint.

4 Results and Analysis

The preliminary analysis of our first GPT journal session yields three dominant themes: (1) Task Automation, (2) Learning Transformation, and (3) Productivity & Motivation. These themes illustrate how GenAI entered the participant's work network and altered business-model elements.

- Task Automation: The participant initially used GenAI (a summarization assistant) to automate routine parts of document handling. Over time, this led to a translation of roles: GenAI became a quasi-colleague responsible for synthesis tasks. In ANT terms, the process moved through stages: the problem was defined (GenAI summarization), the participant enrolled the AI by feeding texts, and gradually the AI's output was accepted into the workflow. BMC-wise, the key activities shifted summarizing tasks were streamlined, and key resources expanded to include the AI tool itself. The participant noted, "What used to take hours of reading now takes minutes," effectively changing the cost structure (time was saved).
- Learning Transformation: A striking pattern emerged in cognition. The participant described a shift from memorization to comprehension: rather than recalling facts, they now focused on understanding concepts aided by GenAI. The AI acted as a reflective partner, prompting deeper processing of information. This aligns with Pretorius and Pretorius (2025) who found ChatGPT can enhance insight generate on, and with Dedema and Ma (2024) who report creative benefits from AI-assisted brainstorming. In our case, the BMC "value proposition" evolved: the AI did not replace learning but enriched it, adding new educational value. This theme shows GenAI's non-human agency: it became a co-learner, demonstrating how human skills and machine capabilities co-construct new forms of knowledge (an ANT enactment of co-becoming).
- Productivity and Motivation: The third theme was increased efficiency and engagement. The participant felt "more motivated" knowing that tedious steps were handled by AI. This echoes the findings of Noy and Zhang (2023), who observed that access to ChatGPT led to substantial productivity gains by enhancing both the efficiency and quality of professional writing. The psychological boost from cognitive load, also emerged, suggesting that GenAI's disruption can be positive: automating drudgery freed the participant to tackle creative tasks. In business-model terms, productivity gains lower operational costs and possibly enables new services or faster delivery. The participant planned to use the saved time for extra training, implying an indirect value proposition for skill development.

To illustrate the dynamics of GenAI integration, Fig. 1 presents the GenAI summarization process mapped through the lens of Actor-Network Theory. This figure captures how GenAI progressively integrates into organizational workflows by undergoing the five translation stages outlined in ANT: problematization, interessement, enrolment, mobilization, and stabilization, thereby evolving from a peripheral tool to a central actor in value creation.

The researcher also played the dual role of designer and analyst of the journaling tool. This reflexivity meant being aware that the GPT prompts could steer responses. We maintained transparency by recording all prompt changes and using multiple coders to verify theme consistency. The fact that GenAI generated our data as well as our analysis framework is itself illustrative: it shows how AI can be a partner in research, co-constructing the findings just as we aim to reveal its co-construction of organizational work.

This prototypical case demonstrates that GenAI's presence does more than automate, it translates into a

new actor-network configuration. GenAI ceased to be background machinery and instead participated in value creation, knowledge work, and network stabilization.

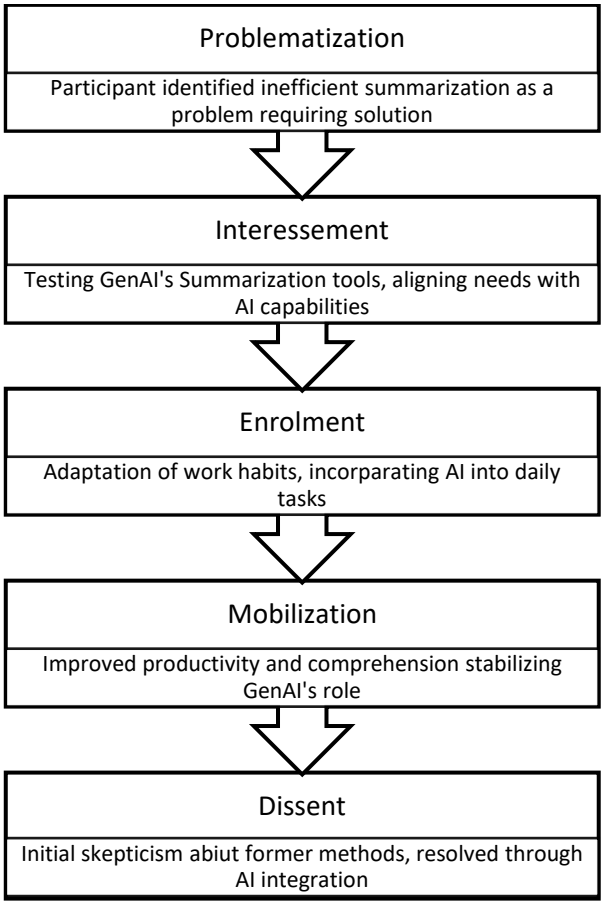


Figure 1. GenAI summarization process based on ANT

5 Discussion

Our findings reveal clear parallels with existing research, reinforcing our dual-lens approach. The observed productivity gain and cognitive shift are consistent with Noy and Zhang (2023), who found that generative AI tools significantly improved task efficiency and the quality of professional outputs. Similarly, the emergence of GenAI as a "reflective partner" confirms Pretorius and Pretorius's (2025) insight that ChatGPT can enhance reflexivity and insight in qualitative research. From an ANT perspective, the stages of integration we identified mirror the classic translation process: GenAI was problematized (as a needed resource), interessement occurred (the participant accepted the AI's help), and the network stabilized as GenAI became accepted (enrolment). These patterns echo Xu et al.'s (2022) e-commerce study where actor roles shifted through translation to maintain network stability. Finally, the impacts on BMC components (e.g., value proposition, cost structure) align with calls in the business-model literature for adaptive BMCs in the face of digital technologies.

From an academic perspective, this research makes three key contributions. **First**, it extends *Actor-Network Theory* into the domain of *generative AI*, offering empirical evidence that GenAI is not a passive tool but an active, enrolling agent in organizational networks. **Second**, it demonstrates how ANT and BMC can be productively combined to map both the relational and strategic dimensions of digital transformation. **Third**, it introduces and validates *GPT-based journaling* as a novel, reflexive qualitative method that aligns methodologically with the hybrid human-machine subject under investigation.

On the practical side, our findings offer actionable guidance for managers, innovation strategists, and organizational designers. We recommend that GenAI adoption be framed not as a technology deployment, but as a strategic redesign process. Organizations can redefine business model elements by embedding GenAI into key resources and activities, using it to unlock new value propositions such as hyper-personalized services, faster ideation, and reduced operational overhead. For instance, HR departments may reconfigure onboarding workflows using GenAI chat interfaces, while marketing teams may deploy GenAI as co-creators in campaign generation, altering not only output speed but also content strategy.

Moreover, organizations should approach GenAI integration as a network negotiation process, enrolling stakeholders, realigning responsibilities, and explicitly mapping new actor roles to avoid resistance and ensure sustained value. By understanding GenAI as a socio-technical actor, rather than a simple IT tool, firms can implement change initiatives that are more inclusive, adaptive, and innovation-driven.

Incorporating GenAI also requires a reconfiguration of organizational agility. Traditional change management models that rely on rigid, linear processes are ill-suited to the fluidity GenAI enables. Instead, organizations must adopt AI-responsive operating models, dynamic frameworks built on modularity, real-time feedback, and iterative learning. For example, project teams can embed GenAI into agile sprints for ideation, experimentation, and reflection, effectively creating rapid innovation loops. GenAI thus becomes not just a resource, but a rhythm, an embedded cadence of continual sensemaking and redesign.

Finally, we encourage decision-makers to treat GenAI not just as a cost-saving mechanism but as a catalyst for cultural transformation. This includes investing in training programs that foster human-AI collaboration, rethinking organizational routines to emphasize co-learning, and designing governance frameworks that mitigate risks while enabling experimentation.

Cross-functional integration will become a competitive imperative. GenAI must not remain

isolated within IT or innovation departments, it needs to be embedded across core business functions. In finance, it can be used for real-time forecasting and strategic modeling; in operations, it can simulate logistics optimizations; in compliance, it can help standardize documentation flows. Such integration calls for new collaborative infrastructure: AI steering committees, shared governance protocols, and interdepartmental AI design charters. Only when GenAI is horizontally embedded does its full organizational potential emerge.

We also note some challenges which mirror the literature. Concerns about AI bias and overdependence (Wagman et al., 2025) emerged; participants occasionally questioned the accuracy of AI summaries, aligning with prior fears about data reliability. Moreover, governance gaps noted by LeanIX (2024) suggest that as GenAI's footprint grows, companies must proactively manage it to avoid uncontrolled disruptions. Thus, a practical recommendation is to pair the enthusiasm for GenAI-driven innovation with robust policies: treat GenAI as a strategic asset that requires role definitions, monitoring, and alignment with organizational goals.

The practical implication is clear: productive disruption from GenAI is most impactful when it is intentionally orchestrated, embedded in business model innovation, and treated as a strategic actor within evolving organizational networks.

Ultimately, GenAI invites a reimagining of value itself, not merely as output, but as experience, intelligence, and co-authorship. It enables the creation of AI-native services that are insight-driven, adaptive, and uniquely generative services that could not exist without the hybrid capacity of human-machine cognition. Strategic leaders must move from the logic of substitution ("what tasks can AI do?") to the logic of synthesis ("what new forms of value can emerge from human-AI partnership?"). This is the frontier of competitive advantage in the GenAI era.

6 Conclusion and Recommendations

This study sets out to investigate how Generative Artificial Intelligence operates as a nonhuman actor capable of transforming organizational networks and reconfiguring core elements of business models. Through the application of Actor-Network Theory (ANT) and the Business Model Canvas (BMC), and the innovative use of GPT-based journaling as a methodological tool, the aim has been successfully achieved. We have demonstrated that GenAI goes beyond being a technological instrument to become an active participant in shaping new forms of organizational value creation.

The findings reveal that GenAI enhances productivity through task automation, initiates cognitive transformation by supporting shifts from memorization to comprehension, and functions as a reflective partner that encourages learning and motivation. These observations affirm that GenAI takes part in the translation processes described by ANT, where it is gradually enrolled as a stable element in organizational workflows. Simultaneously, its impact on business model elements such as key activities, resources, value propositions, and cost structures illustrates its disruptive and constructive influence, validating the utility of the BMC framework.

Despite these promising insights, this study is not without limitations. The scope of the analysis was restricted to one GPT-based journaling session complemented by a limited number of audio reflections. While reflexive and triangulation strategies were applied, the dual role of the researcher as tool designer and analyst may introduce interpretive bias. Moreover, the findings, while rich in insight, may not be generalizable without broader and more diverse datasets.

Future research should expand the participant pool across industries and geographies and adopt longitudinal designs to capture the evolution of GenAI's organizational role over time. Comparative studies between firms with mature versus emerging GenAI integration practices could yield valuable contrasts. Quantitative assessments of GenAI's performance impact, coupled with policy-oriented studies on governance and ethical deployment, will further strengthen this emerging field.

This research reinforces the view that GenAI is not merely a technology to be adopted, but an actor to be understood, engaged with, and strategically integrated. Its potential as a productive disruptor lies in the hands of organizations that are prepared not only to embrace disruption but to design innovation through it.

References

Bachmann, N., Thienemann, A. K., Tüzün, A., Brunner, M., Tripathi, S., Pöchtrager, S., & Jodlbauer, H. (2025). The evolution of the business model canvas for digital sustainability. *Procedia Computer Science, 253*, 1012–1023. https://doi.org/10.1016/j.procs.2025.01.163

Bertella, G., & Tomassini, L. (2024). A relational approach to event sustainability: Applying actor–network theory and Foucauldian discourse analysis to a music event. *Event Management, 28*(7), 987–1003. https://doi.org/10.3727/152599524X17113237364163

Böhm, R., Karger, C. P., Lugrin, J.-L., Reiss, N., Zillner, S., & Mühlhoff, R. (2023). People devalue generative AI's competence but not its advice in addressing societal and personal challenges. *Scientific Reports, 13*(1), 18403. https://doi.org/10.1038/s41598-023-44917-4

Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St. Brieuc Bay. In J. Law (Ed.), *Power, action and belief: A new sociology of knowledge?* (pp. 196–223). Routledge & Kegan Paul.

Cresswell, K. M., Worth, A., & Sheikh, A. (2010). Actor-Network Theory and its role in understanding the implementation of information technology developments in healthcare. *BMC Medical Informatics and Decision Making, 10*, 67. https://doi.org/10.1186/1472-6947-10-67

Cui, H., & Yasseri, T. (2024). AI-enhanced collective intelligence. *arXiv preprint* arXiv:2403.10433. https://arxiv.org/abs/2403.10433

Dedema, M., & Ma, R. (2024). The collective use and perceptions of generative AI tools in digital humanities research: Survey-based results. *arXiv.* https://arxiv.org/abs/2404.12458

Eisape, T. (2019). The platform business model canvas: A proposition in a design science approach [Master's thesis, University of Twente]. https://essay.utwente.nl/79739/

Gutiérrez, J. L. M. (2023). On actor-network theory and algorithms: ChatGPT and the new power relationships in the age of AI. *AI and Ethics, 4*(4), 1071–1084. https://doi.org/10.1007/s43681-023-00314-4

Hendrasto, N., Chairiyati, F., Haidar, A., Herindar, E., & Rehman, H. M. (2024). The impact of knowledge about Business Model Canvas (BMC) on entrepreneurial interest of young generation in Jabodetabek. *Indonesian Journal of Business and Entrepreneurship, 10*(1), 146–156. http://dx.doi.org/10.17358/IJBE.10.1.146

Jensen, T., & Sandström, J. (2019). Organizing rocks: Actor–network theory and space. *Organization, 27*(5), 701–716. https://doi.org/10.1177/1350508419842715

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory.* Oxford University Press.

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers.* Wiley.

Pretorius, L., & Pretorius, C. (2025). Exploring ChatGPT's potential as a qualitative research partner: Researcher and participant perspectives on AI-generated insights. Monash University. https://doi.org/10.26180/28386053

Rahman, K. O., Rezvi, R. I., Nasrullah, F., Islam, M. S., Hasan, M., Khanam, A., & Akash, A. H. (2025). Business Model Canvas: Business analytics on gas stations with C-stores in United States. *Journal of Business and Management Studies, 7*(1), 180–185. https://doi.org/10.32996/jbms.2025.7.1.13

Ryan, T., Ryan, N., & Hynes, B. (2024). The integration of human and non-human actors to advance healthcare delivery: Unpacking the role of actor-network theory, a systematic literature review. *BMC Health Services Research, 24*, 1342. https://doi.org/10.1186/s12913-024-11866-4

SAP LeanIX. (2024). *SAP LeanIX AI Survey Results 2024.* LeanIX GmbH. https://www.leanix.net/en/download/sap-leanix-aisurvey-results-2024

Schlimbach, T., & Asghari, H. (2020). Digital canvas: A structuring framework for digital transformation. In A. Belliger & D. J. Krieger (Eds.), *Organizing networks: An actor-network theory of organizations* (pp. 45–62). transcript Verlag.

Wagman, K. B., Dearing, M. T., & Chetty, M. (2025). Generative AI uses and risks for knowledge workers in a science organization. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–19). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713827

Xu, N., Xu, C., Jin, Y., & Yu, Z. (2022). Research on the operating mechanism of e-commerce poverty alleviation in agricultural cooperatives: An actor network theory perspective. *Frontiers in Psychology, 13*, 847902. https://doi.org/10.3389/fpsyg.2022.847902

Zhang, W., Gao, J., Shi, Y., Xuan, K., Li, T., Yang, T., & Sun, W. (2024). Q-compensated image-domain least-squares reverse time migration through preconditioned point-spread functions. *Geophysical Journal International, 229*(2), 1234–1250. https://doi.org/10.1093/gji/ggad123

Digitalization of Accounting in the 21st Century – A Bibliometric Analysis

Zvonimir Vukić
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
zvukic24@student.foi.hr

Abstract. *Digitalization of accounting is starting to emerge as a main research issue in the context of business process re-engineering, technological progress, and professional accommodation of accounting to the digital world. This bibliometric analysis tries to present a map of current research on digitalizing accounting, determine main authors, articles, and topics, and examine their citation frequencies. The results indicate that the most quoted papers address the digitalization of public services, the skills of accountants in an automated world, and the role of technology in the accounting profession of the future. This analysis adds to the knowledge of how digitalization is influencing the practice of accounting and what are the research directions to pursue.*

Keywords. digitalization, accounting, research process, bibliometric analysis, VOSviewer

1 Introduction

The accounting profession has faced significant changes through history, adapting to shifting business environments and technological advancements. From paper-based books to computer networks, each technological revolution has transformed the practice of accounting. Now, people are facing with another major transformation driven by digital technologies. According to Stoica and Ionescu-Feleaga (2021a) digitalization is revolutionizing traditional accounting operations by streamlining routine tasks and emphasizing technological capabilities. This shift goes beyond mere automation, it's a complete overhaul of accounting procedures, outcomes, and the whole profession.

In recent years, the global landscape of digital adoption has expanded dramatically, especially since the COVID-19 pandemic, which accelerated digital change across industries. In their paper Ben Letaifa and Hadj Salem (2023) describe how digitalization became a vital survival factor for accounting companies during the pandemic, focusing on increased efficiency and adaptability in terms of implementation costs and training needs. This is part of a broader response to global technology trends that are reshaping business processes worldwide.

The importance of digitalizing accounting should not be underestimated. Balakrishnan and Rahman (2022) highlight that digitalization is central to achieving sustainable development goals through automation, transparency, and reduced capital costs. Hrebenyk et al. (2023) further emphasize that digitalizing accounting procedures is crucial for enhancing data security, particularly in challenging environments. These arguments indicates that digitalization in accounting is not just a technological advancement but a strategic imperative with critical implications for organizational resilience and sustainability.

The following chapters emphasizes the literature review and the research methodology. The fourth chapter then outlines the research findings, including descriptive statistics, visual analysis, as well as discussions on the study's implications and its limitations. The paper concludes with a comprehensive conclusion of the conducted research.

2 Theoretical review

The digitalization of accounting has become a central topic in both professional and academic discussions, focusing on automation, sustainability, and firm performance. AbuAkel and Ibrahim (2023) emphasize the role of perceived relative advantage, top management support, and IT infrastructure in fostering trust and implementing e-filing systems. Similarly, Afifa et al. (2024) identify managerial support and technology availability as essential factors driving the adoption of digital systems among accountants.

Stoica & Ionescu-Feleaga (2021b) highlight how digitalization is transforming the traditional role of accountants by automating routine tasks and increasing the demand for technological skills. Langmann and Kokina (2021) further illustrate this shift by showcasing Robotic Process Automation (RPA) as a tool that enhances efficiency in tasks such as journal postings, payables, and travel expense monitoring, achieving up to 90% time savings. Furthermore, Robotic Process Automation (RPA) is revolutionizing

the accountants' roles in organizations, such as their role as identifiers, explainers, trainers, maintainers, and analysts in automated processes. The research process entails a case study of interviews of eight organizations that were adopting RPA. Analysis indicates that the accountants' role is changing from executors to process overseers, determining the skills required to adopt RPA effectively (Kokina et al., 2021).

Agostino et al. (2022) analyse the condition and future of research on digitalization, accountability, and public services accounting on the basis of a 232-article survey. The most striking concerns of digital accountability like dialogical accountability, boundary blurring and role, and the urgency of social equality are pointed out. The significance of increased emphasis on data translation and its implication is also emphasized by the study. The study paves the way for more studies into these matters in the public sector and how they affect the practice of accounting. The impact of digitalization extends to professional skills and accounting education. Shuhidan et al. (2023) and Marioara et al. (2022) stress the importance of integrating ERP systems, cloud computing, and digital literacy into curricula to prepare future accountants for evolving demands. However, challenges persist, Arfaoui and Kammoun (2023) identify the lack of digital infrastructure and resistance to change as significant barriers in Tunisia during the COVID-19 pandemic.

Digitalization also plays a broader role in achieving Sustainable Development Goals. Balakrishnan and Rahman (2022) argue that it enhances transparency while reducing capital costs. The digital transformation of accounting is reshaping the profession, bringing forth new opportunities and challenges across legal frameworks, human resources, and the adoption of advanced technologies. Legal and institutional frameworks are equally important for advancing digital transformation. Remlein et al. (2022) discuss how legislation can drive Poland's adoption of digital technologies, while Zainuddin et al. (2025) highlight blockchain's potential to improve accountability. Legislation on data protection (e.g., GDPR), electronic invoicing, and digital signatures directly impacts how accounting information is processed and stored (Stoica & Ionescu-Feleaga, 2021b). In a similar context, Shygun and Chystova (2022) refer to how digital tools enable tax enforcement and visibility at ease, highlighting the need for the most sophisticated regulatory environments.

In practice, technologies like machine learning, AI, blockchain, and cloud computing are revolutionizing accounting processes. Jassem (2024) goes further by positioning AI not only as an automation tool but also as a strategic instrument for foresight and decision-making within Industry 5.0. Artificial intelligence in accounting transforms data analysis, fraud detection, and decision-making. Baker (2024) and Pal et al. (2024) show how machine learning improves accuracy and operational efficiency. These developments raise new ethical and procedural questions requiring further research. HR issues are another field of accounting digitalization. According to Grosu et al. (2023) and Arfaoui and Kammoun (2023), resistance and lack of training are the concerning issues. Enhancing accountants' knowledge of financial principles and technological skills is becoming increasingly important.

Similarly, Baidybekova et al. (2021) and Bathla et al. (2024) emphasize the transformative effects of secure digital foundations on accounting workflows. Digitalization is also reshaping public sector accounting. Quang Huy and Kien Phuc (2024) explore its role in public sector servitization, while Kartikasari et al. (2024) examine its application in monitoring the financial flows of enterprises supported by local governments. Recent research emphasizes that successful digitalization of accounting depends on systematic organizational procedures (Puspitawati et al., 2024) and skilled staff, improving the quality of financial reporting with quicker and more reliable results (Jabor & Hamdan, 2023). Learning material should be customized to address digital transformation needs (Oleiwi, 2023), whereas system compatibility and stakeholder pressure are motivators and training needs are barriers (Valentinetti & Rea, 2024). While digitalization promises, it generates the need to manage risks and stay informed (Arslan, 2024). In the public sector, it enhances transparency in the budget (Sergeev & Sergeev, 2023) facilitates green innovation in robot transportation via IoT (Zadorozhnyi et al., 2024), and enhances enterprise competitiveness via modernization (Babakhanova et al., 2024). Finally, bibliometric analyses by Stoica and Ionescu-Feleagă (2024) and Zainuddin et al. (2025) reveal growing research interest in digital accountability but caution that cybersecurity risks and continuous training remain critical concerns for successful implementation.

3 Methodology

This paper primarily aims to review the literature published up to March 16th, focusing on research related to the digitalization of accounting. Desk research was made to achieve this goal, accompanied by bibliometric analysis on secondary data gathered from the Scopus and WoS core collection databases.

The bibliometric analysis process followed these stages:

1. **Research design** – While exploring the field of accounting, the author frequently discovered the concept of digitalization, which has emerged as a new term in today's business environment. Upon examining existing literature, it became evident that bibliometric analysis papers on this topic were insufficient. Recognizing this gap, the author conducted a bibliometric analysis of relevant studies available in both databases.

2. **Defining the search query** – The author selected primary keywords, "digitalization" and "accounting" and explored various combinations of these terms. The finalized search query for Scopus and WoS databases included: *"digitalization of accounting*" OR "digitalization in accounting*" OR "accounting digitalization*" OR "digitalization accounting*".*
3. **Research results in both databases** – The initial search query in Scopus resulted with total of 64 papers across all research areas until March 16th. After applying three inclusion criteria: language (English – 63 papers), publication years (2021-2025 – 60 papers) and subject area (1) Business, Management and Accounting; (2) Economics, Econometrics and Finance, 40 papers were selected. Similarly, the WoS database initially contained 28 papers across all fields. After filtering by language (English - 26 papers), publication years (2021-2024 - 24 papers), and Web of Science categories (1) Business Finance; (2) Business; (3) Economics; (4) Management 18 papers were identified.
4. **Refining initial search results** – All 58 papers were examined to ensure they specifically addressed the digitalization of accounting. Through detailed reviews of titles, abstracts, and main texts, 52 papers met the criteria.
5. **Identifying and removing duplicates** – During the initial review, 12 duplicate papers were found, appearing in both databases. After eliminating duplicates, 40 unique papers remained as the source for data analysis.
6. **Data analysis** – The final stage involved analyzing the data using descriptive statistics and performing bibliometric mapping with VOSviewer software (version 1.6.10). This step visualized results such as networks of co-occurrences, co-citations, and bibliographic matches among the papers in Scopus.

4 Research results

4.1 Results of descriptive statistics

The main goal of this paper is to give a detailed review of the literature published until March 16th, 2025 within both databases focusing on digitalization of accounting. The paper achieved its goal through conducted desk research. From 2021 to 2025, all 40 papers found in Scopus and WoS were published in English. These papers involved 122 authors from 11 countries and represented 86 affiliations, highlighting significant international collaboration. The publications included 22 original research articles, 10 book chapters, 3 conference papers, 3 review articles, and 2 proceeding papers. The year with the most citations was 2022, totaling 195 citations, while the highest number of publications occurred in 2024 with 14 papers, reflecting growing interest in this field. The most frequently cited paper was *"Digitalization, accounting and accountability: A literature review and reflections on future research in public services"* written by Agostino, Saliterer, and Steccolini, which was cited 117 times between 2022 and 2025. Geographically, Romania led with seven publications, followed by Ukraine and Malaysia with four each. Other notable contributors included India, the United States, Germany, and Vietnam with three papers each. Several other countries, including Indonesia, Italy, Kazakhstan, and Tunisia, have 2 papers published, while Bahrain, Belgium, China, Cyprus, France, Ghana, Jordan, Moldova, the Netherlands, New Zealand, Oman, Poland, Russia, Saudi Arabia, the United Arab Emirates, the United Kingdom, and Uzbekistan each published one paper.

The findings reveal a broad and diverse global interest in the topic, which is crucial for understanding the impact of digitalization of accounting and accountability processes, especially in the public sector. An examination of citation counts reveals that 2022 saw the highest number of citations, indicating rising interest at that time. However, the trend continued with a notable increase in publications in 2024, featuring a record 14 new papers. This sustained growth is likely driven by rapid digital advancements in the public sector, increased investment in technologies promoting transparency and accountability, and heightened research interest fueled by ongoing global digitalization. The geographical distribution of publications highlights a wide range of contributing countries, with significant contributions from Eastern Europe and Asia, as well as major Western nations like the United States, Germany and France. This diversity underscores the universal relevance of digitalization, accountability, and accounting across different economic, political, and cultural contexts. Furthermore, the steady increase in citations per paper reflects growing recognition and influence within the industry.

Overall, these statistics demonstrate the rapidly expanding and evolving nature of research on digitalization, accounting and accountability, particularly in public services. They also highlight how these concerns are becoming integral to modernizing public administrations worldwide.Table 1. shows the most important findings of the ten most cited papers in the Scopus and WoS CC database on the topic of digitalization of accounting. The distribution of the number of all published papers in both databases, as well as the total number of citations in the period from 2021 to 2025 is shown graphically in Fig. 1.

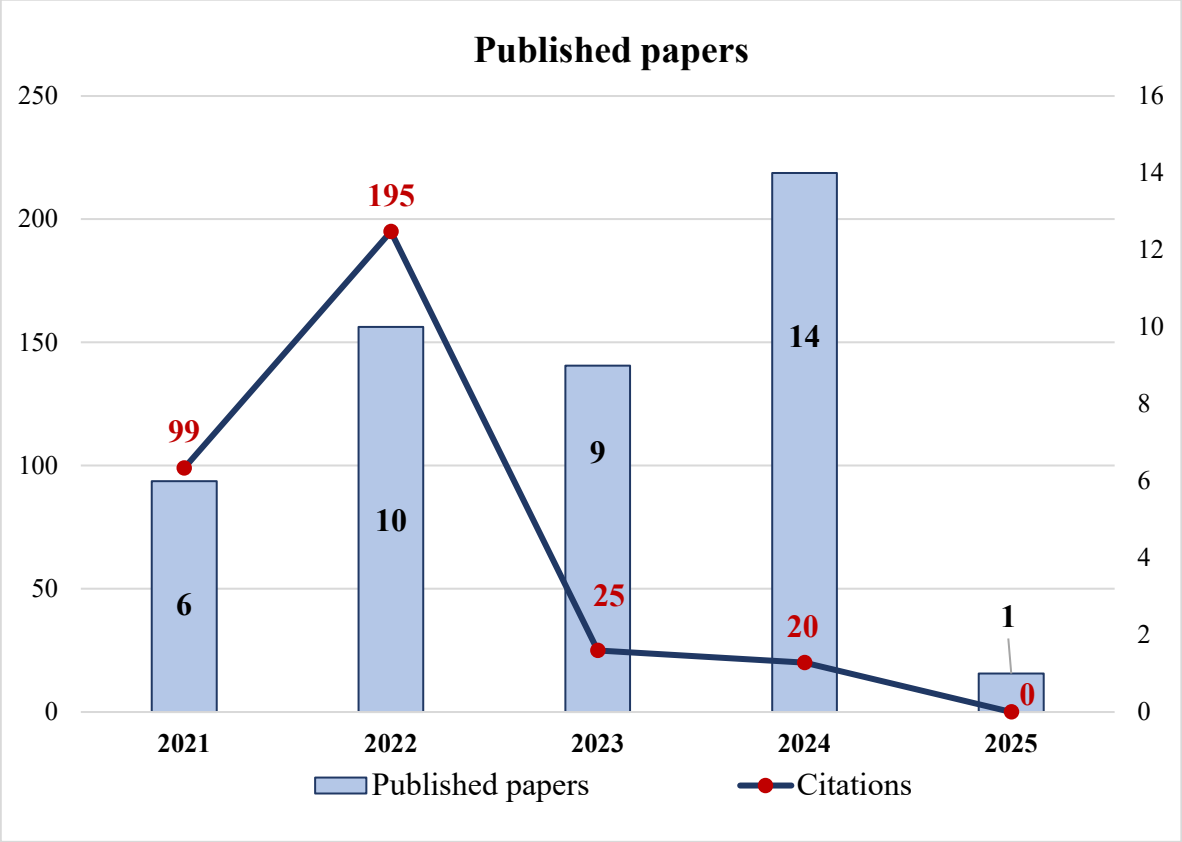


Figure 1. Distribution of the number of published papers and citations by year (Source: Authors' work)

Table 1. The most important findings of the ten most cited papers in the Scopus and WoS CC database on the topic of digitalization of accounting

PAPER	AUTHOR(S)	CITATIONS (YEAR)	FINDINGS
1. Digitalization, accounting and accountability: A literature review and reflections on future research in public services	Agostino D.; Saliterer I.; Steccolini I.	117 (2022)	Digital tools enable dialogic and horizontal accountability, but at the same time, they can increase bureaucratic barriers and responsibility overload.
2. Accountant as digital innovator: Roles and competencies in the age of automation	Kokina J.; Gilleran R.; Blanchette S.; Stoddard D.	67 (2021)	Robotic Process Automation (RPA) transforms the roles and competencies of accountants in the digital age.
3. Fintech accounting and industry 4.0: future-proofing or threats to the accounting profession?	Fülöp, MT; Topor, DI; Ionescu, CA; Capusneanu, S; Breaz, TO; Stanescu, SG	32 (2022)	The digitalization of financial services brings significant benefits, such as the automation of routine tasks and increased accuracy, but it also presents challenges like job loss and the need for new skills.
4. Digitalization as a Vector of Information Systems Development and Accounting System Modernization	Chyzhevska L.; Voloschuk L.; Shatskova L.; Sokolenko L.	17 (2021)	Digitalization causes structural changes in economic sectors, increases company productivity, and requires the modernization of systems and mechanisms to improve competitiveness and management.
5. Testing accountants' perceptions of the digitization of the profession and profiling the future professional	Grosu V.; Cosmulese C.G.; Socoliuc M.; Ciubotariu M.-S.; Mihaila S.	17 (2023)	Digitalization brings significant changes to the accounting profession, requiring continuous improvement and adaptation of accountants to ensure the relevance of the profession in the digital age.
6. Implications of the “momentum” theory of digitalization in accounting: Evidence from Ash Cloud	Zhang M.; Ye T.; Jia L.	10 (2022)	The development of the "momentum" theory of digitalization in accounting (DIA) emphasizes that the characteristics of companies and leaders are key drivers of digital transformation.
7. Analysis of trends in the implementation of digitalization in accounting (Ukrainian case)	Petchenko M.; Fomina T.; Balaziuk O.; Smirnova N.; Luhova O.	9 (2023)	Digitalization enables the improvement of accounting process efficiency, including automation, data transfer accuracy, and remote access to information.
8. Digitalization in accounting-Warmly embraced or coldly ignored?	Jans, M; Aysolmaz, B; Corten, M; Joshi, A; Peteghem, M	9 (2022)	The digitalization of accounting enables process automation, increases accuracy and transparency, and facilitates decision-making based on data.
9. Digitalization in small and medium enterprise: a parsimonious model of digitalization of accounting information for sustainable innovation ecosystem value generation	Pham, QH; Vu, KP	9 (2022)	The digitalization of accounting information (DOAI) improves the quality of financial reports, increases the effectiveness of business decisions, and promotes sustainable innovation practices.
10. No longer second-class citizens: Redefining organizational identity as a response to digitalization in accounting shared services	Klimkeit D.; Reihlen M.	8 (2022)	Robotic Process Automation (RPA) transforms the organizational identity of accounting companies, enabling greater relevance and a strategic role for companies in business.

Source: Authors’s work based on Scopus and Clarivate Web of Science, March 16th, 2

4.2 Visualization of bibliometric analysis

This section outlines the bibliometric analysis, focusing on keyword co-occurrence, bibliographic coupling of documents to explore the connections between published papers and their authors, and citation metrics in the Scopus database. The decision to exclude papers from the Web of Science database was based on the limited number of unique papers available, only six of them, which seemed to be insufficient for meaningful analysis.

The bibliometric analysis on the topic of digitalization in accounting identified a total of ten keywords that appeared at least twice in the Scopus database. These keywords were grouped into four distinct clusters, each represented by a different color (red, green, blue, and yellow), comprising 17 links with a total link strength of 17 (Fig. 2).

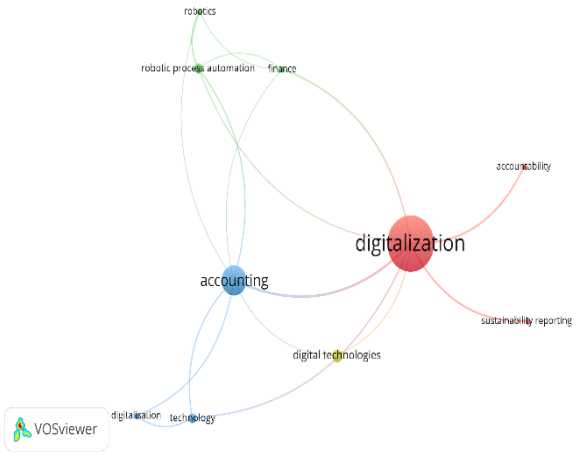


Figure 2. Visualization map of keywords co-occurrence (Source: Author's work)

The first cluster, highlighted in red, consisted of three keywords: accountability, digitalization, and sustainability reporting. The second cluster, represented by green, also contained three keywords: finance, robotic process automation, and robotics. The third cluster, marked in blue, encompassed three keywords: accounting, digitalisation, and technology. Lastly, the fourth cluster, colored in yellow, included only one keyword: digital technologies. Among these, the most frequently occurring keywords, which were pivotal to the search query, were "digitalization"/ "digitalisation" (two spellings of the same term, found in both cluster 1 and cluster 3) with 19 occurrences, 9 links, and a total link strength of 12, as well as "accounting," which had 9 occurrences, 7 links, and a total link strength of 7.

Following these, "digital technologies" emerged as the third most frequent keyword, occurring 4 times with 1 link and a total link strength of 1. Ranking fourth, with an occurrence level of 3, were the keywords "robotic process automation" (4 links, total link strength: 3) and "technology" (3 links, total link strength: 3). The fifth position was occupied by four keywords, each with an occurrence level of 2: "finance" (4 links, total link strength: 2), "robotics" (3 links, total link strength: 2), "accountability" (1 link, total link strength: 2), and "sustainability reporting" (1 link, total link strength: 2).

From Fig. 3, it can be seen that the available papers are grouped into six distinct clusters. The first cluster, highlighted in red, contains five papers, all published between 2022 and 2024. The second cluster, marked in green, includes four papers, also published within the same timeframe (2022–2024). The third cluster, shown in blue, consists of four papers. The fourth cluster, colored yellow, comprises three papers and features the oldest paper in the dataset: "*Accountant as digital innovator: Roles and competencies in the age of automation,"* written by Kokina, Gilleran, Blanchette, and Stoddard in 2021. This paper is among the top five most cited papers in this research area, with 67 citations. The fifth cluster, represented in purple, contains three papers, while the sixth cluster, shaded light blue, includes two papers.

This visualization underscores the significance of two key papers as foundational works for subsequent research on digitalization in accounting. The first is "*Digitalization, accounting and accountability: A literature review and reflections on future research in public services"* authored by Agostino, Saliterer, and Steccolini in 2022, which has received 117 citations. The second is "*Accountant as digital innovator: Roles and competencies in the age of automation"* by Kokina et al., published in 2021. Both papers serve as pivotal references for this field of study.

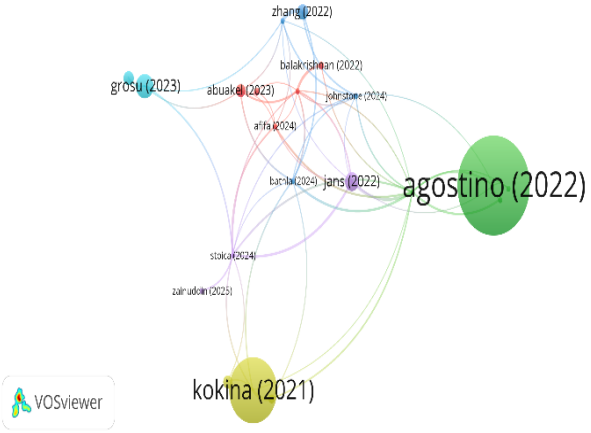


Figure 3. Visualization map of bibliographic coupling of documents (fractional counting) (Source: Author's work)

4.3 Research discussion and limitations

The role of digitalization of accounting is undeniably significant in today's digital business environment.

This is evident both from the theoretical framework presented in this paper and the bibliometric analysis conducted to achieve the research goal. Following all stages of the research process, as outlined in the methodological section, it is clear that the frequent occurrence of keywords highlights a strong and critical connection between digitalization and accounting. Based on studies published since 2021, it can be inferred that interest in this field is steadily increasing. This growth may be attributed to advancements in digital technology during the Covid-19 pandemic lockdown or possibly other factors, which time will reveal. At present, only a limited number of researchers, institutions, and countries are actively exploring this area. However, it is anticipated that this will change as interest continues to grow.

Although this study contributes valuable insights into the digitalization of accounting, it is important to emphasize that this paper only summarized existing literature. This paper could serve as a solid foundation for future researches in which a new concept or theoretical perspective could be defined by expanding research process on other relevant databases and on other research field in upcoming years. The integration of legal, HR, and AI-related perspectives into the bibliometric review creates an interdisciplinary foundation for further research. Future studies could employ multiparameter analytical tools such as cluster analysis and factor analysis to investigate the interdependence of digitalization drivers, regulatory frameworks, workforce readiness, and technological adoption levels. This would allow a more granular and multidimensional un understanding of digital transformation in accounting.

Nevertheless, the conducted research has certain limitations. Firstly, the fundamental limitation lies in the fact that the bibliometric analysis relies on only two databases, Scopus and WoS. While both databases are respectable, future studies should broaden their scope by incorporating additional scientific databases. The limitations of this research also include the search query used, as it inherently constrained the authors. Based on the initial query, which identified a significant number of papers, 92 in total (64 from Scopus and 28 from WoS), the author established criteria deemed most suitable for the topic of the paper, namely the digitalization of accounting. While this approach may have led to a reduced number of papers available for bibliometric analysis, which can be considered a limitation, a thorough review of the 58 selected papers confirmed that the criteria were appropriate. Ultimately, 40 unique papers were found to align with the researched topic.

5 Conclusion

The paper emphasizes the growing importance of digitalization in accounting, particularly for enhancing automation and company performance. Despite its advancements, digitalization in accounting has yet to fully realize its potential due to several challenges and limitations. However, it has opened up significant opportunities for research and development in this field, which is the main focus of the study. Digitalization in accounting may increase the future importance of related fields like digital advertising. The study offers a theoretical overview of accounting digitalization and presents findings from desk research using Scopus and WoS databases. Through bibliometric analysis of 40 publications, the research highlights a strong and significant connection between digitalization and accounting, as evidenced by keyword occurrences. This outcome aligns with predictions made in the theoretical overview for the period from 2021 to 2025, indicating a likely increase in scientific publications on this topic. The rigorous six-stage data preparation ensured the bibliometric analysis's success, highlighting the study's reliability and the promising future of digital accounting research.

References

AbuAkel, S. A., & Ibrahim, M. (2023). The Effect of Relative Advantage, Top Management Support and IT Infrastructure on E-Filing Adoption. Journal of Risk and Financial Management, 16(6). Scopus. https://doi.org/10.3390/jrfm16060295

Afifa, M. A., Van, H. V., & Van, T. L. H. (2024). Digitalisation in accounting according to accountants' perception: An extended TAM model. International Journal of Business Innovation and Research, 34(1), 64–89. Scopus. https://doi.org/10.1504/IJBIR.2024.138140

Agostino, D., Saliterer, I., & Steccolini, I. (2022). Digitalization, accounting and accountability: A literature review and reflections on future research in public services. FINANCIAL ACCOUNTABILITY & MANAGEMENT, 38(2), 152–176. https://doi.org/10.1111/faam.12301

Arfaoui, F., & Kammoun, I. (2023). Did accounting education remain resistant to digitalization during COVID-19? An exploratory study in the Tunisian context. Journal of Accounting Education, 65. Scopus. https://doi.org/10.1016/j.jaccedu.2023.100874

Arslan, M. (2024). Digitalisation in accounting: Is it a risk or an opportunity? In Impact of Digitalization on Reporting, Tax Avoidance, Account., and Green Finance (pp. 112–133). IGI Global; Scopus. https://doi.org/10.4018/9798369316788.ch005

Babakhanova, N. U., Sultanova, S. M., Djumanova, A. B., Kodirov, M. K., & Seytbekova, S. T. (2024). IMPROVING ACCOUNTING PRACTICES TO INCREASE THE COMPETITIVENESS OF ENTERPRISES. In Adv. Ser. Manage. (Vol. 33A, pp. 221–228). Emerald Publishing; Scopus.

https://doi.org/10.1108/S1877-63612024000033A026

Baidybekova, S. K., Shomshekova, B. K., Kydyrbayeva, E. O., Sharipov, A. K., & Kasseinova, M. I. (2021). Accounting and financial reporting in the context of globalisation and digitalisation: Problems and improvements. Public Policy and Administration, 20(2), 248–259. Scopus. https://doi.org/10.13165/VPA-21-20-2-08

Balakrishnan, C., & Rahman, H. (2022). Nexus of Digitalization, Social, and Governance Factors. In Contrib. Finance. Account.: Vol. Part F230 (pp. 181–200). Springer Nature; Scopus. https://doi.org/10.1007/978-3-031-11545-5_10

Bathla, S., Himanshu, & Ahmad, F. (2024). Digital disruption: Unveiling antecedents propelling the revolution in contemporary accounting practices. In Journal of Accounting and Organizational Change. Emerald Publishing. https://doi.org/10.1108/JAOC-02-2024-0051

Ben Letaifa, W., & Hadj Salem, W. B. (2023). The Digitalization of Accounting Firms: Survey Research from Tunisia. IBIMA Bus. Rev., 2023. Scopus. https://doi.org/10.5171/2023.550358

Chyzhevska, L., Voloschuk, L., Shatskova, L., & Sokolenko, L. (2021). Digitalization as a Vector of Information Systems Development and Accounting System Modernization. In Studia Universitatis Vasile Goldis Arad, Economics Series (Vol. 31, Issue 4, pp. 18–39). De Gruyter Open Ltd. https://doi.org/10.2478/sues-2021-0017

Fulop, M. T., Topor, D. I., Ionescu, C. A., Capusneanu, S., Breaz, T. O., & Stanescu, S. G. (2022). FINTECH ACCOUNTING AND INDUSTRY 4.0: FUTURE-PROOFING OR THREATS TO THE ACCOUNTING PROFESSION? In JOURNAL OF BUSINESS ECONOMICS AND MANAGEMENT (Vol. 23, Issue 5, pp. 997–1015). VILNIUS GEDIMINAS TECH UNIV. https://doi.org/10.3846/jbem.2022.17695

Grosu, V., Cosmulese, C. G., Socoliuc, M., Ciubotariu, M.-S., & Mihaila, S. (2023). Testing accountants' perceptions of the digitization of the profession and profiling the future professional. Technological Forecasting and Social Change, 193. Scopus. https://doi.org/10.1016/j.techfore.2023.122630

Hrebenyk, T., Ivashchenko, M., & Shcherbyna, O. (2023). Management of an organisation's accounting department: Overcoming the challenges of martial law in Ukraine. Scientific Bulletin of Mukachevo State University. Series Economics, 10(3), 9–21. Scopus. https://doi.org/10.52566/msu-econ3.2023.09

Jabor, L., & Hamdan, A. (2023). Digitalization in Accounting and Financial Reporting Quality: Literature Review. In Contrib. Manag. Sci.: Vol. Part F1640 (pp. 793–801). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-981-99-6101-6_59

Jans, M., Aysolmaz, B., Corten, M., Joshi, A., & van Peteghem, M. (2022). Digitalization in accounting–Warmly embraced or coldly ignored? Accounting, Auditing and Accountability Journal, 36(9), 61–85. Scopus. https://doi.org/10.1108/AAAJ-11-2020-4998

Jassem, S. (2024). Artificial Intelligence in Accounting Practices in the Industry 5.0 Era from a Dynamic Capabilities Perspective: Role of Strategic Foresight, Agility, and Flexibility. In Stud. Syst. Decis. Control (Vol. 545, pp. 149–160). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-031-65203-5_14

Kartikasari, E. D., Zainuddin, S. A. B., & Jamil, A. B. (2024). Transformation of Accounting Information Systems in the Management of BUMDesa Financial Performance in the Digitalization Era. In CSR. Sustain. Ethic. Gov.: Vol. Part F3302 (pp. 537–543). Springer Nature; Scopus. https://doi.org/10.1007/978-3-031-53877-3_42

Klimkeit, D., & Reihlen, M. (2022). No longer second-class citizens: Redefining organizational identity as a response to digitalization in accounting shared services. Journal of Professions and Organization, 9(1), 115–138. Scopus. https://doi.org/10.1093/jpo/joac003

Kokina, J., Gilleran, R., Blanchette, S., & Stoddard, D. (2021). Accountant as digital innovator: Roles and competencies in the age of automation. Accounting Horizons, 35(1), 153–184. Scopus. https://doi.org/10.2308/HORIZONS-19-145

Langmann, C., & Kokina, J. (2021). RPA in accounting. In Robot. Process Automation: Manag., Technol., Appl. (pp. 243–261). De Gruyter; Scopus. https://doi.org/10.1515/9783110676693-013

Marioara, I., Valentin, B., Delia, D., & Amalia, N. Ş. (2022). Perception of Students and Master Students from the Western Part of Romania Over the Digitalization Process in the Accounting Education. Studies in Business and Economics, 17(1), 52–72. Scopus. https://doi.org/10.2478/sbe-2022-0004

Oleiwi, R. (2023). THE EXTENT TO WHICH TEXTBOOKS FULFILL THE REQUIREMENTS OF DIGITAL TRANSFORMATION IN ACCOUNTING AND AUDITING. International Journal of Professional Business Review, 8(5). Scopus. https://doi.org/10.26668/businessreview/2023.v8i5.1509

Petchenko, M., Fomina, T., Balaziuk, O., Smirnova, N., & Luhova, O. (2023). ANALYSIS OF TRENDS IN THE IMPLEMENTATION OF DIGITALIZATION IN ACCOUNTING (UKRAINIAN CASE). Financial and Credit Activity: Problems of Theory and Practice, 1(48), 105–113. Scopus. https://doi.org/10.55643/fcaptp.1.48.2023.3951

Pham, Q. H., & Vu, K. P. (2022). Digitalization in small and medium enterprise: A parsimonious model of digitalization of accounting information for sustainable innovation ecosystem value generation. In ASIA PACIFIC JOURNAL OF INNOVATION AND ENTREPRENEURSHIP (Vol. 16, Issue 1, pp. 2–37). EMERALD GROUP PUBLISHING LTD. https://doi.org/10.1108/APJIE-02-2022-0013

Puspitawati, L., Shaffira, N., & Bansah, P. F. (2024). Skills and Functioning Departmentalization in Improving The Effectiveness of Digitalization of Accounting Systems. Australasian Accounting, Business and Finance Journal, 18(5 Special Issue), 87–106. Scopus. https://doi.org/10.14453/aabfj.v18i5.06

Quang Huy, P., & Kien Phuc, V. (2024). Digitalization accounting information system and servitization in public sector organizations—Moderating effect of recoverable slack. Cogent Business and Management, 11(1). Scopus. https://doi.org/10.1080/23311975.2024.2420762

Remlein, M., Jastrzębowski, A., & Obrzeżgiewicz, D. (2022). The impact of legislation on the development of accounting digitisation in Poland. Zeszyty Teoretyczne Rachunkowosci, 46(4), 115–127. Scopus. https://doi.org/10.5604/01.3001.0016.1305

Sergeev, L. I., & Sergeev, D. L. (2023). Digitalization in the Processes of Classification of the Country's Budget Expenditures. Finance: Theory and Practice, 27(1), 195–207. Scopus. https://doi.org/10.26794/2587-5671-2023-27-1-195-207

Shuhidan, S. M., Awang, Y., Taib, A., Rashid, N., & Hasan, M. S. (2023). Technology Readiness Among Future Accountants Towards Digitalization of Accounting Profession. In INDONESIAN JOURNAL OF SUSTAINABILITY ACCOUNTING AND MANAGEMENT (Vol. 7, Issue 1, pp. 1–12). UNIV PASUNDAN. https://doi.org/10.28992/ijsam.v7s1.878

Stoica, O. C., & Ionescu-Feleaga, L. (2021a). Digitalization in Accounting: A Structured Literature Review. In A. Dima (Ed.), RESILIENCE AND ECONOMIC INTELLIGENCE THROUGH DIGITALIZATION AND BIG DATA ANALYTICS (pp. 453–464). SCIENDO. https://doi.org/10.2478/9788366675704-045

Stoica, O. C., & Ionescu-Feleaga, L. (2021b). The accounting practitioner as a driver of digitalization pace. In PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON BUSINESS EXCELLENCE (Vol. 15, Issue 1, pp. 768–782). SCIENDO. https://doi.org/10.2478/picbe-2021-0072

Stoica, O. C., & Ionescu-Feleagă, L. (2024). A Bibliometric Analysis Regarding the Digitalization of the Accounting Profession. In Dima A.M. & Vâlcea S. (Eds.), Springer Proc. Bus. Econ. (pp. 11–20). Springer Nature; Scopus. https://doi.org/10.1007/978-3-031-58967-6_2

Valentinetti, D., & Rea, M. A. (2024). Factors influencing the digitalization of sustainability accounting, reporting and disclosure: A systematic literature review. Meditari Accountancy Research. Scopus. https://doi.org/10.1108/MEDAR-02-2024-2385

Zadorozhnyi, Z.-M., Muravskyi, V., Shevchuk, O., Muravskyi, V., & Zadorozhnyi, M. (2024). Digitization of Accounting in the Innovative Management of Autonomous Robotic Transport. In MARKETING AND MANAGEMENT OF INNOVATIONS (Vol. 15, Issue 3, pp. 110–126). SUMY STATE UNIV, DEPT MARKETING & MANAGEMENT INNOVATIVE ACTIVITY. https://doi.org/10.21272/mmi.2024.3-09

Zainuddin, S. A., AlJahsh, M. A. I., Abdullah, T., Nasir, N. A. M., Zainol, N. R., Nawi, N. C., Ahyaruddin, M., & Yiannaki, S. M. (2025). Global Research Trends in Digitalization and Accountability: A Bibliometric Analysis. In Stud. Syst. Decis. Control (Vol. 568, pp. 711–721). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-031-71526-6_62

Zhang, L., & Ren, J. (2022). Virtual Influencers: The Effects of Controlling Entity, Appearance Realism and Product Type on Advertising Effect. In G. Salvendy & J. Wei (Eds.), DESIGN, OPERATION AND EVALUATION OF MOBILE COMMUNICATIONS, MOBILE 2022 (Vol. 13337, pp. 298–305). SPRINGER INTERNATIONAL PUBLISHING AG. https://doi.org/10.1007/978-3-031-05014-5_25

Protecting privacy on social media platforms: a case study of Croatian users

Anamarija Horvat, Renata Mekovec
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
ahorvat21@student.foi.unizg.hr, renata.mekovec@foi.unizg.hr

Abstract. *This study identifies three areas of research that are investigated in relation to privacy protection when using social media platform: (1) technical area, (2) area related to the user and (3) area related to privacy models. A survey was done to find out how much privacy awareness Croatian users of social media platform have. The results show that a lot of users of social media platforms share their first and last names, photos and videos, and their birth date. The research uncovered also significant security gaps like a limited awareness of basic security measures and a significant lack of awareness about the rights users have in the context of privacy protection.*

Keywords. privacy concerns, social media platform, digital threats

1 Introduction

Billions of people utilize major social media platforms such as Facebook, Instagram, TikTok, Snapchat, YouTube, Twitter, and LinkedIn. A total of 5.52 billion people were using the internet at the start of October 2024, where number of internet users have increased by 151 million (+2.8 percent) over the past 12 months. According to Kepios's latest analysis, the number of social media platform users is equating to 63.8 % of all the people on Earth (Kemp, 2024).

Social media platform's phenomenal expansion has provided their owners with unprecedented access to and control over users' lives. Social media platforms collect information like personal data, demographic information, interests, behaviors, and activities elsewhere on the internet (Federal Trade Commission, 2024). Some of this information is willingly supplied via posts and profile information. Information also may be released unknowingly through tracking cookies, which track a user's online activities such as webpage visits, social media platform sharing, and purchase history. All information is then collected and organized into user categories, which data brokers sell for marketing purposes.

On other hand it should be stressed that social media platforms have drastically changed perceptions about privacy. While technology allows self-expression and virtual interactions, it also introduces hazards such as profiling, targeted advertising, and mass monitoring. It is true that cyberbullying, and data breaches have a significant impact on confidence in the digital economy. Trust in online services, and social media platforms declines when people suffer from data breaches or are the targets of cyberbullying and online bullying. People may be less inclined to consume digital goods, reveal personal information, or participate in online activities as a result of this diminished confidence (ENISA, 2024).

The combination of AI, analytics, behavioral science, social media platform, Internet of Things (IoT) and other technologies makes an incredible opportunity for bad actors to create and spread highly effective, mass-customized disinformation, also known as malinformation (Plummer et al., 2023). Americans have little faith that social media platform executives will responsibly handle user privacy. Approximately 77% of Americans don't trust social media platform executives to speak up with their mistakes and accept accountability for misuse of their platforms. Furthermore, they have lost faith in the government's capacity to control them: 71% of people don't think the government will hold these tech CEOs responsible for their mistakes. Although the majority of Americans (85%) believe that parents bear a major portion of the duty for safeguarding their children's online privacy, they also believe that tech corporations (59%) and the government (46%) bear this obligation (Pew Research Center, 2023). Only half Europeans believe their digital rights are well secured, over a third feel the EU isn't doing enough, and 45% are unsatisfied with the protection of children in the digital environment. The lowest percentage, at just 40%, believes they have full control over their own data and digital legacy. Older respondents are more likely to believe the EU does not protect their rights adequately in the online environment. Those aged 40-54 are the most likely (38%) to think this, followed by those aged 55 and up (36%). Younger respondents (15-24) are less likely to support this opinion (32%). For example, 73% of those aged 15-24 believe that getting more freedom of expression and information online e.g., via online

platforms, social networks, search engines is well protected in their country (European Union, 2023).

While some experts stated that people are extremely worried about their online privacy, others suggested that, while being sceptical of social media platform practices, users were willing to reveal personal information in exchange for tiny advantages (Khan et al., 2023). The "privacy paradox" phenomenon proposes that users of social media platforms who are highly concerned about their privacy do not always apply their concerns to their usage practices. In other words, despite their high level of privacy concerns, social media platform users continue to utilize the platform in their daily lives (Hew et al., 2019).

The challenge of protecting privacy when utilizing social media platforms is one that both researchers and the IT sector struggle with. As stated by TechTarget (2024) common social media platform privacy issues include: (1) data mining for identity theft, (2) vulnerabilities in privacy settings (loopholes), (3) location settings, (4) harassment and cyberbullying, (5) fake information, and (6) malware and viruses.

The research described in this paper aims to better understand the current perspective on privacy protection on social media platforms, from an academic and social media platform user's standpoint.

Therefore, this research is to address following research questions:

1. what is the latest research on protecting privacy on social media platforms?
2. how can privacy be protected on social media platforms?
3. what is the level of privacy awareness among Croatian social media platform users?

The primary contribution of this work is to identify most recent research on privacy protection when using social media platforms in order to use this knowledge to improve user privacy awareness. This will be performed by reviewing the most recent academic results, investigating practical privacy protection methods, and analysing the level of privacy awareness among Croatian social media platform users.

2 Recent studies regarding privacy on social media platforms

Researchers concentrate on various privacy areas, privacy issues, and privacy protection strategies used by social media platform users. Thus, they focus on fatigue due to social media platform (Dhir et al., 2019), usage intention (X. Fan et al., 2021), (Neves et al., 2024), (Garima & Sheokand, 2024), users' lurking behavior (Liu et al., 2024), most important preferences of users regarding social media platform apps (Alshehri & Alamri, 2022), (Bracamonte et al., 2024), (Tang & Ning, 2023), the effectiveness of social media platform advertising (Jung & Heo, 2024), privacy differences across various regions (Farooq et al., 2024), (Meso et al., 2021), (Nuzulita & Subriadi, 2020).

The aim of this study was to find the most important studies on privacy challenges when using social media platform and identify the main research area addressed by the aforementioned studies.

2.1 Review methodology

The research articles included in this study were published between 2019 and 2024 and retrieved from Scopus and Web of Science. These databases were chosen because they contain the most recent full-text peer-reviewed publications, have advanced search capabilities, and cover the most published research papers. The following keywords were used in a keywords-based search method: "social media" AND "privacy concern". The quantity of citations of the publications was taken into account while selecting important papers, and some of the most widely referenced articles in our study (Table 1) were selected as seed papers. In the second step, a graph-based search strategy was utilized to locate more relevant papers, with papers found during our initial search serving as seeds. To make the graph more focused on subject, we begin adding articles that we find important. Citation analysis tools Inciteful (https://inciteful.xyz/) was used to create citation network graphs, and to explore and add most important recent related papers to the database. In total 204 items in all were gathered in this manner.

Table 1. Most widely referenced articles in our study – seed papers

Title	Cited
The age of mobile social commerce: An Artificial Neural Network analysis on its …(Hew et al., 2019)	199
Antecedents and consequences of social media fatigue (Dhir et al., 2019)	190
Privacy concerns and benefits of engagement with social media-enabled apps: ….(Jozani et al., 2020)	170
Effect of penitence on social media trust and privacy concerns: The case of … (Ayaburi & Treku, 2020)	119
The role of trust and privacy concerns in using social media for e-retail services: …(Alzaidi & Agag, 2022)	82
Does role conflict influence discontinuous usage intentions? Privacy concerns.. (X. J. Fan et al., 2021)	49
Social media users' online subjective well-being and fatigue: A network … (Kaur et al., 2021)	47

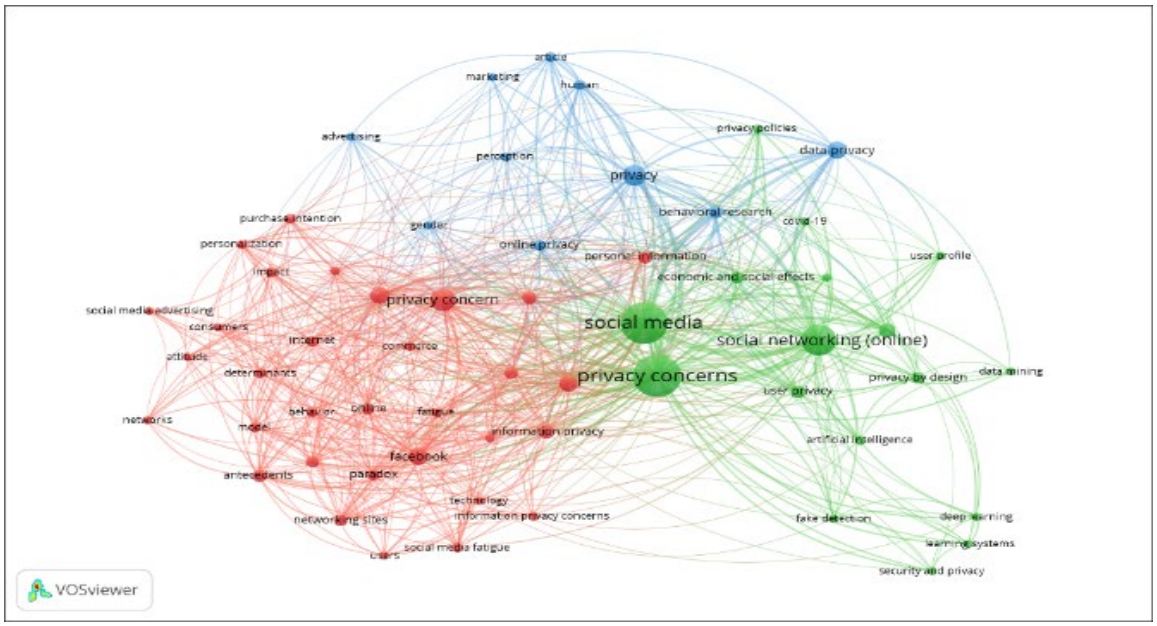


Figure 1. Keyword co-occurrence network map

Fig. 1 displays a keyword co-occurrence map of selected articles (204 articles/items identified in second step) derived by bibliometric analysis in VOSViewer. In the network visualization, items are represented by their label and by a circle. The size of the label and the circle of an item is determined by the weight of the item. The higher the weight of an item, the larger the label and the circle of the item is. The colour of an item is determined by the cluster to which the item belongs. Lines between items represent links.

On the visualization of related keywords, it can be seen the grouping of keywords around three clusters. In this manner it can be answered on research question: what is the latest research relevant to privacy on protecting privacy on social media platform? Considering the structure of keywords in each cluster, it can be defined that research focus is on: (1) technical area, (2) area related to the user and (3) area related to privacy models. The first cluster, the **technical area**, includes technology that is increasingly being used in all aspects of life, thus affecting social media platforms as well. It is thus evident that research related to privacy in the field includes issues related to artificial intelligence, data mining, learning systems, as well as fake detection and privacy by design. The second cluster, the **area related to the user**, is oriented towards users of social media platforms, so research in this area is focused on perception, behavioral research, gender and personal information. The third cluster, **the area related to privacy models**, includes research related to antecedents and determinants of privacy concerns, paradox, model and impact.

3 Methods for privacy protection on social media platform

In the context of increasing digital threats, privacy protection on social media platforms has become a crucial topic. This chapter presents key methods of protection that social media platform users can apply, along with an analysis of recommendations from relevant sources and experts.

3.1 Using complex and unique passwords

The complexity of passwords and avoiding predictable patterns are crucial for security. According to MacKay (MacKay, 2024), passwords that combine uppercase and lowercase letters, numbers and symbols and exceed 12 characters in length significantly reduce the risk of compromise. Passwords that include personal information such as names or birth dates should be avoided, as such data is easily accessible through social media platform, making the account more vulnerable to attacks.

3.2. Implementation of two – factor authentication (2FA)

Two – factor authentication adds an extra layer of protection. According to Counting Up (How to Maintain Privacy on Social Media | Countingup, 2022), the introduction of one – time login codes, in addition to regular passwords, reduces the risk of unauthorized access. This measure provides additional security because an attacker cannot gain access to the account without possessing the codes specific to the user's device, which significantly reduces the risk of hacking.

3.3. Adjusting privacy settings

Controlling privacy settings and regularly adjusting the visibility of personal data are essential protective measures. According to recommendations given by Data Privacy Manager (How to Protect Your Privacy on Social Media? – Data Privacy Manager), users should limit the visibility of their profiles to options such as „friends“ and „only me“ to reduce the risk of information misuse. This adjustment allows users greater control over who can see their information, thereby reducing the possibility of privacy compromise.

3.4. Access management: forced logout and login monitoring

Forced logout from devices and regular checking of active logins are important security step. MacKay (MacKay, 2024) emphasizes that control over active sessions allows users to detect unknown devices and quickly remove unauthorized users from the account. This measure ensures that users have insight into active devices, which improves access control and reduces the risk of account compromise.

3.5. Deleting old and inactive accounts

Old, inactive accounts often remain vulnerable to attacks because users rarely update their security settings. Bizga (2023) recommends deleting these accounts and requesting from platform administrators to remove all associated personal data from these accounts. By doing so, users reduce the „attack surface” since inactive accounts are often not protected by modern security measures.

3.6. Preventing excessive sharing of personal data

The European Commission (Komisija, n.d.) recommends refraining from sharing excessive personal data, including sensitive information such as exact locations and personal photos. This practice can prevent potential misuse, especially when it comes to sensitive data about locations and daily activities. Additionally, the Commission advises adopting a selective approach to accepting friend requests and engaging with unknown individuals to further reduce the risk of data misuse.

Establishing strong privacy protection on social media platforms require users to consistently adhere to security guidelines and utilize available tools. Regularly updating privacy settings, deleting old accounts, using complex passwords and implementing two – factor authentication significantly reduces the risk of personal data compromisation. These measures, combined with education about threats, such as phishing attacks and cyberbullying, are essential for enhancing security on social media platforms. In the broader context, digital security awareness and responsible digital behavior become crucial components in combating increasingly sophisticated cyber threats, thereby contributing to overall safety on online platforms.

4 Analysis of research results

As part of the preparation for the undergraduate thesis „Privacy protection on social media platform“ a research study was conducted to examine the security practices, privacy awareness and attitudes of Croatian social media platform users regarding digital threats and personal data protection (Horvat, 2024). The primary objective of the study was to: (1) assess the respondents’ level of awareness about security threats, (2) identify patterns in social media platform usage and protective measures and (3) analyse the ways personal data is shared on these platforms.

Questionnaire consist form 35 questions where answers were proposed. First group of questions were demography characteristic like gender, age and working status, usage of social media platform (awareness, daily usage in hours, device from which is social media platform accessed and what social media platforms is used by respondents). Second group of questions were connected to information that respondents are sharing via social media platform (Which types of information do you share on social media platforms?), with whom are they sharing information (Who are your followers on social media platforms, or with whom do you share information?), who are they follow (Whose friend requests/follow requests via social networks do you accept?) what information are they revealing (Whether and how often you publish your location on statuses, posts, videos or photos you post on social media platforms?). Third group of questions was connected to security measures that respondents use:

1. Do you use the same password for more social media platforms?
2. Are your passwords longer than 12 characters?
3. Do you use a combination of upper and lower case letters and a combination of numbers and/or characters and symbols (eg ?, !, ,, ., *, +, -) in your password for social media platforms?
4. Do you, and how often, connect different social media platforms through the same account (e.g. connecting Facebook and Instagram accounts)?
5. How old were you when opened your Facebook profile?

Fourth group of questions encompass following items:

1. Have you ever experienced a situation where your privacy was compromised when using social media platform, such as hacking profile or identity theft?
2. Are you familiar with the term GDPR (General Data Protection Regulation)?
3. Are you familiar with the term "Digital Fingerprint"?
4. Have you ever received a certain type of spam messages via social media platform (e.g., advertisements for products and services, messages that encourage certain religious or political opinion)?
5. Have you ever been a victim of "phishing", or online identity theft? (A person persuades a victim to reveal personal information via a link to a fake website and then uses the victim's information and poses as that person)
6. Have you ever been a victim of stalking via social media platform?
7. Have you ever been a victim of bullying via social media platform (e.g., insults, mockery, threats)?

Fifth group of questions encompass items connected to awareness of protecting during online activities:

1. Do you know what two-factor authentication is and do you use it when logging in to your social media platform?
2. Are you deleting profiles on social media platforms that you no longer use?
3. Do you know that only uninstalling certain social media platform applications will not delete your profile and posts that did you post?
4. Did you know that there are tools today that allow people to, by simply entering any photo you've posted on social media platform, find out when the photo was taken, where it was taken, and find the people in it?
5. Do you know how to set the visibility of personal data on social media platform posts and profiles to "only me" or "friends" and remove the option to find information about you using your email or social media platform link?
6. Do you know how to find devices that are logged into your profile and log out of those that don't belong to you?
7. Do you know that you have the right to ask the data controller a copy of the data collected about you, to whom all these data are sent, what is the purpose of this sending and request deletion of data after a certain period of time, if there is no reason for them to be detained?
8. Do you know what AZOP (Personal Data Protection Agency) is and what you can report to it?
9. Are you using any kind of antivirus protection?
10. Do you think that more should be said and people should be taught about privacy protection on social media platforms?

The research sample was made using snowball sampling method, and have consisted of Croatian users, providing insights into security practices within a specific cultural and legal framework. However, the conclusions were limited to the regional context, serving as a basis for comparison with global trends in digital security.

The sample included 140 respondents from the Republic of Croatia, of which 65% were women and 35% men. The respondents were predominantly young, with half of the total sample being aged between 18 and 25 years, 17.9% aged between 36 and 45 years, while the remaining portion of the sample was older than 46 years. The largest segment of respondents consisted of college students (47.1%), followed by employed individuals (37.9%), with high schoolers, unemployed and retirees being represented in smaller numbers (12.9%, 1.4% and 0.7% respectively). This structure provided insight into the behavioral patterns and attitudes of the population most active on social media platform in Croatia, with a particular focus on young users and their security and privacy practices.

The research revealed that 98.6% of respondents use social media platforms, while only 1.4% are aware of it but do not use it. The time spent on social media platform is also significant: 40.7% of respondents use social media platform between two and four hours daily, 24.3% from one to two hours, 23.6% spend between four and six hours on social media platform and the remaining 10% spend more than six hours on them daily, thus increasing their exposure to potential threats. Access to social media platform is predominantly through mobile devices (92.9%), while the use of laptops and desktop computers is minimal (3.6% and 2,1% respectively), highlighting the need for specific security measures tailored to mobile platforms.

The most frequently used social media platforms among respondents are WhatsApp (96.4%), YouTube (82.9%), Instagram (78.6%) and Facebook (68.6%), while platforms like X (Twitter) (12.9%) and Skype (9.3%) are less popular. This segment of the research indicated that the surveyed users are primarily focused on platforms that enable communication and multimedia content sharing.

The results show that 86.4% of respondents publicly share their first and last names on social media platform, 74.3% share photos and videos, while 58.6% share their birth date. Additionally, 36.4% of respondents share their phone number, increasing the

possibility of identity theft or contact misuse by malicious users. None of the respondents indicated sharing financial information or personal identification numbers, which suggests a basic level of awareness regarding the protection of sensitive information. However, it also highlights vulnerability due to sharing data such as phone numbers and birth dates.

The questions aimed at understanding privacy threats covered respondents' perceptions of risks such as hacking, cyberbullying and phishing attacks. The results show a high level of concern among respondents: 70% express worry about potential privacy breaches, while 50% are concerned about hacking and the misuse of personal data. Despite these concerns, the research uncovered significant security gaps; 43.6% of respondents use the same passwords across multiple platforms and 60.7% use passwords shorter than twelve characters, making them vulnerable to brute – force attacks and other forms of cyber threats. Only 68.6% of respondents use antivirus protection, indicating a limited awareness of basic security and privacy measures.

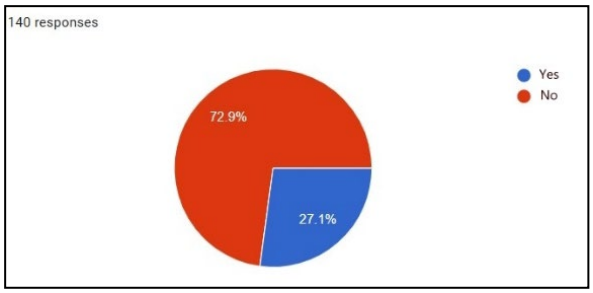


Figure 2. 5. Have you ever been a victim of "phishing", or online identity theft? (Horvat, 2024)

Questions related to the perception of threats, such as cyberbullying, profile hacking and phishing, indicate a limited understanding of these dangers. Fig. 2 shows results of question related to profile hacking or identity theft. Respondents answered that as many as 27.1% of them experienced profile hacking or identity theft, which proves that they should be educated about ways to protect against repeated attacks. Also, approximately 45% of respondents reported that they have personally experienced, or know someone who has experienced, cyberbullying, while 52% understand the danger of phishing attacks but are unsure how to recognize them. These results highlight the need for specific educational interventions aimed at recognizing and protecting against digital threats.

Questions regarding users' legal rights revealed a significant lack of awareness about the rights users have in the context of privacy protection. Specifically, 58.6% of respondents did not know they could request a copy of their data stored by social media platforms, while 56.4% were unaware of the role of the Croatian Personal Data Protection Agency (AZOP). This low level of awareness highlighted the need for additional informational campaigns to help users better understand their rights and the legal framework for privacy protection in the digital space.

Fig. 3 shows the results related to the final question of the research, specifically about their opinion on the importance of education and discussions about privacy protection on social media platform. As many as 96.4% of respondents believe that there should be more discussions and education about privacy protection on social media platform. This indicates that the respondents are aware of the need for greater public awareness of online privacy protection methods and that the research itself has prompted them to reassess their own privacy.

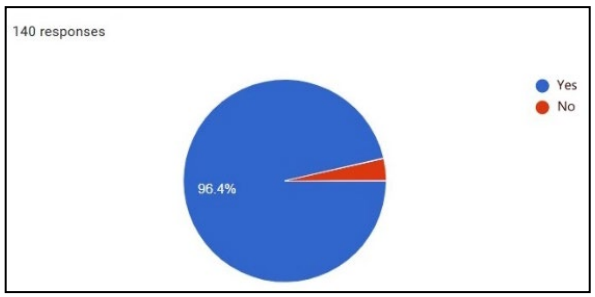


Figure 3. 10. Do you think that more should be said and people should be taught about privacy protection on social media platforms? (Horvat, 2024)

5 Conclusion

The overall results of performed research indicate a high level of social media platform usage among Croatian users, but reveal an insufficient level of applied security and privacy measures and an insufficient level of awareness of threats such as hacking, cyberbullying and phishing attacks. Despite concerns about security and privacy, a significant portion of respondents do not undertake basic security measures, such as using unique and complex passwords and multi – factor authentication. Additionally, the research reveals a low level of awareness regarding rights to access and control over personal data.

Given the identified vulnerabilities, it is recommended to introduce specific educational initiatives focused on basic security practices and user rights. Campaigns should include recognizing threats such as phishing attacks and cyberbullying, using complex and unique passwords, the importance of antivirus tools and the implementation of multi – factor authentication. Additionally, it is necessary to educate users about legal rights concerning data protection and privacy control, in collaboration with initiatives such

as AZOP, enabling users to better understand and apply security measures.

The results of this research, though limited to a national sample, provide insights into security attitudes that align with European trends. Younger generations exhibit high engagement on social media platform but limited awareness of advanced security practices. The findings of the research are an important contribution to understanding national attitudes towards digital security and privacy, highlighting the need for an interdisciplinary approach and international cooperation in promoting a culture of digital security on a global scale.

References

Alshehri, A., & Alamri, S. (2022). Exploring Social Media Privacy Preferences in Saudi Arabia. *International Journal Of Computer Science And Network Security*, *22*(1), 725–730. https://doi.org/10.22937/IJCSNS.2022.22.1.95

Alzaidi, M. S., & Agag, G. (2022). The role of trust and privacy concerns in using social media for e-retail services: The moderating role of COVID-19. *Journal Of Retailing And Consumer Services*, *68*. https://doi.org/10.1016/j.jretconser.2022.103042

Ayaburi, E. W., & Treku, D. N. (2020). Effect of penitence on social media trust and privacy concerns: The case of Facebook. *International Journal Of Information Management*, *50*, 171–181. https://doi.org/10.1016/j.ijinfomgt.2019.05.014

Bizga, A. (2023). *Why you should delete old accounts you no longer use*. https://www.bitdefender.com/en-us/blog/hotforsecurity/why-you-should-delete-old-accounts-you-no-longer-use

Bracamonte, V., Orito, Y., Fukuta, Y., Murata, K., & Isohara, T. (2024). Perception of Privacy Tools for Social Media: A Qualitative Analysis Among Japanese. In D. V. S.De.C. & S. P. (Eds.), *Proceedings of the International Conference on Security and Cryptography* (pp. 151–162). Science and Technology Publications, Lda. https://doi.org/10.5220/0012762000003767

Dhir, A., Kaur, P., Chen, S., & Pallesen, S. (2019). Antecedents and consequences of social media fatigue. *International Journal of Information Management*, *48*, 193–202. ENISA. (2024). *Foresight Cybersecurity Threats for 2030* (Issue March). https://doi.org/10.2824/349493

European Union, Special Eurobarometer 532, The Digital Decade, https://europa.eu/eurobarometer/api/deliverable/download/file?deliverableId=88016

Europska Komisija, *Zaštitite svoju privatnost na društvenim mrežama*. https://ec.europa.eu/croatia/secure_privacy_on_social_networks_hr

Fan, X. J., Jiang, X. Y., Deng, N. Q., Dong, X. B., & Lin, Y. X. (2021). Does role conflict influence discontinuous usage intentions? Privacy concerns, social media fatigue and self-esteem. *Information Technology & People*, *34*(3), 1152–1174. https://doi.org/10.1108/ITP-08-2019-0416

Fan, X., Jiang, X., Deng, N., Dong, X., & Lin, Y. (2021). Does role conflict influence discontinuous usage intentions? Privacy concerns, social media fatigue and self-esteem. *Information Technology & People*, *34*(3), 1152–1174. https://doi.org/https://doi.org/10.1108/ITP-08-2019-0416

Farooq, A., Salminen, J., Martin, J. D., Aldous, K., Jung, S. G., & Jansen, B. J. (2024). Exploring Social Media Privacy Concerns: A Comprehensive Survey Study Across 16 Middle Eastern and North African Countries. *IEEE ACCESS*, *12*, 147087–147105. https://doi.org/10.1109/ACCESS.2024.3463869 WE - Science Citation Index Expanded (SCI-EXPANDED)

Federal Trade Commission. (2024). *A Look Behind the Scenes Examining the Data Practices of Social Media and Video Streaming Services* (Vol. 4, Issue 2). https://doi.org/10.36950/apd-2016-011

Garima, & Sheokand, K. (2024). Demystifying the Effect of Social Media Advertising on Purchase Intention of Millennials: Role of eWOM and Privacy Concerns. *JOURNAL OF CREATIVE COMMUNICATIONS*. https://doi.org/10.1177/09732586241246403

Hew, J.-J., Leong, L.-Y., Tan, G. W.-H., Ooi, K.-B., & Lee, V.-H. (2019). The age of mobile social commerce: An Artificial Neural Network analysis on its resistances. *Technological Forecasting & Social Change*, *144*, 311–324. https://doi.org/https://doi.org/10.1016/j.techfore.2017.10.007

Horvat, A. (2024). *Zaštita privatnosti na društvenim mrežama | Repozitorij Fakulteta organizacije i informatike*. https://repozitorij.foi.unizg.hr/islandora/object/foi:8166

How to Maintain Privacy on Social Media | Countingup. (2022). https://countingup.com/resources/how-to-maintain-privacy-on-social-media/

How to Protect Your Privacy on Social Media? – Data Privacy Manager. Retrieved December 19, 2024, from https://dataprivacymanager.net/how-to-protect-your-privacy-on-social-media/

Jozani, M., Ayaburi, E., Ko, M., & Choo, K. K. R. (2020). Privacy concerns and benefits of engagement with social media-enabled apps: A privacy calculus perspective. *Computers In Human Behavior*, *107*. https://doi.org/10.1016/j.chb.2020.106260 WE - Social Science Citation Index (SSCI)

Jung, A. R., & Heo, J. (2024). How Did You Get My Information? The Influence of Personal Information Gathering Methods on Social Media Advertising Effectiveness. *International journal of human-computer interaction*, *40*(20), 6579–6590. https://doi.org/10.1080/10447318.2023.2258019

Kaur, P., Islam, N., Tandon, A., & Dhir, A. (2021). Social media users' online subjective well-being and fatigue: A network heterogeneity perspective. *Technological Forecasting And Social Change*, *172*. https://doi.org/10.1016/j.techfore.2021.121039

Kemp, S. (2024). *The Global State of Digital in 2024 — DataReportal – Global Digital Insights*. https://datareportal.com/reports/digital-2024-october-global-statshot

Khan, M. I., Loh, J., Hossain, A., & Talukder, M. J. H. (2023). Cynicism as strength: Privacy cynicism, satisfaction and trust among social media users. *Computers In Human Behavior*, *142*. https://doi.org/10.1016/j.chb.2022.107638

Liu, X. Y., Feng, R., Chen, X. B., & Yuan, Y. (2024). "Left on read" examining social media users' lurking behavior: an integration of anxiety and social media fatigue. *Frontiers In Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1406895

MacKay, J. (2024). *How to Protect Your Personal Information on Social Media*. https://www.aura.com/learn/how-to-protect-your-personal-information-on-social-media

Meso, P., Negash, S., & Musa, P. (2021). Interactions between Culture, Regulatory Structure, and Information Privacy across Countries. *Journal of Global Information Management*, *29*(6). https://doi.org/10.4018/JGIM.20211101.oa49

Neves, J., Turel, O., & Oliveira, T. (2024). Explaining Social Media Use Reduction As an Adaptive Coping Mechanism: The Roles of Privacy Literacy, Social Media Addiction and Exhaustion. *Information Systems Management*. https://doi.org/10.1080/10580530.2024.2332187

Nuzulita, N., & Subriadi, A. P. (2020). The role of risk-benefit and privacy analysis to understand different uses of social media by Generations X, Y, and Z in Indonesia. *Electronic Journal of Information Systems in Developing Countries*, *86*(3). https://doi.org/10.1002/isd2.12122

Pew Research Center. (2023). *How Americans View Data Privacy*.

Plummer, D., Karamouzis, F., Alvarez, G., Hill, J., Sallam, R., McMullen, L., Sicular, S., Ramsey, M., Andrews, W., Bittman, T., Gupta, R., Scheibenreif, D., Raskino, M., Henein, N., Moyer, K., Furlonger, D., Johnson, B., Markkanen, A., Jones, L., … O'Donohue, R. (2023). *Gartner's Top Strategic Predictions for 2024 and Beyond — Living With the Year Everything Changed*. *November 2023*. https://www.gartner.com/doc/reprints?__hstc=254338199.dfb32abfb3189962387122c6e2bbd286.1704311186728.1704311186728.1704311186728.1&__hssc=254338199.8.1704311186728&__hsfp=1654019965&id=1-2FZ2U5YZ&ct=231218&st=sb&submissionGuid=4c21bbc0-5ce2-42d7-a424-8517

Tang, Y., & Ning, X. (2023). Understanding user misrepresentation behavior on social apps: The perspective of privacy calculus theory. *Decision Support Systems*, *165*. https://doi.org/10.1016/j.dss.2022.113881

TechTarget. (2024). *6 Common Social Media Privacy Issues | TechTarget*. https://www.techtarget.com/whatis/feature/6-common-social-media-privacy-issues

Indirect User Experience Evaluation: Insights from Employees Working on a Public Sector Information System

Krešimir Kavran
InfoDom Ltd.
Andrije Žaje 61, 10000 Zagreb, Croatia
kresimir.kavran@infodom.hr

Boštjan Šumak
University of Maribor
Faculty of Electrical Engineering and Computer Science
Koroška cesta 46, 2000 Maribor, Slovenia
bostjan.sumak@um.si

Abstract. *Customer satisfaction and customer feedback are essential to software manufacturers for a better understanding of customer needs.*
This paper focuses on one Croatian software development company that relies on its employees to communicate with customers and propose necessary actions for improving customer satisfaction.
A questionnaire was used to collect the data. It aims to determine customer involvement in developing new features and what information can be extracted from customer feedback.
The results show that senior employees can make decisions based on customer feedback. Additionally, these employees involve customers in the development process.

Keywords. Customer satisfaction, customer feedback, user experience, user involvement, data collection

1 Introduction

When creating a new product or improving an existing one, developers and designers must consider the end users, their needs, and how they will or want to use a product. End users care about doing their tasks and achieving their goals, not how the product was made or the company that created it (Saffer, 2010).

For that purpose, user experience metrics are essential. From these metrics, the company and its employees can get a deeper insight into the acceptance of their product. Metrics add structure to the development and evaluation process and inform decision-makers (Tullis & Albert, 2013).

When creating a product, the focus must be on the end users. The user-centered design process emphasizes the needs and perspectives of end users. It is a process in which the needs, wants, and limitations of the end users of a product are given attention. In the user-centered design process, users are involved in the design process (Allen & Chudley, 2012), similar to the involvement of the end-users in Agile software development. Agile focuses on lightweight working practices, frequent deliveries, and customer collaboration (Kupiainen et al., 2015).

The selected information system within the scope ot this research is a private information system for state administration bodies and, as such, is not available to everyone. Depending on the organization, there may be restrictions on data collection and the ability to communicate with users about their feedback in a mass manner. As a result, the software development company has limited options for gathering user feedback and opinions about the information system.

This research paper aims to examine user experience through a proxy, i.e., by asking the company's employees who communicate with users about their users' opinions. Additionally, this paper examines employees' positions and their views on their work and work environment to identify potential improvements for enhanced involvement in development and increased customer participation in development.

For this purpose, the following research questions were specified:

- Research Question 1 (RQ1): Are employees in direct communication with users? If so, what positions or functions do they hold within the company?
- Research Question 2 (RQ2): Are users involved in the development process?
- Research Question 3 (RQ3): How is the feedback about user experience being managed?

This paper is structured as follows. Section 2 provides background information about user experience and related work to this study. Section 3 describes the research methodology, when and how the research was conducted. Section 4 presents the study's results. Section 5 contains a discussion. Section 6 describes limitations and section 7 describes threats to validity. Section 8 concludes this paper.

2 Background and Related Work

In the past, computers were used mainly by a core of technically oriented users who were not only willing to accept the challenge of overcoming poor usability but sometimes welcomed it as a barrier to protect the craft from uninitiated “outsiders”. Poor usability was good for the mystique, not to mention job security (Hartson & Pyla, 2012).

As more people began to use computers, the public was generally slow to realize they could demand a better user experience. Poor designs can indeed look so bad to users that they are forced to assume they could not be that bad unless they were deliberate. Today, end-users want to use technology to learn, be entertained, and connect with others, looking beyond sheer functionality or usability to achieve emotional satisfaction (Hartson & Pyla, 2012).

In modern software development, user involvement throughout the development process is considered best practice. It leads to increased development productivity and user satisfaction with the product. Users’ involvement relates to their participation in activities associated with specifying, elaborating, prioritizing, reviewing, and verifying the requirements, as well as testing and verifying developed features. The effectiveness of user involvement can vary considerably in different projects, and it can be challenging to achieve (Buchan et al., 2017).

Evaluating the usability based on users' reports is a challenging task. Several factors, including an increase in the number of users (clients) or the context of the user task, influence the final report (Alomari et al., 2020).

To track progress and monitor user involvement, metrics need to be chosen. These metrics facilitate a deeper understanding of the project's current state and inform future decision-making.

There are many papers regarding user experience. Also, based on these papers, systematic literature reviews were conducted.

Amirova et al. (2019) examined what factors affect customer satisfaction and how it can be measured. The most significant results were that it is vital to identify customer needs and roles, incorporate customer feedback into planning and development, and foster long-lasting relationships with customers. The most used techniques for measuring customer satisfaction are interviews, feedback reports, questionnaires, and surveys. The authors stated that it is essential to communicate with the customer.

Ntoa et al. (2021) proposed a framework for evaluating user experience in an intelligent environment. The proposed framework contains 103 metrics for measuring the acceptance and usability of information systems and user experience. The research paper (Majumder, 2025) stated that some users do not know how to express their concerns about the product or interface design, and proposes a solution for tracking users when using the product to collect feedback.

The research paper (Hinderks et al., 2022) aimed to investigate user experience management in Agile software development. The conclusion of that paper stated that no approaches directly address user experience management.

The conclusion from published papers on user experience and user involvement in the projects is that user-centered design works well with Agile software development.

3 Research Method

This study used a questionnaire to collect data to determine which employees communicate with customers based on their role and what are employees' conclusions from that communication.

Challenges in creating the information system for state administration bodies and collecting user feedback include limited access to end-users, legal regulations, and security restrictions. Meaning that initiating communication from the manufacturer of the information system to end users is not allowed, and communication is only allowed with selected persons from state administration bodies (customer representatives). For these reasons, a questionnaire was created to determine what can be determined about user experience based on limited access.

The questionnaire consisted of twenty-eight questions divided into four groups (see Appendix A for the complete questionnaire). Before creating the questionnaire, existing literature and published research papers were examined.

The first group consisted of six questions about employees' work experience in the company and their views on their roles. Four questions were closed (multiple-choice), and two were open (employees could write anything). The last two questions in this group were based on the research paper (Buchan et al., 2017).

The second group consisted of seven questions regarding communication with customers and their involvement in the development process. Four questions were multiple-choice, and three were open-ended. This group's third and fourth questions were selected based on the conclusions drawn from the research paper (Amirova et al., 2019). The fifth and sixth questions are based on the research paper (Buchan et al., 2017).

The third group had twelve questions about employees’ conclusions regarding user experience with the information system, based on customer feedback and user experience evaluation. Nine questions were multiple-choice, and three were open-ended. This group of questions is the most important because it contains questions that determine user options and experiences regarding the user interface and integration with other services, such as automated data collection and notification via email or other communication channels.

The first three questions in the third group are based on (Saffer, 2010). These questions focused on examining user satisfaction with the general usage of the information system, the user interface, and its integration with other services.

The fourth question in the third group is based on the research paper (Alomari et al., 2020) to determine user opinions on the information system immediately after education for new users. The third group's sixth, seventh, and last three questions are based on (Tullis & Albert, 2013). The eighth and ninth questions were chosen according to (Allen & Chudley, 2012).

The fourth group consisted of three questions about customer requests for new features or improvements to existing ones. All three questions were multiple-choice questions. The first question in this last group was chosen according to (Hartson & Pyla, 2012). The second question is the result of the analysis of the research paper (Vanhanen et al., 2018).

In total, twenty questions were closed (multiple-choice), and eight were open questions. For the eight questions, respondents were asked to provide a grade on a scale of one to five. Grade one represents the worst grade, while grade five represents the best grade. Additionally, each of these questions has an alternative answer if the respondent was unable to provide an answer.

The questionnaire was available from October 22, 2023. till November 5, 2023. Employees who worked on the project for at least six months were contacted via email with an explanation of the research and a link to the questionnaire. Based on the criteria, thirty employees were chosen. The questionnaire was anonymous, allowing employees to provide honest answers. However, it is also a limitation because it does not guarantee that all employees will complete the questionnaire or provide an answer to all questions. Ultimately, fourteen employees (46.6 %) completed the questionnaire.

4 Results

This section presents the results of the questionnaire, including answers to the research questions. It is divided into seven subsections. The first four subsections present the results of the questionnaire, while the last three subsections contain answers to the research questions.

4.1 Employee's Work Experience

The first group of questions was designed to gather basic information about employees and their work experience on the project. Most of the employees who completed the questionnaire were male (71.43 %).

Most employees in the company have multiple roles. Of all employees who filled out the questionnaire, 50 % have more than one role. Table 1 lists all roles and the number of employees (respondents) assigned to each role. The most frequently assigned role is Implementer, whose responsibility is to configure and implement new features on clients' servers after deployment. The Product Manager, Product Owner, and Scrum master roles have only one assigned employee, and all three roles are assigned to the same employee.

Table 1. Roles and number of respondents

Roles	Number of respondents
Implementer	6
Business analyst	5
Administrator	4
Developer	4
Customer care	3
Management	3
Tester	3
Technical Product Manager	2
Product Manager	1
Product Owner	1
Scrum master	1

Fig. 1 shows the number of employees on the project per year. Six respondents have been working on the project for less than a year. Three respondents worked on the project between five and ten years. The majority of the employees on the project are new to the company.

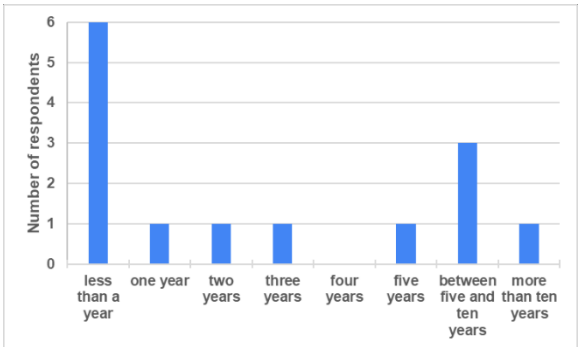


Figure 1. Number of respondents per time on the project

In the questionnaire, respondents were asked to self-evaluate their knowledge of the information system. Fig. 2 shows the number of employees per grade. The lowest grade represents no knowledge of the information system, while the highest grade represents that the employee is thoroughly knowledgeable about the information system. New employees (employed between six and twelve months) self-evaluated their knowledge with grades three and four.

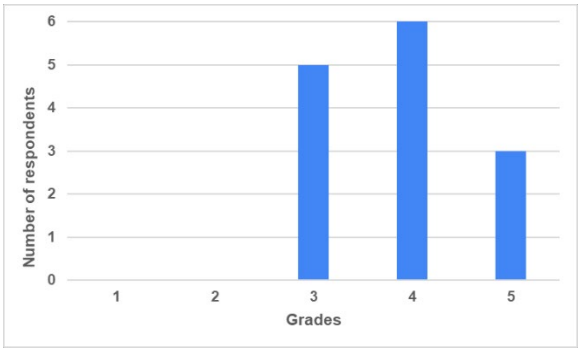


Figure 2. Number of respondents per grade

This group contained two open questions regarding employees' work. The first of these questions was about improving their understanding of assigned roles and work assignments. Most answers focused on improving the onboarding process, which is unsurprising given the high number of new employees. Other answers were better definitions of assignments for employees with multiple roles and improved communication between employees.

The second open question was how their work contribution and results are evaluated by their supervisors. The goal of this question is to determine the metrics on which respondents are focused regarding their contribution. For example, are they focused solely on completing their tasks, communicating with users, or working as a team? The majority of answers were time spent on solving assignments. In two answers, it was stated that they do not know how their work is evaluated.

Employees' answers to those two questions suggest that the company does not invest much time into onboarding and documentation for new employees. Additionally, it involves enhancing internal documentation and procedures for work evaluation and communicating these to employees.

4.2 Communication with Users

This subsection presents the results for the second group of questions, which aim to determine whether employees have direct contact with clients and whether users influence the development process.

Regarding direct communication with clients, 85.71% of respondents answered positively. Respondents who answered negatively are developers.

Regarding whether users provide feedback, most answers were positive (78.57%), but 14.29% of employees reported not having information or being unaware of users' feedback. Only 7.14 % of answers were negative.

Most answers were negative (42.86 %) regarding whether users actively participate in the development. Only 21.43 % of answers were positive. About a third of employees (35.71 %) were unsure whether users participate in development or could not answer the question. There is no direct connection between employees who responded positively to this question and their job roles.

To assess employees' knowledge of the chosen information system, a question was asked about its years of use. The results are shown in Fig. 3. The current major version has been in use for seven years (at the time this paper was written), but the information system has been operational for over ten years.

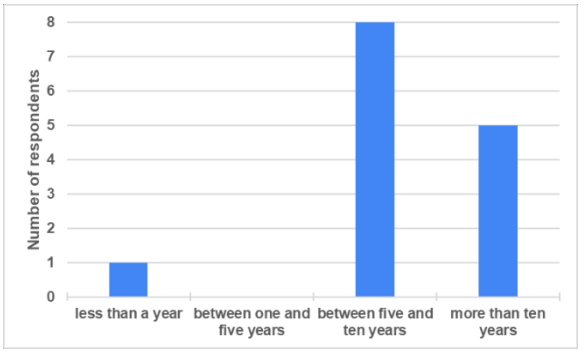


Figure 3. Period of usage of the information system

Only one respondent answered that the information system has been in service for less than a year. That respondent also worked on the project for less than a year. The questionnaire does not determine why the respondent chose the answer “less than a year”, but it can be assumed that it is due to an inadequate or incomplete onboarding process. Other respondents chose the answer “more than five years”, which is in accordance with the actual situation.

This group contained three open questions. The first open question was about the users' role in the information system's development. This question was optional, and only ten respondents answered. All answers can be divided into two groups.

One group contains six answers in which users provide feedback on new features and report defects. The other group includes four answers in which respondents stated that users suggest improvements to the information system or request new functionalities. Four respondents have not answered. The conclusion for this question is that 42 % of respondents consider users to be testers, not customer representatives who have purchased an information system and pay for system maintenance and upgrades.

The second open question was how interaction with the users contributed to the development of the information system. Eleven respondents provided answers to this question. Like the previous question, all answers can be divided into two groups.

The first group, which consisted of eight responses, determined that interaction with users improved their understanding of what users want and how to improve the information system. The second group, comprising three responses, stated that interaction with the users is poor, whether users try to find out how to work around restrictions, or that their feedback is not considered.

Ten respondents answered the third open-ended question, which asked how employees differentiate between good and bad user contributions. Five responses were positive, in which respondents stated that when users understand how the information system works and what can be achieved, their suggestions and feedback are a good contribution. Three responses were negative because users do not want to be involved in the development or lack the necessary knowledge in computer science. Two responses indicated that respondents were unable to answer because they lacked access to the information.

4.3 Employees' Observations and Conclusions

This subsection presents results for employees' conclusions about users' experiences with the information system. All conclusions are based on user feedback and communication with them.

Respondents were asked to grade the user experience for the selected information system in six categories. Table 2 contains the results (all values are displayed as percentages). The first column in Table 2 contains categories. The following five columns show the percentage of answers per grade. The lowest grade is one, and the highest grade is five. The last column in Table 2 shows the percentage of answers without a grade.

Table 2. Respondents' grades of the information system based on user feedback

	1	2	3	4	5	
Ease of use	0	7	57	22	7	7
User interface	0	0	43	29	14	14
Notification methods	0	7	36	29	7	21
Education for new users	0	0	7	29	14	50
User support	0	7	21	7	36	29
Total user satisfaction	0	0	14	64	7	15

Employees (respondents) assigned grades to each category based on user feedback and their interactions with users. Depending on the employee's role, they communicate with the user for various purposes. For that reason, it is interesting to see what conclusions can be extracted.

In the “Ease of use” category, none of the respondents stated that the information system is challenging, e.g., none chose the lowest grade. Most answers (57.14 %) were in the middle grade, meaning that the information system is neither too complicated to use nor too easy to use, leaving many possibilities for improvement.

The category “User interface” got better grades than the previous category. All grades range from middle to high grade. This category has more respondents who were unable to answer.

Regarding user satisfaction with the notification methods in the information system, 21.43% of respondents were unable to answer. The average grade of respondents who answered this question falls in the middle range, leaving a strong possibility for improvement.

The “Education for new users” category received the highest grades, but also had the most unanswered questions, as respondents were unable to answer. Not all employees are responsible for educating new users. It depends on their role. It is expected that there will be more answers without the grade. Employees who hold an education graded user satisfaction with the education with high grades, resulting in the highest average grade.

The category “User support” has the most answers without a grade after the category “Education for new users” because not all employees support the end users. The same applies to the average grade. It got the second-highest average grade after the “Education for new users” category.

The last category was “Total user satisfaction” which includes employees’ options for user satisfaction with the entire information system, its components, and support activities. Only 14.29 % of respondents did not answer this question. Respondents who provided answers gave it a high grade.

This group of questions contained three more multiple-choice questions. Regarding the question of whether users are especially satisfied with certain functionalities of the information system, 35.71 % of respondents answered positively, and 7.14 % of respondents answered negatively. The majority of respondents (57.14 %) did not provide answers because they do not communicate with the users.

In the following question, respondents were asked to name functionalities with which users are especially satisfied if the previous question was answered positively. All answers were about integration with other government services and services from the same company, enabling more accessible work for users and automating business processes.

The question of whether users notice differences between versions of the information system had three possible answers. None of the respondents chose a negative answer. A majority of answers (71.43 %) were positive. Additionally, some respondents (28.57%) did not answer because they either did not communicate with the users or were not in communication with the users long enough to answer the question.

The following question was an open question in which respondents were asked to identify how users learn about new features or improvements in recent versions of the information system. All answers can be summarized into two responses. The first is release

notes, which are part of every version and list new features and improvements. The second is by using the information system. Users who regularly use the information system often notice differences without reading the release notes.

The last multiple-choice question in this group was how often users report errors to the company. Figur 4 presents the results for this question. Most answers were for the middle grade, meaning that users sometimes report errors which result with the same errors in multiple versions of the information system before they are found and resolved.

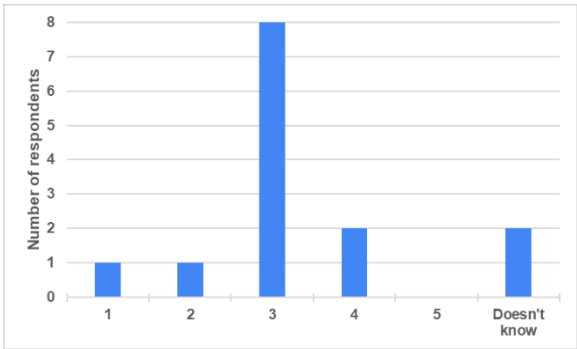


Figure 4. How often do users report errors in the information system?

The company has dedicated software testers who test new versions of the information system before it is delivered to the users. Due to the complexity of the information system, it is challenging to simulate all possible combinations and user actions. The result is that sometimes errors are carried over to the new version and are deployed at the user servers. For this reason, it is important to have users who report errors. If not, the same error will be present in multiple versions before the tester or developer discovers it, or when the user eventually reports that error.

The last question in this group was to list the most significant problems faced by users and provide a brief description. Respondents answered as follows:

- User habits,
- Amount of data,
- Users with poor computer knowledge,
- Users fear using information systems,
- Users forgetting how the information system works,
- Users who do not understand error messages when working with the information system.

User habits can be challenging if users are accustomed to specific processes in solving tasks, e.g., using the information system. Some users prefer routine and repetitive tasks, and introducing changes to existing functionalities (e.g., altering the user interface) can lead to user complaints, regardless of whether the change simplifies their job or streamlines business processes. For this group of users, introducing changes to the information system affects their user experience and productivity within the system (Guo, 2022).

The amount of data generated every month is enormous, as every request, response, document, and piece of citizen information is recorded. This can be an issue for end users with high-level permissions in the information system, who have access to more modules than regular end users. Regular end users have limited access to the information system and stored data. Searching and sorting the data is supported in the information system. Still, the users with elevated permissions who use it for the first time can get overwhelmed by the amount of data if they are unfamiliar with the built-in tools.

Another problem is users with poor computer knowledge. These users are unfamiliar with the technology, making it difficult to explain the specific features of the information system. To help solve this problem, the company can simplify its most frequently used features (if possible).

A combination of limited computer knowledge and an overwhelming amount of data can produce fear in users when using an information system. This leads to many inquiries to customer support and even between end users about the correct way to solve a task. The solution to this problem is constant education for all end users (new and existing users).

End users also forget how to perform specific tasks in the information system because they don't do them often, or they are new to the system. The solution is constant education and open communication with the customer support team.

The information system has many validations for every action, and sometimes, end users are overwhelmed with messages. This can produce fear in the users and prevent them from solving their tasks efficiently.

Most user problems can be resolved through ongoing education on information system usage. The responses suggest that the education process requires improvement, with a focus on maintaining customer satisfaction (Vanhanen et al., 2018).

4.4 User Requests

This subsection presents the results for the last group of questions in the questionnaire, composed of three questions. The first question was to determine whether user requests contain functionalities already present in the information system. Nobody chose the answer "No", but the answer "Yes" was selected by 64.29 % of respondents. Some respondents (35.71 %) could not answer. That means users do not understand the functionalities already available in the information system.

The second question was about respondents' opinions on user requests. Table 3 shows the results. The question had predefined answers, but also allowed input of additional answers. Only one respondent provided an additional answer stating that the user had

valuable input for improvement. The remaining respondents selected answers from the predefined list. Respondents were allowed to choose more than one answer.

Table 3. Respondents' opinions on user requests

Answers	Responses
The client has conflicting ideas.	3
The requirements are too general.	3
The client has a limited understanding of the requirements that they or the end-users need.	8
User feedback on the quality of the delivered functionalities of the information system often comes with a delay or is absent altogether.	1
The client expects minimal investment in work to develop the desired functionalities.	6
Someone from the company made an unrealistic promise to the user.	5
Someone from the company delays communicating the negative consequences to the user.	5
The client is slow to answer questions during the development of new functionalities.	2
The client has no requests for new functionalities or upgrades of existing ones.	2
The client provided valuable input to improve the information system.	1
Does not know	4

The last question was about converting ideas into tasks for the development team, specifically, what ideas are more straightforward to convert into tasks. Half of the respondents stated that ideas generated within the company are more easily converted into development tasks. That means discussing the idea and defining requirements and acceptance tests is more accessible than the same discussion with clients about their requests. The answer that user requests are more easily converted into tasks was selected by 28.57 % of respondents. Some respondents (21.43 %) could not provide an answer.

4.5 RQ1: Communication with Users

Employees communicate with clients, but not all employees do so. The only employees who do not correspond with the clients are those with only technical roles, for example, "Developer". All roles can be divided into two groups:

- Business role (Product Owner, Product Manager, Business analyst, Management, etc.),
- Technical role (Developer, Tester, etc.).

Employees with technical roles do not need to communicate with the users. The company utilizes the Scrum framework to organize its work and development. According to Scrum, the development team needs to be protected from unnecessary disruptions that could derail their focus during a Sprint. Employees with technical roles focus on delivering the Sprint Goal, but meaningful, structured communication with users, when it supports the Sprint Goal, is important. Employees with business roles serve as the interface between the development team and users, and their primary responsibility is to communicate with users (Andrei et al., 2019).

Respondents stated in the questionnaire that they have multiple roles in the company and the development process. Only a few employees have a single role, and these employees have been with the company for a short time, having yet to learn about the development process. This would not be a problem if employees with multiple roles had only business or technical roles, rather than a mixture of conflicting roles.

In this company, it is common for employees to have roles from both groups. The combination of roles can create tensions among employees. The suggestion is to hire new employees and eliminate additional roles for current employees. Employees should have only one role to perform their jobs efficiently.

4.6 RQ2: User Involment

According to the results, users are not directly involved in the development process. However, when employee roles are considered, employees with more critical business roles have answered positively. The remaining respondents hold technical roles, are not in direct contact with users, or have not responded.

This leads us to conclude that users' involvement in the development process depends on the employee. Employees with more extended tenure in the company have greater responsibility (and multiple business roles) and strive to directly involve users in the development process. The rest of the respondents do not directly involve users in the development process.

Company policy on communicating with users should be revised, and employees should undergo additional training to effectively incorporate users into the development process, including testing and discussing new ideas.

4.7 RQ3: User Experience Feedback

As shown in Table 2, it is possible to form an opinion on a specific information system area from user feedback. However, the information that can be extracted depends on the roles of employees and the target users.

End users only see and care about the user interface. They are generally unaware of the inner workings of the information system and typically do not want to

know how it functions or what its architecture is. That means user feedback is limited to what they see on the computer or smartphone screen. Additionally, this presents a challenge in responding to user requests for new functionalities or enhancements to existing ones. User requests state the end goal and what they want to see on the screen, but sometimes lack crucial information about how to get specific information or what operations are necessary to perform.

Employee opinions, therefore, are limited to user interface and graphical design. In the questionnaire, respondents were asked to grade specific information system areas. These areas do not include information system architecture or performance.

The first three areas aimed to determine whether it is possible to form an opinion about information system usage and user interface. The following two areas focused on education and customer support. The last area is overall user satisfaction with the information system.

Some respondents did not provide answers to these questions. It depends on their role. These respondents could not give a grade because they are not directly communicating with the users or cannot form an opinion on indirect communication (written reports or development tasks).

As a conclusion of this section, it is possible to extract information from user feedback and form an opinion based on it. However, user feedback is limited to the graphical design and user interface, and therefore, employees' views are also limited. These employees' views and provided grades can be used to further improve the information system.

5 Discussion

The questionnaire was designed for one specific project but can be applied to other projects in the same company. Additionally, it can be applied to any project, regardless of who the manufacturer is.

The questionnaire aims to determine whether it is possible to obtain helpful information from user feedback and what conclusions employees draw based on it. The results confirm that user feedback can provide insight into the project's current state and be used for future decision-making. However, the questionnaire revealed problems within the company regarding employees' work obligations and internal communication.

The primary challenge in understanding user feedback or requests is the inexperience of younger employees. Their limited knowledge and experience lead to internal miscommunication, which prevents helpful information from reaching all relevant parties and fosters an unproductive work environment.

Additionally, a significant issue is the overlap of conflicting roles for senior employees, which hinders the successful transfer of knowledge to new employees.

Regarding measuring user experience, the questionnaire revealed that it is possible to quantify user experience solely based on user feedback. However, this approach is limited to tasks that users need to perform within the information system. In other words, users focus only on the part of the information system that they need and use, discarding everything else.

The company relies on employees' interpretation of user feedback and requests to assess the usefulness of functionalities and inform decision-making for future development. To improve this practice, it would be recommended to implement a system for recording and assessing users' feedback and reports. The benefit would be the accessibility of information to every involved employee and a better understanding of what the user (customer) wants or needs.

The company needs to estimate user experience before development. This way, it can be determined whether the desired outcome was achieved (Hinderks et al., 2022).

The third group of questions in the questionnaire is the most important, as it contains questions that allow employees to evaluate specific parts of the information system based on user feedback.

User experience is not only a beautiful user interface. User experience includes the whole package, i.e., user interface, integration services, additional plugins, and support. The questions in this third group aim to evaluate user experience for all previously mentioned components, but depending on the project, it may be unclear which components are considered. End users for this project were unaware of what lies beneath the surface, and in general, they provided feedback on the result, i.e., whether they completed their task or not.

It is the task of employees to separate user feedback by component when possible and deliver that information to other employees.

Everything written in this paper indirectly measures user experience. No method for continuous, direct measurement of user experience is implemented in the company. If the company chooses to continue using only communication between employees and customers (indirect measurement of user experience), it is suggested that employees be better educated in communication to improve their data gathering and extracting relevant information.

Additionally, in such cases, it would be beneficial for the company to encourage junior employees with business roles to contact customers more frequently, rather than waiting for customers to initiate conversations, via phone, email, or any other available communication channel. The results showed that senior employees with business roles incorporate users in the development process.

Employees would need to be more proactive and formulate their questions for users in a way that gathers as much relevant information as possible. They should document the answers received and forward their

conclusions to all other relevant employees and management. For this purpose, it is recommended that the company either develop its own solution for recording and documenting communication with users or use an existing solution.

The researchers' opinion is that for the selected project, the questions are adequate. Additionally, questions were selected after reviewing existing research and are grounded in existing research.

While examining existing research papers on this topic, it was found that most research papers focus on customer satisfaction and user interface, i.e., what the user sees on the screen. These papers typically examine user experience through surveys and interviews with users.

The questionnaire included employees and their opinions on user satisfaction with the information system. What is not covered by the questionnaire and what cannot be covered subsequently are the employees' opinions about each user and the organization with which the employees are in contact. It is suggested that a form be created to allow employees to record the reason for and success of each communication after contact with users.

In this way, all other employees would have insight into the course of communication with the individual user and organization and could draw their own conclusions, which may differ from those of the employee who communicated with the user. When only one employee communicates with the same user, it is possible that their prejudices may influence the final opinion, and not all necessary information may be conveyed to other employees. The introduction of the form enables a broader discussion, thereby facilitating better decision-making.

The source of customer feedback is interaction with the users. In the research paper (Fabijan et al., 2015), which contains the results of a literature review in data collection, it is stated that most often, the initial source of customer feedback originates from direct interaction with the customer by using techniques based on active user involvement.

Interaction with customers, such as interviews or face-to-face interactions, is time-consuming and, therefore, challenging to manage in a fast-paced business environment where process efficiency is crucial (Fabijan et al., 2015). To counter this challenge, it is suggested that employees' efforts in completing the form for reporting communication with users be recognized, and, if possible, they should be rewarded for filling out the reports correctly and on time.

The suggested form would enable employees to report all relevant information from users in a unified way. Using the given form, all reports would contain the same structure and relevant information for decision-making.

6 Limitations

As previously mentioned, access to end users is restricted, except for customer representatives and end-user-initiated communication for any reason. This limitation reduces the possibilities for collecting user experience feedback. For this reason, the company needs to collect and analyze every communication with users, whether it be face-to-face, email, or phone.

Limitations in this study include the number of employees working on a project and the number of respondents. At the time of conducting the research, thirty employees were working on the selected project that fulfilled the inclusion criteria. However, not all employees completed the questionnaire.

Fourteen of the thirty employees (46.6 %) filled out the questionnaire. Different results could be obtained if all intended employees had accessed the questionnaire and answered the questions.

Also, if all employees accessed the questionnaire, their answers could confirm the results or disprove the obtained results.

7 Threats to Validity

The main threat to validity is the possibility of author bias. The questionnaire author is affiliated with the company chosen for this study, which presents a potential influence on the results. To eliminate bias, the questionnaire was reviewed by another researcher who is not affiliated with the company.

Existing research papers and published books were consulted to minimize bias when creating the questionnaire. In that way, questions would not be suggestive toward a specific conclusion.

The researcher excluded himself from any discussion regarding the questionnaire with other employees.

Analysis of questionnaire results was conducted with only employees' answers included. Everything mentioned in this paper is contained in the employees' answers. The researcher did not add anything outside the obtained answers.

8 Conclusion

This paper presents the results of a questionnaire conducted in a Croatian software development company. The questionnaire aimed to determine user involvement in the development process and to identify the information that can be obtained from user feedback, which influences employees' perceptions regarding future improvements to the information system.

Employees who have direct contact with the users or have access to meeting reports can form an opinion.

Employee conclusions can be used to further improve the information system.

Additionally, the results of the questionnaire indicate that some employees interact with users. Employees with specific roles communicate with users and shield the development team from external influences.

Furthermore, this paper offers suggestions for enhancing business processes within the company, which can improve employee efficiency and collaboration.

The conclusion of this paper is that user feedback can be utilized by employees to assist management in making informed decisions about the further development of the information system.

Future research will involve collaborating with other teams within the company that face similar challenges in working with state administration bodies. It will incorporate additional methods, such as employee interviews, for data collection. With the expansion of data collection, a larger amount of data will be obtained. The goal of future research is to propose mechanisms and tools for automating the collection of data, data analysis, and tracing the results of decisions based on the obtained data.

References

Allen, J. & Chudley, J. (2012.). Smashing UX Design: Foundations for Designing Online User Experiences. John Wiley & Sons Ltd.

Alomari, H. W., Ramasamy, V., Kiper, J. D., & Potvin, G. (2020.). A User Interface (UI) and User eXperience (UX) evaluation framework for cyberlearning environments in computer science and software engineering education. Heliyon, 6(5)

Amirova R., Khomyakov, I., Mirgalimova R., & Sillitti, A. (2019.). Software Development and Customer Satisfaction: A Systematic Literature Review. International Conference on Objects, Components, Models and Patterns TOOLS 2019, 136-149

Andrei, B.-A., Casu-Pop, A.-C., Gheorghe, S.-C., & Boianggiu, C.-A. (2019.). A study on using waterfall and agile methods in software project management. Journal of Information Systems & Operations Management, 13(1), 125-135

Buchan, J., Bano, M., Zowghi, D., MacDonell, S., & Shinde, A. (2017.). Alignment of Stakeholder Expectations about User Involvement in Agile Software Development. 21st International Conference on Evaluation and Assessment in Software Engineering EASE2017, 334-343

Fabijan, A., Olsson, H. H., & Bosch, J. (2015.). Customer Feedback and Data Collection Techniques in Software R&D: A literature review. 6th International Conference on Software Business ICSOB 2015, 139-153

Guo, Y. (2022.). Does User Preference Matter? A Comparative Study on Influencing Factors of User Activity Between Government-Provided and Business-Provided Apps. Frontiers in Psychology, 13

Hartson, R., & Pyla, P. S. (2012.). The UX Book: Process and Guidelines for Ensuring a Quality User Experience. Elsevier Inc.

Hinderks, A., Mayo, F. J. D., Thomaschewski, J., & Escalona, M. J. (2022.). Approaches to manage the user experience process in Agile software development: A systematic literature review. Information and Software Technology, 150

Kupiainen, E., Mäntylä, M. V., & Itkonen, J. (2015.). Using metrics in Agile and Lean Software Development – A systematic literature review of industrial studies. Information and Software Technology, 62, 143-163

Majumder, A., S. (2025.). Eye-Tracking and Biometric Feedback in UX Research: Measuring User Engagement and Cognitive Load. arXiv.2505.21982

Ntoa, S., Margetis, G., Antona, M., & Stephanidis, C. (2021.).User Experience Evaluation in Intelligent Environments: A Comprehensive Framework. Technologies, 9(2)

Saffer, D. (2010.). Designing for Interaction: Creating Innovative Applications and Devices, Second Edition. New Riders.

Tullis, T., & Albert, B. (2013.). Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics, Second Edition. Elsevier Inc.

Vanhanen, J., Lehtinen, T. O. A., & Lassenius, C. (2018.). Software engineering problems and their relationship to perceived learning and customer satisfaction on a software capstone project. The Journal of Systems and Software, 137, 50-60

Appendix A: Questionnaire Questions

Employee's work experience:

- Gender: male, female
- Select your role or roles: Business analyst, Product Owner, Product Manager, Technical Product Manager, Scrum master, Developer, Administrator, Tester, Implementer, Customer care, Management, other (multiple choice)
- Number of years of work on the current project: less than a year, one year, two years, three years, four years, five years, between five and ten years, more than ten years

- On a scale from 1 to 5, rate your knowledge of the information system. Grade 1 represents no information system knowledge, and grade 5 represents excellent knowledge.
- What can be done, in your opinion, to improve your understanding of your role (position) on the project and help you contribute more effectively?
- How do you think your contribution to the project is evaluated?

Communication with clients:

- Do you communicate directly with users? (Yes, No)
- Do users provide feedback? (Yes, No)
- What is the user's role in the information system's development?
- Do users actively participate in the development of the information system? (Yes, No, I do not know)
- In your opinion, how does interaction with users contribute to the development of the information system?
- How do you distinguish useful (good) user contributions to development from insufficient (bad) user contributions to information system development?
- How long (on average) has the information system been in use? (Less than a year, Between one and five years, Between five and ten years, More than ten years, I do not know)

Employee's observations and conclusions:

- Based on communication with users and user feedback, evaluate how successfully users navigate the information system functionalities on a scale from 1 to 5. Grade 1 represents difficult navigation in the information system, while grade 5 represents easy navigation.
- Based on communication with users and user feedback, evaluate user satisfaction with the user interface on a scale of 1 to 5. Grade 1 represents complete dissatisfaction, and grade 5 represents complete satisfaction with the user interface.
- Based on communication with users and user feedback, evaluate user satisfaction on a scale of 1 to 5 with how they are informed about initiated and completed actions in the information system. Grade 1 represents complete dissatisfaction, and grade 5 represents complete satisfaction with the notification system.
- On a scale from 1 to 5, evaluate the user's satisfaction with the information system immediately after the end of the education for new users. Grade 1 represents complete dissatisfaction, and grade 5 represents complete satisfaction with the information system.
- In your opinion, evaluate the user's satisfaction with customer support on a scale from 1 to 5. Grade 1 represents complete dissatisfaction, and grade 5 represents complete satisfaction with customer support.
- Based on communication with users and user feedback, evaluate the overall user satisfaction with the information system on a scale from 1 to 5. Grade 1 represents complete dissatisfaction, and grade 5 represents complete satisfaction with the information system.
- Based on communication with users and user feedback, list the most significant problems users encounter when using the information system.
- Based on communication with users and user feedback, determine whether the information system has functionalities with which users are particularly satisfied. (Yes, No, I do not know)
- If you answered positively to the previous question, please list the functionalities of the information system with which users are particularly satisfied.
- Do users notice differences or changes in the system between versions of the information system? (Yes, No, I do not know)
- Explain the answer to the previous question.
- On a scale from 1 to 5, evaluate how often users report errors in the information system. Grade 1 represents no error reporting, and grade 5 represents frequent error reporting.

User requests:

- Do the user requests for upgrades include functionalities that already exist in the information system? (Yes, No, I do not know)
- Select the statements that describe user requests for the development of new functionalities or upgrades of existing functionalities and communication with the user regarding their requests: “The user has conflicting ideas.”, “The requirements are too general.”, “The client has a limited understanding of the requirements that they or the end-users need.”, “User feedback on the quality of the delivered functionalities of the information system often comes with a delay or is absent altogether.”, “The client expects minimal investment in work to develop the desired functionalities.”, “Someone from the company made an unrealistic promise to the user.”, “Someone from the company delays communicating the negative consequences to the user.”, “The user is slow to answer questions during the development of new functionalities.”, “The user has no requests for new functionalities or upgrades of existing ones.”, “I cannot answer.”, “Other” (multiple choice)
- According to your experience, which functionalities are more straightforward to convert into concrete tasks, test, and maintain? (“Functionalities created at the request of the user”, “Functionalities created as an idea of a member or members of the team involved in the development and maintenance of the product”, “I do not know”)

Digital Transformation in the Gaming Industry

Igor Pihir, Ivan Kardašić
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
igor.pihir@foi.unizg.hr, ikardasic@student.foi.hr

Abstract. *Digital transformation, as a new paradigm that shifts mind and changes businesses, is meant to allow an organisation to jump towards a more competitive, innovative and sustainable business model. DT paradigm is strategy-oriented and customer-centric; it is based on introducing organisational and technological changes in products/ services, business processes, and/ or business models by implementing contemporary information and communication technologies. DT is a paradigm of today's business and the new digital world. It is applicable across industries and present in all of them. This paper examines the changes in the video games industry, or gaming industry, marked as digital transformation, across the last 25 to 50 years. It explains pre-digitalisation, digitalisation and finally the digital transformation period, as a fine example or fine case study through time. Since this industry can provide evidence of DT in change of processes, products/services and business models or altogether, that makes it a very fine case study for digital transformation. The research was made on several examples of digital transformation of process, product or business model to visualise the change in every aspect and link examples of real products/services, business processes and business models.*

Keywords. digital transformation, video games industry, change, process, product, service, business model

1 Introduction

The gaming industry, or video games industry, relies on developing video games and supporting goods/ artefacts as a part of the creative and entertainment industry. Innovation of business models and gaining a competitive edge are mandatory for this rapidly developing and growing global industry.

The video games industry today spans PC computers, consoles, mobile devices, and handheld consoles. Key components of the industry include game development; publishing (both physical and digital); hardware (including consoles, handheld consoles, joysticks, steering wheels, peripherals, headsets and other accessories); esports and competitive gaming (gaming experience); and community and support services (additional internal society and roles around game titles) (Oguguo, 2024).

In the video game industry, two main players are key to bringing a game to market: the publisher and the game developer. Publishers finance and promote the game, while developers, who can be either independent individuals or large companies, are tasked with the actual game creation. Many publishers also have in-house or vertically integrated developer operations (Thomes, 2015). This brings the key element into the market, but all other elements, as stated before, like hardware, e-sports and competitive gaming, community, and supported services, are also an important part and a significant revenue source for the industry.

The evolution of the global video game industry is characterised by rapid technological innovation and changing business models (Özalp, 2024). Business model innovation (BMI) is crucial to enhancing competitive advantage, and our understanding of how organisations innovate their business model when pursuing contrasting goals (Micheli & Jansen, 2024).

Digital transformation (DT), on the other hand, is a new paradigm that shifts minds and changes businesses, is mented to allow an organisation to jump towards a more competitive, innovative and sustainable business model.

DT is fraught with challenges and disruptions (Tomičić-Pupek et al, 2023; Kutnjak, Pihir, 2025). It is strategy-oriented and customer-centric, and it is based on introducing organisational and at the same time technological change, in all or some of the fields as products/ services, business processes, and/ or business models by implementing contemporary information and communication technology (Pihir, 2022). DT is a paradigm of today's business and the new digital world. It applies across industries and is present in all (Kutnjak et al, 2019). Since the video games industry can provide evidence of DT in change of processes, products/ services and business models, or altogether, this makes it a very fine case study for digital transformation, which is a theme of this research paper. DT is also determined by determinants of digital transformation (Pihir et al, 2019), which will be connected to the video game industry at the conclusion, describing why it is a good case study.

This paper is structured as follows. Section II overviews the methodology and research aspects of the paper. Section III describes the gaming industry and video games trends brief overview. Section IV provides insights into the characteristics of digital transformation and results of the research according to digital transformation in the gaming industry by transformation of processes, products/ services or business models in subsections. Section V concludes the paper by providing a discussion about what are the advantages of such changes are, from the perspective of key determinants of digital transformation.

2 Methodology

The methodology and resources used in this research are based on real cases over time, examples, literature sources and real-world case studies based on the author's own experience or literature research, market studies, and professional and scientific literature that research the gaming industry and the entertainment industry, with a focus on the video games industry. Models made in the paper as scientific contributions are made on several video games industry history analyses combined with knowledge of the authors and personal experience in gaming and the field of digital transformation, making a differentiation and classification of periods and differences across the process, product/ services and business models across industry and period. Models as results of this research and developed by icons from Noun Project (2025) and by storytelling method described in (Tomičić-Pupek et al, 2019). All models are based on description from dependable literature review sources referenced before models in section 4. and subchapters.

3 Gaming Industry and Video Games Trends

The gaming industry trends and market shares per device division have been rapidly growing in the last decade. According to the DRDC article by Arlan (2022), referring to an IDC report from 2022, the video game market is growing in a linear trend. In 2014, it was at 30 billion dollars versus 222 billion dollars in the last reported year, 2022, and mobile platforms had a share of 60% of that cumulative revenue.

In Fig. 1 we can observe game industry market by devices in billions of dollars (Arlan, 2022), from 2014 till 2022, on all platforms: PC (personal computers, laptops and notebooks), mobile (mobile phones and tablets), home consoles (Sony PlayStation, Nintendo, etc.) and handled consoles (small, portable self-contained video game consoles like Game Boy, Switch, Steam Deck). Fig. 1 shows large growth, thanks to mobile platforms. PC and consoles. are in growth but very small, since the market here is at full strength. For example, as can be seen in Fig. 1, the PC video games market grew from 2014 to 2022 from 28 to 40 billion dollars, which is not bad, to a growth of almost 42% in the last 10 years. On the other hand, mobile phone and tablet video games went from 30 billion in 2014 to as much as 136 billion in the year 2022, which is a growth of 350%. That is tremendous growth, and it shows a large potential in the future. Similar growth shows also other reports specialised in this field (Grand View Research, 2025), and also Ernst & Young group reports from 2022 (Porter et al, 2022). predict not only financial growth figures till 2030 but also models of future gaming elements connected to more 3D and VR gaming.

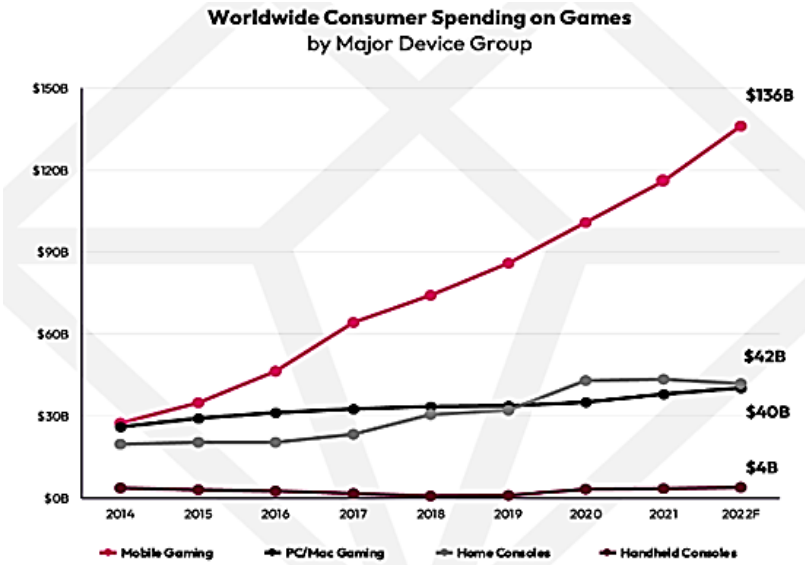


Figure 1. Game industry market by devices in billions of dollars (Arlan, 2022)

This industry reach is not only connected to video game development and publishing, which creates a large market in the entertainment industry, but it also boosts the need for buying the appropriate hardware, peripherals, and gadgets to support players' game experience and add visual, sound, and tactile elements to games, see illustration of this hardware in Fig. 2.

In that manner, for a gaming experience on any device (PC, console, mobile phone, handheld consoles), additional speakers, headsets, joysticks, VR glasses, steering wheels and other equipment are needed to make the experience closer to reality. In Fig. 2, all these assets are presented, and all of them make this industry more competitive since a lot of these peripherals are adapted for specific games or game series and console series. All of it adds value for customers and their customer experience.

Erst & Young future predictions study (Porter et al, 2022) presents that the evolution of gameplay itself will become more complicated, with a lot of new equipment made for games like VR and peripherals that some will be core to the new gaming experience. The use of haptic suits in gaming is also expected to grow significantly in the coming years, and overall wearable devices with sensors will provide feedback to users and collect input from their physical response to gameplay.

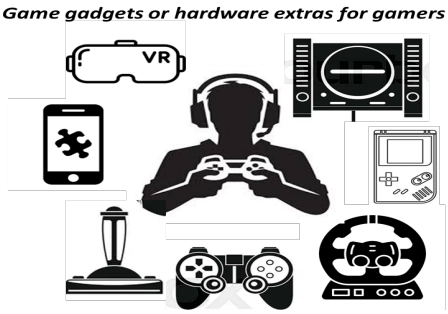


Figure 2. Game gadgets and hardware extras for gamers as additional sales boosters for gaming industry (authors)

Since the video game industry is very competitive and developer and groups of developers and publishers combine their products (games) and platforms (consoles and others peripherals) like PlayStation, Xbox, Nintendo or others are developing a community of players and sales not only games but also hardware and peripherals connected to group of games titles or platform series e.g., Nintendo Wii games. According to older research (Thomes, 2015), when this pre-digital transformation phase was active, Thomes stated that the industry exhibits strong cross-group externalities between software publishers and gamers. Console manufacturers are aware of these positive externalities and invest in first-party titles to make their consoles more attractive to gamers (games come to a certain console first, then are published with a delay on PC or other consoles.).

The video game industry is fragmented and generates significant economic benefits, for example, in the USA, Japan, Finland, Poland and other countries with a large video games industry. By Özalp (2024), diverse pathways to building video game industry clusters emphasise the role of cross-industry skill transfer, intellectual property, and government support. Where traditional video game industry hubs like the USA and Japan he a crossover of capabilities from the entertainment and hardware industries.

For example, this means that games like driving simulations, e.g., NFS, Formula 1, Collin McRae Rally make almost always a need for additional steering wheel, or in flying simulation games, e.g., MS Flight Simulator joystick is needed, or today a VR equipment for new editions. These peripherals make hundreds of dollars in addition to the game and make it more realistic.

4 Digital Transformation of Gaming Industry

Innovation in hardware, software, digital transformation and business models has redefined the boundaries of game development and player experiences (Oguguo, 2024). According to Piers Harding-Rolls, former Head of Games Research, “The gaming industry has benefited hugely from the mainstream adoption of games apps on smart devices, and the transition from physical distribution to digital distribution in established parts of the market such as console gaming. Games are now truly mass market and a very important part of the entertainment sector. Growth has been fuelled by the dominance of free content and in-game monetisation, which expands the adoption of games but also removes the cap on spending for those gamers that are really engaged in the experiences” (Arlan, 2022).

This means that new business models are running in the video game industry, the product or service of the game itself is changing to free games and in-game monetisation. From a process perspective distribution process is changing significantly, and all that means, this is a nice example of digital transformation happening in this industry, as a change of one part or as a whole. To support this change in industry and across time a storytelling models are developed.

The models focused on the distribution process change in subchapter 4.1, game as a product/service in subchapter 4.2. and business model change in subchapter 4.3. Change is supported with several models/storytelling scenarios, from older pre-digital/physical model, digitalisation included and finally digital transformation. Models illustrate what was/is happening across time, means of implementation and execution. All models are made by authors based on self-experience, case studies, and literature reviews. For evolution of distribution

platforms for example Steam (Sayer & Wilde, 2022), (Le, 2023), historical review of games evolution and change of gaming (Datapath.io, 2017), (Le, 2023), (Karthikeyan, 2022), (Özalp, 2024), games communities and technology selection (Larch, 2023), (Le, 2023). All models are self-made by authors using free icons and photos combined to make figures from the Noun Project (2025) and by using storytelling methodology as described in (Tomičić-Pupek et al, 2019)

Further subtopics and their evaluation are researched, discussed and presented in the next three subchapters presenting digital transformation by type of digital transformation according to (Pihir, 2022). Digital transformation of business process change, digital transformation by product or service, and digital transformation by change of business model.

4.1 Digital transformation of game distribution process

Besides game development and sales, the most important business process in the video game industry is the distribution process. To show the evolution and development of this process over time, during the evolution of the video game industry, several storytelling models are developed and presented the process flow and development.

The models focused on change in time, the implementation and execution of the process. Models present the **gaming industry - digital transformation of the distribution process** from the physical process or pre-digitalisation process (Fig. 3), digitalisation of the process (Fig. 4) and digital transformation of the distribution process (Fig. 5).

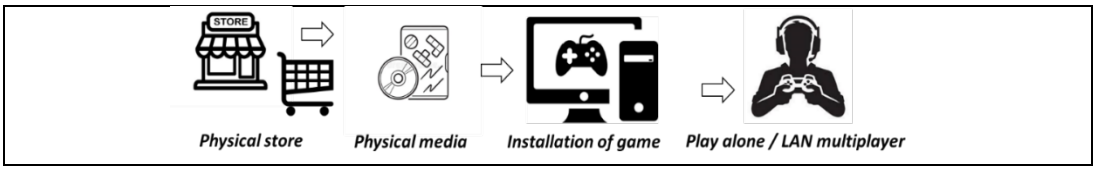


Figure 3. Game distribution process - Physical process, years 1980'/90'-2010'+ (authors)

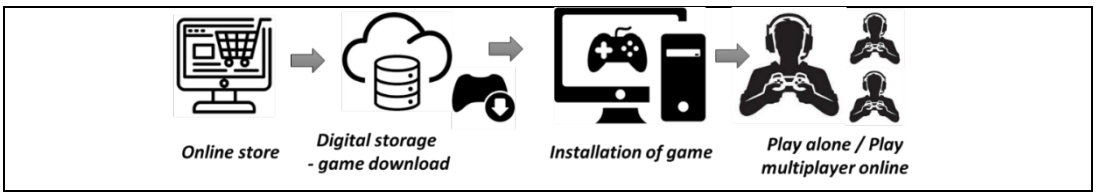


Figure 4. Game distribution process - Digitalization of process, years 2010' – 2020'+ (authors)

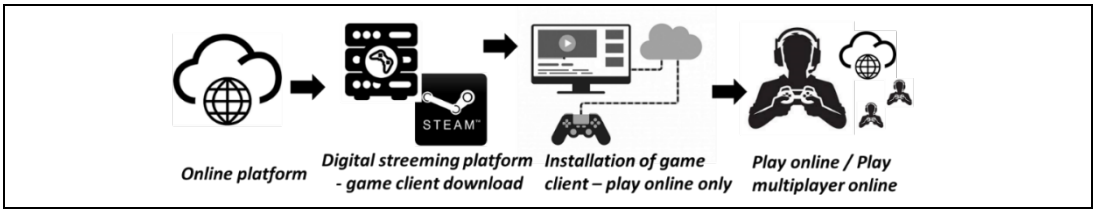


Figure 5. Game distribution process - Digital transformation of process, years 2020' – 202x'+ (authors)

Firstly, Fig. 3 presents the distribution process in the physical normal or pre-digital era. Customers buy games in physical stores (mostly bookstores, electronic shops, or similar). Games are stored on physical media (CD/DVD-ROM or memory card). Customers need to install the game on PC or console, and they can play the game. They play alone, multiplayer is on the same device (second controller) or in local networks. The era is from the early 1980s till about 2010.

Fig. 4 presents the digitalisation period in which games are sold mainly online (physical distribution is still present as in Fig. 3, but not for all games; this is just part of the market; most games are sold online. Sales include the download of a game, like a file, from digital online storage or market, installation of the game on PC or console, and playing the game. Games are played solo or in multiplayer, which is now mostly online (hosted on third-party host servers); local LAN multiplayer is not a dominant type of play. The era is from 2010 to 2020. Later,

Fig. 5 describes the digitally transformed distribution process. Games are sold online, by the platform, not for download or installation, but also for streaming the game. Installation is laminated only to the client while most of the game / or connection to the play environment is still on servers. Games can be installed locally on PC or console, but they can and are mostly played online only. Customers play online in single or multiplayer mode, with a mandatory online connection. The era is from 2020 till now 2025.

A case study for this part could be Half-Life 1 from 1998 is an old physical distribution game case study, or Call of Duty (CoD) in (Fig.3).

Half-Life 2, since 2004. (Fig.4) was the first platform release game coded and distributed online as a download option (sold also in physical distribution for consoles and "older players"). Example are also Microsoft Age of Empires (AoE) with its editions.

For a new digitally transformed distribution process as streaming and playing (subscription games) case study could be World of Warcraft, or for free of charges games: Enlisted, World of Tanks (WoT), World of Warships (WoW) in Fig. 5.

4.2 Digital transformation of game as product/service

Models in this subchapter focus on the game as a product or service that changes over time, through implementation and physical means, from product to product as a service. Evolution can be seen in detail in literature for the evolution of distribution platforms, e.g. Steam (Sayer & Wilde, 2022), (Le, 2023).

Gaming industry - digital transformation of game as product/service from physical product on media or pre-digitalization (Fig. 6), digitalization of product download as a file (Fig. 7) digitalization of product/product as a service – pay and play (Fig. 8) and digital transformation of product/service - game is online and free (Fig. 9).

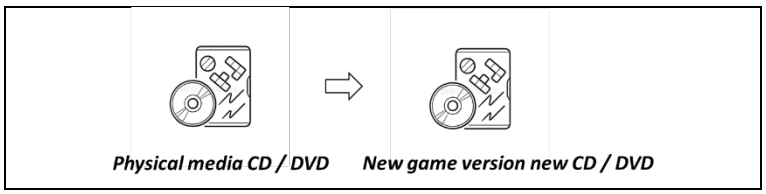


Figure 6. Game as product/service - Physical product on media CD/DVD, years 2000' - 2010'+ (authors)

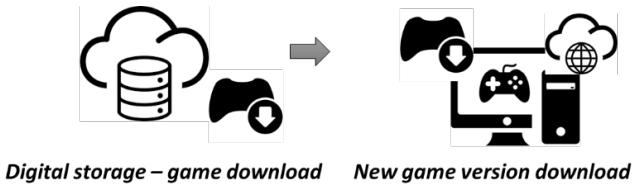


Figure 7. Game as product/service - Digitalization of product download as a file, years 2010' – 2015'+ (authors)

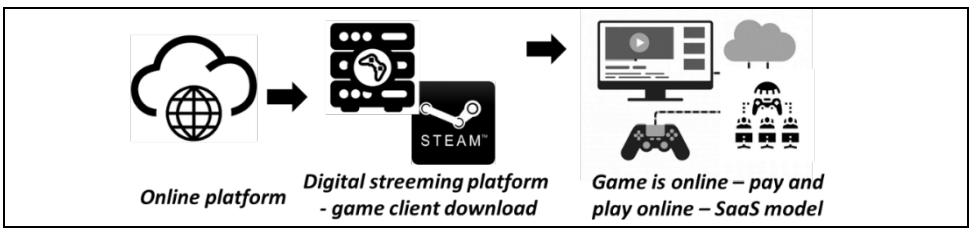


Figure 8. Game as a product/service - Digitization of product product as a service - pay and play, years 2015' - 2020'+ (authors)

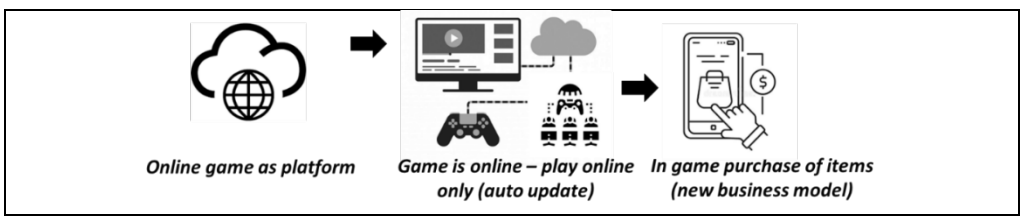


Figure 9. Game as product/service - Digital transformation of product/service - game is online and free, years 2020' – 202x'+ (authors)

Fig. 6. A game as a product/service is stored on physical media sold as a product on media like CD/DVD or memory card. The new game version needs a purchase of new media with new content. The era from 2000 to 2010. Games, e.g., Half-Life, CoD, AoE.

Fig. 7: The game as a product/service is stored digitally as a file on a server. For play, the game needs

to be downloaded, and all new features or extras need to be downloaded again, as a new version. The era from 2010 to 2015 is known as the digitalisation of the product. Fig. 8 game is sold as a service from an online platform where all or part of the game (client) is downloaded to a PC or console. Platform stream games as a service, and the game is mostly present online and played online as a service in a SaaS model. The era from 2015 to 2020 is known as the digitalisation of products exchanged to services.

Fig. 9 game is digital, the game is on a platform. It is played online and only online. A small client is needed on PC or console, but the game is played, and the main content is online. Update of game components is automated and made every day or before the game. The game is normally free. The content of the game itself is free, but some extra content or abilities are purchased within the game. The era from 2020 to the present. Digital transformation of the product/service in the game is fully online and free of charge. Game experience and extra features in the game allow players to buy game resources in-game, but they are not forced to spend real money if they do not want that. Case study games and examples at the end of subchapter 4.3. since they are the same.

4.3 Digital transformation of game as business model

When the digital transformation of the business model is in focus, the video games industry changes over time, and by means of implementation, added value or revenue streams have a lot to offer.

Gaming industry - digital transformation of game as business model from pre-digitalisation of business model - physical business model - physical game sale/media sale (Fig. 10), digitalization of business model - digital game and sequel sale (Fig. 11) digitalization of business model - product as a service - pay and play, (Fig. 12) and digital transformation of business model - game is online and free! Players pay for items in-game (Fig. 13).

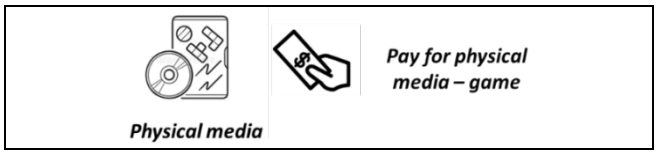


Figure 10. Game as business model - Physical business model - physical game sale / media sale, years 1980'/90'-2010'+ (authors)

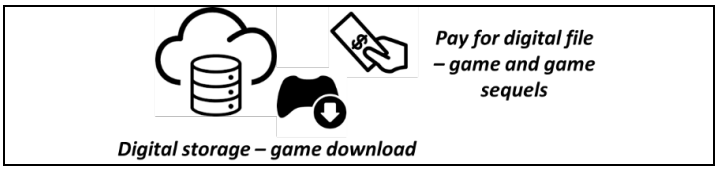


Figure 11. Game as business model - Digitalization of business model - digital game and sequel sale, years 2010' – 2015'+ (authors)



Figure 12. Game as business model - Digitalization of business model - product as a service - pay and play, years 2015' – 2020'+ (authors)

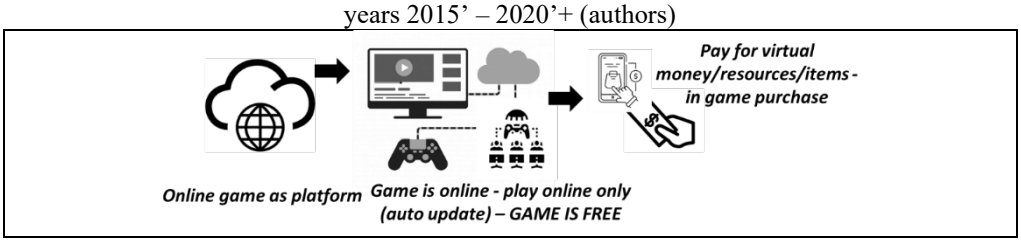


Figure 13. Game as business model - Digital transformation of business model - game is online and free! Player pay for items in game, years 2020' – 202x'+ (authors)

Fig. 10 Physical business model, the game is sold as a product on physical media. The customer pays for physical media. Era from the 1980s to 2010. Fig. 11 Digitalisation of the business model game is sold as a file that needs to be downloaded from digital storage online. The customer pays for a digital file of the game or its sequels. The era from 2010 to 2015. Fig. 12 Digitalisation of the business model of the game is in change from game as a product purchased online to game as a service. The game is digitally streamed from a platform where only the client of the game is downloaded and installed, but the game is played online only or as connected to the platform. Customers pay for online time or a subscription for game/games for some period. Era 2015 till 2020. Fig. 13 Digital transformation of game business model, where the game is only online, and the game is a platform. The game can be played only online or connected, and the game as a product/service is free of charge. Players pay if they want, for extra in-game content or in-game money /resources/items. Era 2020 till now.

A case study for this and the previous part could be the same examples. Any game older, like Half-Life 1, CoD 1and 2, MS AoE 1 and 2, could be in Fig. 6 and Fig. 10. Half-Life 2 from 2004, but most games online downloaded from year 2010+, like Fallout 3, CoD Black Ops, can represent a case study for models in Fig. 7 and Fig. 11. In Fig. 8 and Fig. 12, as digitisation or transformed product to service by streaming and playing (subscription games) could be seen case studies: World of Warcraft, Counter Strike 2, PUBG: Battlegrounds, Warframe, or some of them also in free of charge games Counter-Strike 2, Dota 2, Apex Legends, and Destiny 2. In the category of digitally transformed with in-game purchase and on a free-of-charge platform: Enlisted, World of Tanks, World of Warships, sea Fig. 9 and Fig. 13.

9 Discussion and Conclusion

The video game industry has evolved greatly in the last 25-50 years. Considering the perspectives of types of changes in DT, models from physical product/service to digital, to digital transformation of products/services, have been developed through the paper. The same was done for the video game distribution business process and for the video game industry business model itself.

According to the determinants of DT (Pihir et al, 2019). and the definition of DT (Pihir, 2022), we can conclude that it is an example of digital transformation in which a strategic change with a strategic meaning is made (visible from the trends and sales growth in the introductory analysis). Changes significantly determine the product that later becomes a service. The process is changing significantly from a physical, digitised and ultimately digitally transformed process. The business model is also changing significantly, from charging for the game on the media as a physical medium, to the digitisation of the product into a file that can be downloaded from the online store, then digitisation and change to a service, and ultimately to a free game in which accessories and in-game items can be purchased. All the changes are made by disruptions in time and development, making the change constant in the industry and connected to the final digital transformation, as in (Tomičić-Pupek et al, 2023)

As a second feature/determinant, all these changes in all their forms are focused on users, increasing their number, accessibility and improving the user experience. ICT and process infrastructure, as the third determinant of DT, have been adapted and changed from a pre-digital, digitised and digitally transformed process from physical media, file servers, streaming platforms to online environments and in-game purchases. All this required a shift in thinking, improvement of resources and culture of change among game developers, publishers, sellers and especially in the culture and mindset of the players themselves. The research resulted in the development of a model of these changes by types of digital transformation, creating a good case study for this industry. Numerous changes are clearly shown through periods and the manner of change.

The limitations of the research itself are the small number of examples and the connection to numerous research studies on video game development that are the predominant sources and which are, by nature, professional works related to the time frames within which the changes are positioned and the prevailing model for a particular period. Given that we are dealing with changes over time that we have all witnessed and which are known facts, the brevity of the description and the text that fits into the conference article can be considered limiting factors of this paper.

Concerning student research and its continuation in business models related to digital transformation in the video game industry, the expansion of the research and its continuation are expected.

Acknowledgments

This scientific paper was created in collaboration with a student as a continuation of the research inspired by the final thesis of Ivan Kardašić titled: Research of Digital Transformation in the Game Industry, FOI, Varaždin 2023.

References

Arlan, V. (2022). Of the massive $222 billion video game market, 60% from mobile! And.. what's the fad on Metaverse?. *DRDC June 2022*. Retrieved from https://coachboostgio.com/of-the-massive-222-billion-video-game-market-60-from-mobile-and-whats-the-fad-on-metaverse%EF%BF%BC/#close

Datapath.io. (2017). The History of Online Gaming. Retrieved from https://medium.com/@datapath_io/the-history-of-online-gaming-2e70d51ab437

Grand View Research. (2025). Video Game Market Size, Share & Trends Analysis Report By Device (Console, Mobile, Computer), By Type (Online, Offline), By Region (North America, Europe, Asia Pacific, Latin America, Middle East & Africa), And Segment Forecasts, 2025 – 2030. *Grand View Research GVR-4-68038-527-4*. Retrieved from https://www.grandviewresearch.com/industry-analysis/video-game-market

Karthikeyan, K. (2022). Cloud Gaming: Is it the Future?. Gameopedia. Retrieved from https://www.gameopedia.com/cloud-gaming/

Kutnjak, A., Pihir, I. & Tomičić Furjan, M. (2019). Digital Transformation Case Studies across Industries – Literature Review. In MIPRO 2019, Opatija, Croatia, pp. 1498-1503, 2019. doi: 10.23919/MIPRO.2019.8756911

Kutnjak, A., & Pihir, I. (2025). Key Determinants for Successful Digital Transformation: Insights from Content Analysis and Educational Implications. In K. Daimi, & A. Alsadoon (Eds) *Proceedings of the Fourth International Conference on Innovations in Computing Research (ICR'25). ICR 25 2025. Lecture Notes in Networks and Systems*, vol 1487. Springer, Cham. 350-361. https://doi.org/10.1007/978-3-031-95652-2_30

Larch, F. (2023). eSports History: How it all began. Retrieved from https://www.ispo.com/en/sports-business/esports-history-how-it-all-began#:~:text=But%20on%20October%2019%2C%201972,day%20to%20compete%20in%20Spacewar

Le, L. (2023). A Complete History of Valve: Founding, Hits, and Failures. Retrieved from https://www.historytools.org/companies/a-complete-history-of-valve-founding-hits-and-failures

Micheli, M. R., & Jansen, J.J.P. (2024). Goal orientation and business model innovation in dynamic environments: Evidence from the creative industries. *Creativity and Innovation Management*, 33(2), 197–212. https://doi.org/10.1111/caim.12588

Noun Project. (2025). Free Icons and Stock Photos For Everything. Retrieved from https://thenounproject.com/

Oguguo, P. C. (2024). Innovation and Intellectual Property Use in the Global Video Game Industry. *World Intellectual Property Organization (WIPO) Economic Research Working Paper Series*, No.85/2024, 1-32. Retrieved from https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4816102_code5768558.pdf?abstractid=4816102&mirid=1

Özalp, H. (2024). Heterogeneous development paths to growth and innovation: The evolution of the video game industry across four hubs. *World Intellectual Property Organization (WIPO) Economic Research Working Papers*, No. 84/2024, 1-63. Retrieved from https://doi.org/10.34667/tind.49527

Pihir, I. (2022). Digital Transformation as University Course - Development and Implementation. In *EDULEARN22 Proceedings. IATED, Palma, Spain.,* pp. 7725–7729. doi: 10.21125/edulearn.2022.1799

Pihir, I., Tomičić-Pupek, K., & Tomičić Furjan, M. (2019). Digital transformation playground: literature review and framework of concepts. *Journal of Information and Organizational Sciences*, 43, 33–48. https://doi.org/10.31341/jios.43.1.3

Porter, S., Harrison, J., Ang, A., Golik, R., & Gupta, S. (2022). Gaming Industry Survey 2022 - What's possible for the gaming industry in the next dimension? *Erst & Young*. Retrieved from https://www.ey.com/content/dam/ey-unified-site/ey-com/en-us/insights/media-entertainment/documents/ey-tmt-gaming-metaverse-report-final.pdf

Sayer, M., & Wilde, T. (2022). The 19-year evolution of Steam. Retrieved from https://www.pcgamer.com/steam-versions/

Thomes, T. P. (2015). In-house Publishing and Competition in the Video Game Industry. *Information Economics and Policy*, 32, 46–57. doi.org/10.1016/j.infoecopol.2015.07.005

Tomičić-Pupek, K., Tomičić Furjan, M., Pihir, I., & N. Vrček. (2023). Disruptive business model innovation and digital transformation. *Business Systems Research*, 14, 1-25, 2023. doi.org/10.2478/bsrj-2023-0001

Tomičić-Pupek, K., Tomičić Furjan, M. & Pihir, I. (2019). Storytelling in Teaching Digital Technology Related Courses. In *EDULEARN19 Proceedings. IATED, Palma, Spain*., pp 5338-5344. doi: 10.21125/edulearn.2019.1311

Innovation, Entrepreneurship and Economic Growth

Antonia Kurtović, Kristina Šorić
AI and Simulation in Entrepreneurial Education: Lessons from the RITup Ventures Program

Natali Belušić, Kristina Šorić
Digitalization of Banking Processes and Adaptation for Elderly People

Ivona Kocijan, Kristina Detelj
Digital Resilience of SMEs: Conceptual Distinction and Strategic Factors in the Context of Disruptions

Walter Rath, Berndt Jesenko, Harris Gerzic
Potentials and Barriers of Using Big Data in Austrian SMEs

Idoia de Paz Gómez
Cultural Factors that Inhibit or Facilitate Innovation in Family Businesses: A Systematic Literature Review

Luka Juričinec, Marina Klačmer Čalopa
Digital Transformation of HR Practices: A Systematic Review and Bibliometric Analysis of e-HRM Literature / Digitalna transformacija HR prakse: Sustavni pregled i bibliometrijska analiza literature o e-HRM-u

Marija Vugrinec, Katarina Pažur Aničić
IT project management: Does sustainability matter?

Tanja Šestanj-Perić
The Influence of Blockchain-based Triple-entry Accounting on Accounting Information – a Theoretical and Conceptual Perspective

Ivan Turčić, Ana-Marija Stjepić, Tomislav Jurić
Leveling Up Process Efficiency: AI as a Game-Changer for SMEs

AI and Simulation in Entrepreneurial Education: Lessons from the RITup Ventures Program

Antonia Kurtović, Kristina Šorić
RIT Croatia
Damira Tomljanovića Gavrana 15
10 000 Zagreb, Croatia
{ak4438, kxscad12}@rit.edu

Abstract. *In order to give young innovators the tools they need to succeed in the fast-paced technical and economic environment of today, entrepreneurial education is essential. The RITup Ventures Startup Conference and Academy, which combines workshops, mentorship, and hands-on learning, was created to fill the gap in startup-focused programs for Croatian high school students. The curriculum was created by students in collaboration with RIT Croatia, the global campus of the internationally renowned U.S. university Rochester Institute of Technology (RIT). RITup Ventures introduces high school students to the startup world and helps them create and pitch their own business ideas. Additionally, RITup Ventures consists of four core components: a Startup Conference, a Startup Academy, a Pitching Competition, and pop-up events. However, planning, resource allocation, and time management uncertainties frequently affect the organizational procedures that underpin such projects. Monte Carlo simulation is used in this work to model project completion dates, pinpoint important risks, and improve educational program design. The simulation offers practical insights for enhancing program scalability and reliability by examining critical path dependencies and task length variability. The results show how quantitative techniques like Monte Carlo simulation (enhanced by AI-assisted modeling) can reduce risks and improve project outcomes, underscoring the importance of data-driven decision-making in entrepreneurial education. By demonstrating how analytical methodologies may be included into the design of educational programs, this research adds to the larger conversation on innovation ecosystems and promotes resilience and sustainable growth in youth entrepreneurship efforts.*

Keywords. Entrepreneurship Education, Monte Carlo Simulation, Project Management, Startup Ecosystem, AI-Assisted Decision-Making

1 Introduction

1.1 Project Management in the Event Industry

For events to be delivered successfully, project management is essential, particularly in dynamic settings like startup conferences. Event-based initiatives are distinguished from regular operational tasks by their tight schedules, resource limitations, and complexity (Silvius & Schipper, 2014). Integrated planning across several domains, such as marketing, logistics, content creation, stakeholder coordination, and real-time risk mitigation, is necessary for effective event project management. According to studies, employing structured approaches such as Agile or hybrid models in event planning increases stakeholder satisfaction and improves flexibility (Alvarez-Dionisi et al., 2016). Project management's function goes beyond logistics to include delivering academic impact in the setting of educational conferences, where a variety of participant demands and experiential learning outcomes are crucial. Recent research emphasizes the necessity of more investigation into analytical tools and frameworks for decision-making that assist project managers in settings that are becoming more complex and innovation-driven (Müller, 2023). This research expands on these frameworks by examining how quantitative project planning methods, like Monte Carlo simulation, can enhance the quality and reliability of innovation-focused educational events. The simulation's Python code was generated using OpenAI's ChatGPT, demonstrating how AI can democratize advanced analytics for non-technical project teams. It also presents the RITup Ventures initiative, which was created to fill a void in Croatian high school students' access to entrepreneurial education.

1.2 *RITup Ventures* Case Study

In the rapidly changing world of Artificial Intelligence and other technology advancements, it is becoming crucial to possess entrepreneurial skills such as creative problem-solving, strategic thinking, innovation, leadership, etc (World Economic Forum, 2020). However, developing these skills takes time and

consistent effort, underscoring the importance of cultivating them at a young age (OECD, 2019). Since the majority of efforts concentrate on university-level students or general career guidance, there is a visible lack of entrepreneurial programs aimed at high school students in Croatia. This is consistent with OECD results that emphasize the necessity of incorporating entrepreneurship education into secondary school systems throughout Europe (OECD, 2019). "The score for Entrepreneurial Education at School remained unchanged, with a lowly 2.9, once more the worst of all Croatia's framework scores" (Global Entrepreneurship Monitor, 2023). This noticeable gap in entrepreneurial programs targeted at high school students in Croatia prompted the development of an independent initiative designed to address this need. The initial concept, later on named RITup Ventures, emerged back in February of 2024 as a result of collaboration between RIT Croatia and its students. Building on this foundation, RITup Ventures became a project dedicated to fostering interest for entrepreneurship among high school students. It was made with a goal to educate teenagers about the most important aspects of starting a business. Through various lectures, discussions, and workshops, participants have the opportunity to gain practical knowledge, develop their business ideas, and make a first step into the world of entrepreneurship. RITup Ventures consists of four core components: a Startup Conference (two editions in 2024, 140+ participants), a Startup Academy (3-week workshop series), a Pitching Competition, and pop-up events.

Overall, RITup Ventures has clearly defined short-term, mid-term, and long-term goals, with the focus on sustainable expansion in the future. The short-term goals of the project include encouraging interest of high school students for entrepreneurship, connecting industry experts with high school students, and gathering feedback from participants for further development of the project. Mid-term goals include organization of the second edition of Startup Academy in June and third edition of the Startup Conference in October, and promotion of project results in order to increase the number of participants. Long-term goals include creating a recognizable brand that will become a synonym for youth entrepreneurship education and helping students develop their own startups. Project sustainability is ensured through financial support of RIT Croatia, participant and partner feedback that is systematically used to improve the program to ensure continued relevance and quality of its activities, long-term community growth that will be fostered through engagement with former participants, and promotion of successful teams to inspire a new generation of participants. With these objectives and a clear strategic framework, RITup Ventures aspires to become a leading program for entrepreneurial education among youth in Croatia, while continuously contributing to the development of innovation and the startup ecosystem.

The deliverables and execution of the project are carried out through an iterative approach based on agile principles, which enables flexibility and adaptation at every stage of implementation. The project starts with a planning phase, during which goals of each project component are clearly defined. After that comes the preparation of content and collaboration with both internal and external partners to ensure mentors and experts for the workshops. The execution of the project flows through clearly defined phases: the Conference first serves as a place of inspiration and represents an introduction to the startup world, and then the Academy enables a practical approach and systematized course to help participants develop their business/startup ideas. The Pitching Competition then crowns the project as a way for teams to present everything they learned and the idea they developed, while pop-up events serve as smaller scale events in-between the larger phases of the project. Throughout the whole project, participant feedback is anonymously gathered through surveys in order to improve the program on the go. At the end of each phase, evaluation is done to ensure an even higher quality foundation for the next edition of the project.

From the organizational point of view, the key members include RIT Croatia staff, who take care of the budgeting, marketing campaigns, etc. and a team of 15 RIT Croatia students who coordinate speakers, take care of logistics, provide tech support, prepare the content and materials, find sponsors and partners, contribute to social media promotion, and so on. This illustrates the importance of having a well-coordinated and motivated team in order to set a high standard for the program. However, there are also certain challenges that are faced during the preparation phase. Schedules of all stakeholders have to be aligned, all resources have to be secured within deadlines, tasks have to be properly assigned and delegated, and the content of the program has to be in alignment with the background of the participants (they are all coming with different levels of previous knowledge). All of this requires clear communication, defined expectations, and regular meetups, and it also leaves room for improvements.

To gain deeper insights into the program's operational dynamics and identify those areas for improvement, a Monte Carlo simulation was applied to model and analyze the key processes within the project. This quantitative approach allowed for a systematic examination of uncertainties and performance variability, which then provided data-driven recommendations for future editions of the project. Considering the variability and evolving demand of entrepreneurial educational programs, this paper aims to demonstrate how simulation-based decision support can enhance the design and scalability of innovation-driven initiatives. By combining practical program development with analytical modeling, the research aligns with the goal of fostering innovation through productive disruptions and proposes a replicable

framework for continuous improvement in similar educational and entrepreneurial contexts. The underlying hypothesis assumes that structured application of probabilistic simulation can reveal actionable insights that traditional post-event evaluation methods might overlook, which, at the end, improves both program efficiency and participant experience.

2 Literature Review

A startup can be defined as a team of people (or a young company) that is in the early stage of developing a unique product or service that it aims to bring to the market. The main difference between a regular business and a startup is that startups strive for rapid growth and huge scalability (Startup Genome, n.d.). They also operate under extreme uncertainty and face a high possibility of failure. The most common methodology used in the startup world is Lean Startup (Forbes, 2022). This methodology emphasizes rapid iteration of the product/service, customer feedback, and focusing on building a minimum viable product (most basic version of a product through which founding team can gather data about the preferences of their customers). Lean Startup methodology was applied in the RITup Ventures project as well.

By shifting focus on innovation, problem-solving, adaptability, and risk taking, youth entrepreneurship programs play an important role in economic development. They enable young people to gain essential skills, such as leadership, time management, creativity, and networking, which are all critical for personal and professional development, and later on success (Koh Management, 2023). Moreover, giving a chance to young entrepreneurs can lead to the development of new industries, which can then contribute to the diversification of the economy (Koh Management, 2023).

Agile methodology represents one of the commonly used project management approaches that emphasize flexibility, collaboration, and customer feedback. It consists of sprints - small, manageable, iterative units that represent different projects or parts of the project. This approach is used because it enables teams to quickly adapt to changes and deliver value to customers efficiently (MCI Innsbruck, n.d.). This methodology usually fits startup environments quite well because it can follow rapid changes and high level of uncertainty.

Monte Carlo simulation is a quantitative method used for comprehending the effects of risk and uncertainty in prediction and forecasting models. By using random variable while running simulations, a variety of potential outcomes and the likelihood that they would happen can be provided. By assessing the possible risks and uncertainties connected to different situations, this method aids in the process of making well-informed decisions (Investopedia, n.d.). Monte Carlo simulation can be applied to startup programs to evaluate different project outcomes and enhance planning and execution techniques.

Recent developments in artificial intelligence have created opportunities for enhancing decision-making and project management effectiveness. Automated scheduling, project risk prediction, alternative outcome simulation, and resource optimization are all becoming more and more common uses for AI tools (Dacre and Kockum, 2022). AI reduces manual labor and offers quick analytical support, which helps teams in event planning better handle complexity and ambiguity. This potential is further enhanced by generative AI models, such as ChatGPT, which allow even non-technical individuals to create risk simulations or forecasting models using natural language input. The project team's capacity to adjust under pressure is improved and agile responses are supported by the democratization of sophisticated technologies (Al-Arafat et al., 2024).

3 Methodology

Since the author designed and managed the organizational structure and execution of the RITup Ventures Startup Conference, direct knowledge of the operational framework and challenges of the program were provided by the author as well. The analysis used quantitative modeling, specifically a Monte Carlo simulation, to guarantee analytical objectivity and prevent human bias. The extensive task list, used during the project and kept in Notion (a digital platform for project management and communication), served as the major source of data for the simulation. This database had thorough records of all tasks, both planned and completed, as well as allocated due dates and accountable team members. This task list's structured format defined the scope and order of project activities, making it a suitable basis for risk and performance analysis.

In order to anticipate completion timeframes and identify potential risks, particularly those associated with task duration variability and key path dependencies, the Monte Carlo simulation was used. Monte Carlo provides a probabilistic perspective of project timelines, enabling planners to assess risk, predict bottlenecks, and plan resource buffers appropriately, in contrast to static Gantt charts or basic critical path analysis (Wauters & Vanhoucke, 2016). Because of its adaptability, it is perfect for evaluating one-time, nonlinear projects, such as startup events, where dependencies may change and job durations are varied.

This method is particularly interesting because it incorporates generative AI. Based on project-specific data and intended results, the Python code for the simulation was generated using OpenAI's ChatGPT. This demonstrates how AI is now a useful tool for project modeling, forecasting, and code production in

addition to communication and brainstorming. Project managers without extensive programming experience can now access complicated simulations more easily thanks to AI's capacity to automate script authoring, test distributions, and explain statistical output (Al-Arafat et al., 2024). Five distribution types (Triangular, Normal, Uniform, Log-Normal and Beta) were tested to account for the uncertainty present in human-estimated task durations, and 1,000 iterations were completed for each case. The simulation produced the probability of reaching particular target deadlines (e.g. 40 days and 45 days for the critical path subset, and 85 days and 90 days for the whole task list), as well as the mean and standard deviation of the overall project completion time. In order to mimic the project's entire scope, not simply the crucial path, all tasks were included in the first simulation scenario. This made it possible to identify indirect risks, such as non-essential jobs that can turn into bottlenecks in specific situations. A comprehensive understanding of schedule unpredictability is offered by including the entire task set, which also aids in strategic choices like buffer planning and resource reallocation.

4 Results

The task data gathered throughout the RITup Ventures Startup Conference's planning and execution served as the basis for the Monte Carlo simulation. Three-time estimates were given to each task:

- Optimistic time (O) – the least amount of time if everything goes more smoothly than anticipated.
- Most likely time (M) – anticipated amount of time based on prior experience and typical operating conditions.
- Pessimistic time (P) – the longest anticipated period in the event of delays or issues.

During the project's execution, professional opinion and real-world experience were used to establish these estimations. Following the establishment of the optimistic, most likely, and pessimistic time estimation for every task, the Program Evaluation and Review Technique (PERT) model's formulas were used to determine the expected time and variance. The expected time (ET) is

$$\frac{O + 4M + P}{6} \quad (1)$$

The variance (V) is

$$\left(\frac{P - O}{6}\right)^2 \quad (2)$$

A probabilistic framework for additional simulation was created by using these data to characterize the central tendency and spread of each task's time. The Z-value for each task was established based on the computed expected times and variances, enabling the computation of the likelihood of completing a task by the deadline. This offered a preliminary indication of possible risk hotspots within the timetable. By creating a Critical Path Diagram, jobs were examined concurrently to figure out how they related to the project's overall schedule. Some tasks were deemed non-critical, while others that directly affected the minimum project length were designated as part of the critical route. Simulation results for the entire task set and the critical path alone were distinguished using this categorization. A comprehensive project table (see Fig. 1) describes the combined computations, and a Critical Path Diagram (see Fig. 2) shows the connections between the tasks. The activities are the following:

- A Confirm conference date with RIT Croatia & book auditorium
- B Contact & confirm all speakers
- C Copy for website
- D Marketing campaign
- E Visuals for LinkedIn
- F Posts for LinkedIn
- G Drinks & Snacks
- H Reminders for speakers
- I Check mics availability and sound check
- J Digital agenda
- K Thank you certificates
- L Name tags
- M Goodie bags for speakers
- N Merch giveaway
- O Introduction presentation
- P Prepare speech and guidelines for the team, T-shirts for the team
- R Feedback form & registration form
- S Write questions for the panel discussion and startup spotlights
- T Thank you emails to everyone involved

Table 1. O, M, and P time

	Optimistic time	Most likely	Pessimistic time
A	1	2	3
B	5	10	14
C	1	2	3
D	10	14	15
E	3	10	14
F	3	10	14
G	1	2	3
H	1	2	3
I	4	5	6
J	2	3	4
K	3	5	6
L	3	5	6

M	1	2	3
N	1	2	3
O	3	5	6
P	1	2	3
R	3	4	5
S	1	2	3
T	1	2	3

Table 2. ET, V and critical path

	Expected time	Variance	On critical path
A	2,00	0,11	No
B	9,83	2,25	Yes
C	2,00	0,11	Yes
D	13,50	0,69	Yes
E	9,50	3,36	No
F	9,50	3,36	No
G	2,00	0,11	No
H	2,00	0,11	Yes
I	5,00	0,11	No
J	3,00	0,11	No
K	4,83	0,25	No
L	4,83	0,25	No
M	2,00	0,11	No
N	2,00	0,11	No
O	4,83	0,25	Yes
P	2,00	0,11	Yes
R	4,00	0,11	Yes
S	2,00	0,11	Yes
T	2,00	0,11	Yes

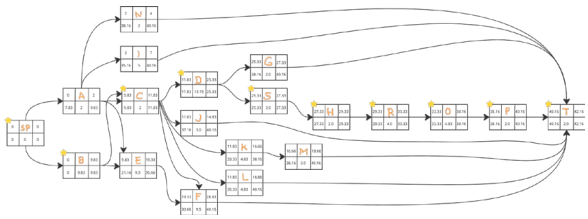


Figure 1. Critical path diagram for the Startup Conference project

For the critical path analysis, the sequence of tasks identified was B–C–D–S–H–R–O–P–T, with a total project variance of 3.86 and a standard deviation of approximately 1.96 days. The simulation estimated the expected completion time of the critical path at 42.16 days. Two deadline scenarios were evaluated to assess the likelihood of on-time project completion. For Deadline 1, set at 40 days, the Z-value was calculated as –1.10, corresponding to a probability of 13.58% that the project would be completed on or before that date. In contrast, Deadline 2, set at 45 days, resulted in a Z-value of 1.45, with a significantly higher completion probability of 92.58%.

Using the anticipated timeframes and variances determined in the first couple of steps, a Monte Carlo simulation was run to further investigate the uncertainty and unpredictability of the project's completion time. Two different scenarios were used in the simulation:

- All project tasks – taking into consideration the project's entire scope.
- Critical path tasks only – the sequence of tasks that directly determined the project's minimal completion time.

Task durations were randomly sampled according to the previously established input distributions, and 1,000 iterations were used to simulate each situation. The simulation also investigated the ways in which the observed task completion patterns fit other theoretical distributions, including Triangular, Normal, Uniform, Log-Normal, and Beta.

With a standard deviation of 4.49 days, the findings for the entire task set showed a mean completion time of 84.62 days. 53.7% of the project was expected to be finished in 85 days, but 88.8% of the project was expected to be finished by the 90-day deadline.

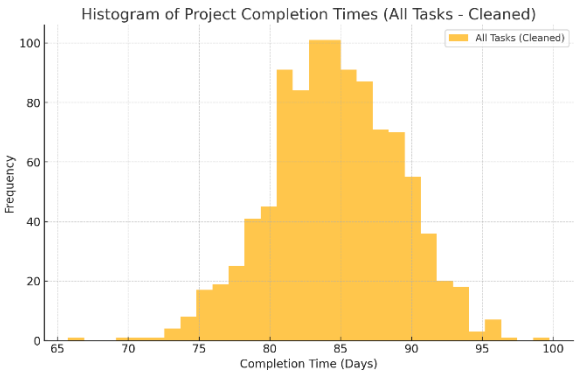


Figure 2. Probability distribution of project completion time for all tasks (Monte Carlo simulation output).

The project's risk profile improved when the analysis was limited to the critical path tasks. The standard deviation decreased to 2.47 days, and the mean completion time decreased to about 41.26 days. There was a 30.6% chance of finishing in 40 days, and a 94.2% chance of achieving a 45-day deadline.

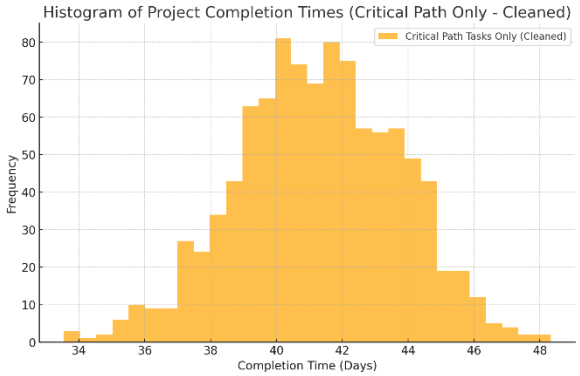


Figure 3. Probability distribution of project completion time for critical path tasks (Monte Carlo simulation output).

Overall, the probability values were calculated by counting the number of simulation iterations in which the project was completed within a specified deadline (e.g., ≤85 days), then dividing that number by the total number of iterations (1,000). This produces an empirical estimate of the project's probability of meeting a specified schedule in the face of uncertainty. One example of how AI might democratize access to sophisticated analytical tools is exactly the capacity to generate this simulation code using ChatGPT. AI-enabled solutions give small teams without professional data scientists a way to apply sophisticated modeling techniques without requiring extensive programming experience. In this instance, ChatGPT demonstrated a novel model of project intelligence in which planners work with AI to model, assess, and enhance delivery methods by serving as a low-barrier interface between project data and Python-based analytics.

The simulation investigated how several statistical distributions fit the observed task duration data in addition to forecasting completion times. For this specific project, the Log-Normal and Beta distributions caught extreme outlier cases, while the Triangular and Normal distributions offered the most accurate and balanced approximation.

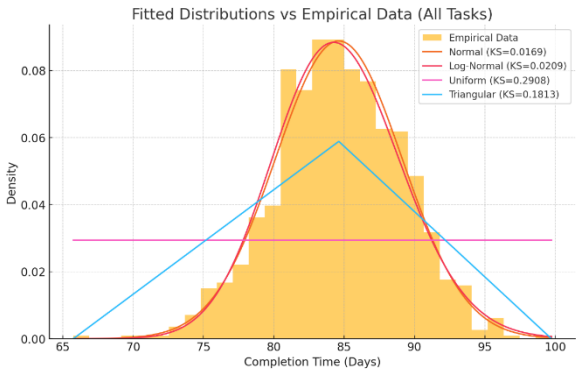


Figure 4. Comparison of fitted distributions to project completion time (all tasks).

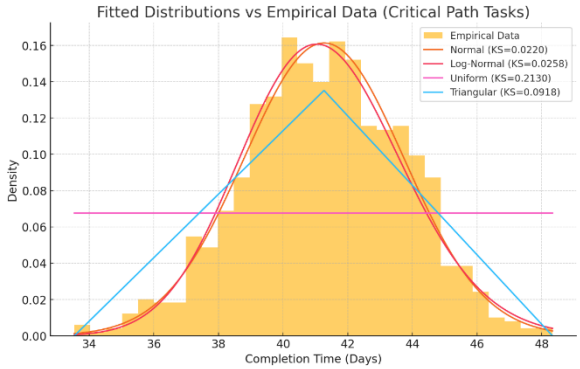


Figure 5. Comparison of fitted distributions to project completion time (critical path only).

5 Discussion

Both quantitative and qualitative insights into the project's delivery process and risk exposure were obtained from the simulation-based analysis of the Startup Conference. A thorough grasp of project dynamics was made possible by the integration of critical path analysis, probabilistic modeling, and actual task data. The findings demonstrate how important task length unpredictability is when organizing an entrepreneurial event. Even with meticulous planning, there was only a 53.7% chance of finishing the entire project in the allotted 85 days, the more accommodating 90-day target was met in 88.8% of cases. This result emphasizes how minor delays in individual tasks, particularly those on the critical path, can have a significant impact on the timetable as a whole. With a 94.2% chance of completing the 45-day target, the simulation showed a more consistent performance profile for the critical path in particular. This implies that the fundamental framework of the project was practical yet accommodating of small disturbances. The comparatively low likelihood (30.6%) of completing in less than 40 days highlights the necessity of buffer time and adaptable resource distribution.

The simulation found:

- One of the main risk factors is task duration variability. Project completion time was disproportionately affected by tasks with estimates that ranged widely between optimistic and pessimistic.
- It was discovered that missing important deadlines was directly increased by delays in essential path jobs, highlighting the significance of accurate estimation and continuous monitoring.
- Risk mitigation methods must be in place for less predictable activities even under structured planning, as evidenced by the 11.2% possibility of surpassing 90 days when considering all tasks.

Similar advantages were discovered in a recent study by Isa and Ahmad Shukri (2019) that used Monte Carlo simulation to manage construction projects.

These advantages included better timeline projections, enhanced transparency in risk analysis, and better detection of bottlenecks. The writers stressed the value of realistic time estimates and probabilistic thinking in planning, just like the RITup Ventures project did. Although their study concentrated on large-scale infrastructure, this scenario shows that, when supported by AI technologies, the same simulation ideas are applicable - and just as effective - in educational event management.

The use of beta and Log-Normal distributions to depict potential extreme cases helped project managers visualize risk in both typical and edge-case scenarios, while the simulation's distribution fitting process further validated the use of Triangular and Normal distributions as helpful approximations for project estimation in educational and entrepreneurial initiatives. These results, taken as a whole, provide credibility to claim that quantitative methods such as Monte Carlo simulations are very successful in enhancing planning, lowering risk, and promoting more intelligent resource allocation in startup-related projects. In addition to improving project durability, this strategy satisfies the need for flexibility and data-driven decision-making in the entrepreneurial ecosystem.

6 Recommendations

The following suggestions are put forth to enhance upcoming RITup Ventures Startup Conference iterations and related activities in light of the insights gained from the simulation results and critical path analysis:

- Emphasis on high-variance tasks – to lessen the possibility of cascading delays, tasks with broad optimistic-to-pessimistic ranges should be given priority for closer supervision and early execution.
- Buffering critical path activities – especially for critical path items, adding intentional time buffers or parallelizing non-dependent operations can boost deadline assurance.
- Continuous risk assessment – as the project moves forward, reevaluate the possibility of on-time delivery by implementing iterative risk reviews at significant milestones using simulation-based techniques.
- Optimize pessimistic forecasts – to reduce overly optimistic planning and increase simulation accuracy, encourage the team to make more accurate and cautious pessimistic forecasts.
- Use distribution insights – to prepare for worst-case scenarios, periodically stress-test the project using Log-Normal and Beta assumptions, but choose Triangular or Normal distributions during early planning stages.

These suggestions can be incorporated into RITup Ventures's operational plan to increase the program's scalability and dependability, guaranteeing a lasting influence on young people's entrepreneurship education and supporting the larger innovation ecosystem.

7 Conclusions

In addition to providing young entrepreneurs with a hands-on learning experience, the RITup Ventures Startup Conference was an invaluable case study for the planning, implementation, and assessment of educational initiatives within the startup ecosystem. The project brought to light the benefits and difficulties of providing entrepreneurship education in a dynamic and uncertain setting by fusing structured project management techniques with real-world event planning. This case study supports a number of best practices for overseeing educational initiatives that promote innovation, including the use of flexible, iterative planning frameworks like Agile, early critical path identification, proactive risk reduction, and unambiguous goal-setting. In order to guarantee the long-term viability of the program, the project also illustrated advantages of open communication, regular feedback loops from partners and participants, and effective community-building techniques. This approach is transferable to other educational programs with high uncertainty (e.g., hackathons, incubators), offering a template for data-driven risk mitigation.

Monte Carlo simulation was incorporated into the project review process, highlighting the importance of data-driven decision-making in situations where uncertainty is inevitable. The simulation was an effective tool for enhancing project design and outcome reliability since it not only shed light on possible dangers but also produced practical suggestions for future optimization. An example of how AI may assist project managers in complicated contexts, even those with low technical expertise, is the incorporation of ChatGPT into the simulation process. A useful model for project delivery in the future, particularly in education and event-based innovation, is provided by the collaboration between human planning and machine-generated analytics.

Beyond the borders of this particular program, the RITup Ventures story demonstrates how entrepreneurial education may change from planning based on intuition to management based on data, better reflecting the realities of contemporary innovation ecosystems. Integrating these analytical techniques into program creation will be crucial for maintaining growth and building resilience as the technological and economic environment that successive generations of entrepreneurs must navigate becomes more complicated. This case study highlights how the use of AI tools and probabilistic simulation represents a productive disruption - one that improves decision-

making, lowers risk, and promotes sustainable growth in entrepreneurial education - as innovation ecosystems increasingly depend on lean, flexible, and tech-enabled approaches.

Acknowledgments

The authors would like to express their sincere gratitude to RIT Croatia for their invaluable support in organizing the RITup Ventures Program.

References

Al-Arafat, M., Kabir, M. E., Morshed, A., & Islam, M. M. (2024). Artificial intelligence in project management: Balancing automation and human judgment. *Future Advances in Engineering and Technology*, 1(2), 18–29.

Alvarez-Dionisi, L., Turner, R., & Mittra, M. (2016). Global project management trends. *International Journal of Information Technology Project Management*, 7(3), 54–73.

Dacre, N., & Kockum, F. (2022). Artificial intelligence in project management: A review of AI's usefulness and future considerations for the project profession. Technical report. https://doi.org/10.13140/RG.2.2.14173.63209

Global Entrepreneurship Monitor (2023). Retrieved from *https://www.gemconsortium.org/economy-profiles/croatia-2*

Forbes. (2022). 10 startup terms every new founder should know. Retrieved from *https://www.forbes.com/sites/abdoriani/2022/09/30/10-startup-terms-every-new-founder-should-know/*

Investopedia. (n.d.). Startup capital definition, types, and risks. Retrieved from *https://www.investopedia.com/terms/s/startup-capital.asp*

Isa, Z., & Ahmad Shukri, F. A. (2019). Monte Carlo simulation-based approach to quantify risks for construction project. Retrieved from *https://www.researchgate.net/publication/361752560_Monte_Carlo_Simulation_Based_Approach_to_Quantify_Risks_for_Construction_Project*

Koh Management. (2023). Importance of youth entrepreneurship in a country and its progress. Retrieved from *https://www.shkoh.com.sg/post/importance-of-youth-entrepreneurship-in-a-country-and-its-progress*

MCI Innsbruck. (n.d.). Entrepreneurship & start-ups glossary. Retrieved from *https://www.mci.edu/en/research/entrepreneurship-start-ups-glossary*

Müller, R. (2023). Exploring the future of research in project management. *Revista de Gestão e Projetos, 14(3), 14–26.*

Organization for Economic Co-operation and Development. (2019). Future of education and skills 2030: OECD learning compass 2030. OECD Publishing. Retrieved from *https://www.oecd.org/education/2030-project/*

Silvius, A., & Schipper, R. (2014). Sustainability in project management competencies: Analyzing the competence gap of project managers. *Journal of Human Resource and Sustainability Studies*, 2(2), 40–58.

Startup Genome. (n.d.). The glossary of ever-evolving startup terms and definitions. Retrieved from *https://startupgenome.com/glossary*

Wauters, M., & Vanhoucke, M. (2016). A comparative study of artificial intelligence methods for project duration forecasting. *Expert Systems with Applications*, 46, 249–261.

World Economic Forum. (2020). The future of jobs report 2020. Retrieved from *https://www.weforum.org/reports/the-future-of-jobs-report-2020*

Digitalization of Banking Processes and Adaptation for Elderly People

Natali Belušić, Kristina Šorić
RIT Croatia
Business Department
Damira Tomljanovića Gavrana 15
10 000 Zagreb, Croatia
{nb3751, kxscad12}@rit.edu

***Abstract**. Digital banking can be used as a convenient way to work with finances, but elderly users often face barriers due to limited digital literacy, cognitive challenges, and unfamiliarity with technology. This paper explores how elderly individuals in Croatia engage with digital banking and how banks address their needs. Primary research includes two surveys: one targeting elderly users and the second one focused on banking professionals. Results show that while banks are aware of the challenges elderly customers face, such as password management, fraud concerns, and app navigation, few provide tailored services or educational support.*
Most elderly respondents actively use digital banking for tasks like checking balances and making payments, but many still rely on family support and lack confidence. The findings also highlight a demand for clearer instructions, simplified interfaces, and better assistance.
Secondary research reinforces the need for accessible design and targeted education. Global examples, such as Barclays' Digital Eagles and Bank of America's fraud education, show effective strategies for supporting elderly users.
This study concludes that improving digital banking access for the elderly is essential for financial inclusion and customer retention. Banks must invest in user-friendly design, personalized education, and dedicated support to bridge the digital divide.

Keywords. digital banking, elderly users, digital literacy, banking support services, digital divide, mobile banking, customer education, inclusive banking

1 Introduction

As digitalization continues to transform the banking industry, it offers users faster, more convenient, and cost-effective ways to manage their finances. However, not all demographic groups benefit equally from these advancements. Elderly individuals, in particular, often find themselves excluded or underserved in the digital banking landscape. In Croatia, as in many parts of the world, this group faces distinct challenges when interacting with online and mobile banking platforms. These challenges stem from a combination of limited digital literacy, cognitive and physical limitations, and a lack of familiarity or confidence with new technologies.

While digital banking has become the norm for younger generations, many older adults struggle with basic tasks such as navigating mobile apps, remembering complex passwords, or identifying fraudulent content. Even though they may be willing to engage with these services, they frequently require assistance from family members or caregivers, leading to a loss of financial independence. Moreover, the lack of tailored support and user-friendly design exacerbates these difficulties, making the experience frustrating or even inaccessible.

This study aims to shed light on the current digital banking experience of elderly users in Croatia and to examine how well banks recognize and respond to their needs. Through both primary and secondary research, the paper investigates the specific barriers faced by this demographic and identifies potential strategies to enhance their digital banking experience.

The paper is composed as follows. The second chapter provides an overview of the relevant literature. The third chapter describes the methodology, including both secondary and primary research. In the fourth chapter, the results of the primary research are presented. The final chapter explores directions for future work and provides the conclusion.

2 Literature Review

Mobile banking enables customers to access their banking information and perform different services like making payments, checking their balance, or checking stock exchange transactions (Tam & Oliveira, 2017).

While mobile banking provides many advantages and an easy-to-use approach, one segment that struggles with the complexity of the Internet and digital features is the elderly. They often do not see the benefits of mobile banking services and are not eager to use them. Additionally, because of their age, the elderly's ability to process and learn new skills is lower

than capable of it. The elderly are still motivated to learn new skills to be able to handle mobile banking, but there is a lack of resources available for them (Castillo-Villar & Castillo-Villar, 2023).

The research conducted by Pejic Bach et al. (2023) confirmed that to help the elderly adapt to Internet services, it is necessary to provide workshops and sessions where the elderly can enhance their digital skills. Moreover, the research suggests that the opinions of family members and close friends impact the elderly's use of the Internet and motivation to learn more about it (Pejić Bach et al., 2023).

Another factor to consider is the time the elderly need to acquire skills related to Internet services. The research suggests that the elderly are better off with practical knowledge that enables them to learn from mistakes. Asking questions also helps the elderly to learn, so being patient is important when showing them new features (Mubarak & Suomi, 2022).

One study showed that the elderly are more likely to use mobile devices and computers if they have a higher education or live with a partner (Petr Balog et al., 2022). Furthermore, when comparing Croatian citizens with American citizens, it was found that American citizens are using digital technologies significantly more than Croatians (Petr Balog et al., 2022).

3 Methodology

The research uses both research methods, combining primary and secondary data. Primary data was collected through two surveys: one targeting bank professionals in Croatia to assess how they address the digital needs of elderly customers, and another directed at elderly individuals to understand their experiences with digital banking. The surveys included both quantitative and qualitative questions to gather insights on challenges, confidence levels, support systems, and usage patterns. Secondary data was gathered through a review of academic literature, industry reports, and case studies, focusing on digital inclusion, accessibility, and best practices in elderly-friendly banking services. This approach provided a comprehensive view of both user experiences and institutional responses.

3.1 Secondary Research

The term that best describes the gap between the elderly's usage of digital banking services is the digital divide. Because of their possibilities access to information and communication technologies, they are less able to access digital banking services. When looking at age as a demographic factor, together with constantly developing digital banking services, there is a gap that is not taken into consideration. Many elderly people have difficulties with eyesight and do not have the skills that enable them to navigate these complex processes. The fact that banks are having fewer physical branches and more online services does not benefit the elderly (Georgieva, 2018).

It is essential to improve current digital banking processes and make the elderly more inclusive in the area of digital banking. One reason for this is to provide the elderly with financial independence. Many elderly people have health difficulties or live in areas that are not close to the physical banks. Additionally, improving the services and educating the elderly can also be beneficial for banks, as the elderly are a significant customer segment. Some important factors to consider include retirement fund management, medical expenses, and money transfers. There are several challenges for the elderly when using digital banking services. The obvious ones include eyesight, interface difficulties such as small text size, multiple steps to execute an order, and complex navigation. Another challenge for the elderly is the security concerns. These include phishing scams and online fraud (Khurshid, 2025).

Research by Gatsou, Politis, and Zevgolis (2017) discovered that the key problems for the elderly are small fonts, small abstract icons, symbols, the need for excessive scrolling, small buttons, difficulties in finding one's location on the site, and cluttered interfaces.

To overcome these challenges, the same research suggests inserting more space between sections, employing larger fonts for headers, reducing the usage of scrolling menus, highlighting clickable and non-clickable features, and providing feedback (Gatsou et al., 2017).

Traditional banks are closing their physical branches more and more, which is especially seen in rural areas. The Bank of Spain revealed that out of 43,000 banks in 2010, just 18,556 were left in 2022. Besides that, the number of employees is lowering, as many services can be done online. Banks need to adjust their processes to fit the different segments they serve, as otherwise, they can become less efficient, profitable, and effective (GDS Link, 2023).

The elderly should try to learn the basics of online banking, choose a user-friendly bank, learn continuously, utilize customer support, and stay updated with features and trends.

One of the top NGOs in India, NIIT Foundation, provides digital training for the elderly, and some topics that are covered include basic computer skills, internet usage, online safety, and usage of digital tools and applications (NIIT Foundation, 2024). Having programs offered to the elderly, such as the one provided by the NIIT Foundation, is essential to make the elderly an inclusive part of the digital banking society.

Bank of America provides resources for the elderly to assist them in navigating digital financial services. Their focus is especially set to provide the elderly with knowledge on recognizing scams, frauds, and setting

up the right security system. They can name a trusted contact or set up a joint account to help them better handle the digital banking services (Bank of America, n.d.).

Another bank that educates the elderly about digital financial services is Barclays through its Digital Eagles program. The elderly are provided with informal conversations while learning digital skills and free modules focused on internet basics and digital safety (Barclays, 2020).

3.2 Primary Research

The primary research consists of two surveys that were conducted among respondents regarding Technology Usage in Banking for the Elderly, and Banking Support for Elderly Customers in Digital Technology.

The survey that was conducted, "Banking Support for Elderly Customers in Digital Technology," studies the approach of international banks in Croatia to the digital banking needs of elderly customers. The findings revealed a gap between the implemented practices aimed at the elderly and awareness of digital services. The data show some measures targeted toward the elderly population, but there is still room for improvement.

4 Results of Primary Research

A strong majority (62.5%) stated that their banks are "very aware" of the difficulties elderly users face when using digital banking tools.

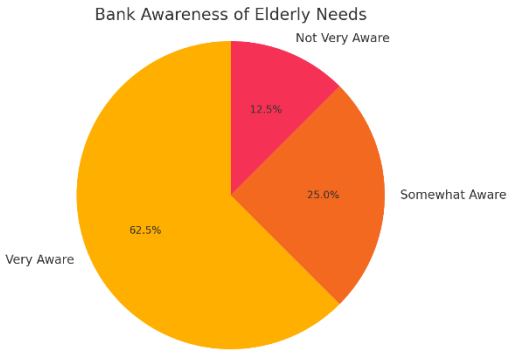


Figure 1. Percentage of bank professionals reporting awareness levels of elderly users' digital banking challenges.

These challenges include remembering passwords (62.5%), concerns about fraud or scams (75%), and navigating apps or websites (50%).

Moreover, only 25% of respondents indicated that their banks consider elderly customers a distinct category in the structure of digital banking services. 37.5% admitted that while the bank does not currently make such considerations for different categories, they plan to change their approach. Another 37.5% reported that their banks do not consider elderly customers separately at all. This disconnection between awareness and segmentation suggests that elderly users are not fully integrated into the digital banking services structure.

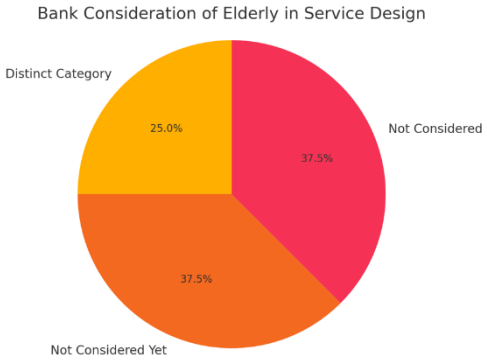


Figure 2. Extent to which banks incorporate elderly users as a distinct group in the design of digital banking services.

Regarding tailored solutions for elderly customers, only 12.5% of respondents stated that their bank has developed digital banking services targeted at elderly customers. On the other side, 25% indicated that only minor adjustments have been made to improve the usage, and 37.5% stated that their institutions treat all users the same regardless of age. Educational programs follow a similar pattern: 50% of respondents noted that their banks do not offer any educational programs specifically for elderly customers, while only 12.5% confirmed regular educational programs.

In terms of support offerings, the most commonly provided support is printed user guides tailored for elderly users, available in 85.7% of cases. Other types of support, such as in-branch workshops, online tutorials, and dedicated customer service hotlines, are less common. Additionally, most banks (75%) address security concerns generally rather than with tailored solutions for elderly customers, potentially leaving this group inadequately protected.

Feedback collection practices are also lacking. Only 12.5% of banks have regular surveys or feedback channels where the elderly population could express their opinions. Most institutions either collect feedback generally or do not collect it. This indicates the inability of elderly users to respond based on their needs and wants.

Planned improvements over the next year include an increased in-branch support for digital banking (62.5%) and digital literacy programs (37.5%), showing some forward improvement. However, the importance of improving elderly digital access is mixed: only 37.5% of respondents consider it a "very important" goal, while an equal percentage reported that it is "not a priority."

The second analysis surveys responses from elderly participants in Croatia, examining their engagement and usage with digital banking services. The data reveals that many elderly individuals are actively using digital banking tools, but barriers related to design, accessibility, and confidence exist.

Demographics show that the majority of respondents are aged between 56–60 (43.9%), followed by 66–70 (14.6%), and 76+ (17.1%). There were 43 participants in the survey. A large proportion of participants reside in urban centers, with 41.5% from Zagreb. The question was designed to ask participants where they live, in terms of country or city, to gather information on whether our participants reside more in urban or rural areas.

In terms of education, respondents are diverse: 28.6% completed elementary school, 23.8% higher education (faculty level), and 16.7% hold master's degrees. This spread reflects a range of digital literacy levels, which directly impacts comfort and ability with online banking.

Regarding device usage, 70% use smartphones to access digital banking services, followed by 20% on tablets and 10% on computers. However, several respondents stated they do not use digital devices at all, instead relying on in-person banking. This indicates a clear digital divide, where a portion of the elderly population remains excluded from online financial services.

Most users (87.5%) use digital banking to check account balances, and 82.5% make payments or transfers. This shows active usage among those who are digitally engaged. However, 30% report needing help, which is usually from family members, to perform banking tasks, while 60% do not require assistance. This reliance on informal support from close members shows a lack of support and education from banks.

In terms of confidence, 51.2% feel "very confident," but nearly 42% report partial or no confidence. Moreover, 51.2% also encounter problems "occasionally" or "very often," especially with app navigation, remembering PINs, and concerns about security.

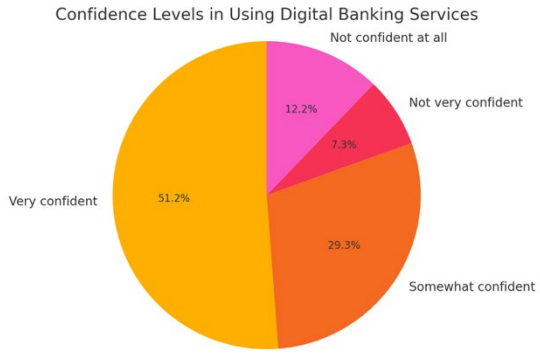


Figure 3. Percentage of elderly users reporting their confidence levels when using digital banking services.

A majority (53.7%) feel that digital banking services are only "somewhat" designed with elderly users in mind, and 29.2% believe they are "not really" or "not at all" elderly-friendly. This shapes the perception that current digital banking platforms lack inclusive design features.

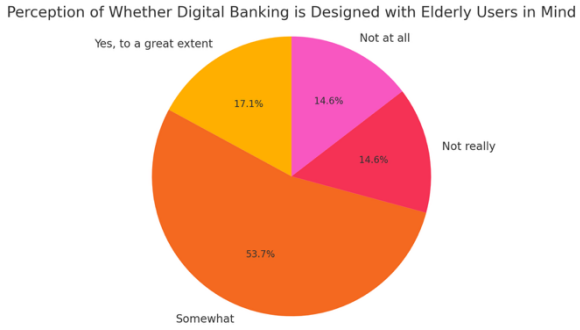


Figure 4. Percentage of elderly users who believe digital banking services are designed with their needs in mind.

When asked about desired improvements, respondents suggested clearer language, virtual assistants, fewer security steps, and simplified tutorials. Additionally, 80.6% said they would be "very likely" to use digital banking more often if it were easier to navigate, demonstrating strong potential for increased adoption if appropriate accessibility measures are implemented.

Finally, when asked about preferred forms of support, the most popular options were phone support (36.8%) and printed guides (31.6%), rather than online tutorials.

5 Limitations and Future Research

This research lacks respondents from the banking sector, as many banks that were asked to participate in the survey declined to do so. The reasons for non-participation differ, but the most common ones include not having the right to fill in or not wanting to reveal the practices of the bank. This impacted the research process as a lack of participants from the bank prevented an adequate amount of data for the analysis. On the other side, this also reflects the current banks' perception and attitudes towards sharing the data regarding the practices they have directed towards the elderly in digital banking services. The future research could reflect more participants from banks or interviews with banks to get more specific data and analysis in the field of banking support for elderly customers in digital technology.

6 Conclusions

In conclusion, this paper reveals that elderly users face barriers when using digital banking services because of limited skills, security concerns, and a lack of inclusivity in the design of these services. Some of the challenges the elderly also face include navigating difficulties on the interfaces, remembering passwords, and general rules not intended for the elderly. While some banks are aware of the challenges that the elderly face, not all implement adequate measures and provide resources. Many banks provide general feedback collection and educational materials that do not benefit the elderly population. The elderly also mostly rely on the help of their family members or close friends, which should also change to make them more confident when using digital banking services. The elderly mostly use digital banking services for checking account balances, making payments, and money transfers. Therefore, it is essential to provide elderly users with adequate education and support because banks will also benefit from having an additional category of customers.

Acknowledgments

This work is part of the Honors Program at RIT Croatia. The authors would like to express their gratitude to RIT Croatia for its continuous support.

References

Bank of America. (n.d.). Elder financial care. Retrieved from https://www.bankofamerica.com/signature-services/elder-financial-services/

Barclays. (2020). Barclays Digital Eagles: Keeping older people connected through lockdown. Retrieved from https://home.barclays/news/2020/04/barclays-digital-eagles--keeping-older-people-connected-through

Castillo-Villar, F. R., Castillo-Villar, R. G. (2023). Mobile banking affordances and constraints by the elderly. Marketing Intelligence & Planning, 41(1), 124-137.

Gatsou, C., Politis, A., & Zevgolis, D. (2017). Seniors' experiences with online banking. In 2017 Federated Conference on Computer Science and Information Systems (FedCSIS) (pp. 623-627). IEEE.

GDS Link. (2023). How to make digital banking more accessible for senior citizens. GDS Link. Retrieved from https://gdslink.com/how-to-make-digital-banking-more-accessible-for-senior-citizens

Georgieva, L. (2018). Digital inclusion and the elderly: the case of online banking. In Proceedings of the LREC 2018 Workshop "Improving Social Inclusion using NLP: Tools, Methods and Resources"(ISI-NLP 2) (Vol. 7, pp. 8-12).

Khurshid, M. (2025). Digital banking problems faced by elderly. Academia Bees. Retrieved from https://www.academiabees.com/digital-banking-problems-faced-by-elderly/

Mubarak, F., Suomi, R. (2022). Elderly forgotten? Digital exclusion in the information age and the rising grey digital divide. INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 59, 00469580221096272.

NIIT Foundation. (2024). Digital literacy for seniors: Empowering the elderly through technology. NIIT Foundation. Retrieved from https://niitfoundation.org/digital-literacy-for-seniors-empowering-the-elderly-through-technology/?utm_source=chatgpt.com

Pejić Bach et al., (2023). Internet usage among senior citizens: self-efficacy and social influence are more important than social support. Journal of theoretical and applied electronic commerce research, 18(3), 1463-1483.

Petr Balog, K., Faletar, S., Jakopec, T. (2022). Digital technology and senior citizens in the city of Osijek: a quantitative study. Libellarium: časopis za istraživanja u području informacijskih i srodnih znanosti, 13(2), 75-94.

Tam, C., Oliveira, T. (2017). Literature review of mobile banking and individual performance. International Journal of Bank Marketing, 35(7), 1044-1067.

Digital Resilience of SMEs: Conceptual Distinction and Strategic Factors in the Context of Disruptions

Ivona Kocijan, Kristina Detelj
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{ikocijan, kristina.detelj}@foi.unizg.hr

Abstract. *Digital resilience is increasingly recognized as a critical organizational capability in today's business environment, characterized by uncertainty, rapid technological change, and frequent crises. Small and medium-sized enterprises (SMEs), due to their limited resources and lower levels of technological readiness, are particularly vulnerable to these challenges.*
This paper provides a systematic analysis of recent scholarly literature with the aim of conceptually distinguishing digital resilience (DR) from related constructs such as digital transformation (DT), digital maturity (DM), and organizational resilience (OR). The focus is placed on identifying key factors that contribute to the development of digital resilience in SMEs, with an emphasis on technological, organizational, and strategic dimensions. The findings suggest that technological preparedness alone is insufficient; organizational culture, leadership, and the capacity for continuous learning play a decisive role. In conclusion, the paper highlights the importance of a holistic approach to building digital resilience, enabling SMEs not only to respond to disruptions but also to transform them into opportunities for innovation and growth.

Keywords. digital maturity, digital resilience, digital transformation, disruptions, organizational resilience, SMEs, VUCA environment

1 Introduction

The modern business environment is increasingly characterized by volatility, uncertainty, complexity, and ambiguity, often described in the literature by the acronym VUCA (Cho et al., 2025). Agility, innovation and adaptability are the new demands of described unstable and uncertain environment, which is a major challenge for organizations, especially SMEs, to overcome. These firms, already at risk of losing their competitive advantage (Cho et al., 2025), often operate with limited resources and personnel, frequently lacking formal organizational structures (Kawane et al., 2024), which hinders their ability to respond to digitalisation challenges.

The COVID-19 pandemic, as an exogenous shock on a global scale, further underscored the critical role of digital adaptation in achieving resilience and maintaining business continuity (Boh et al., 2023; Papagiannidis et al., 2020). Consequently, digitalisation became essential, especially for SMEs, to build resilience and respond to evolving customer and market demands during the crisis (Sinha et al., 2024). In this context, the crisis served as a catalyst for digital transformation (DT), prompting SMEs to rapidly adopt digital technologies and develop innovative business models (Eggers, 2020; Kraus et al., 2021).

Given that SMEs account for more than 90% of all enterprises globally and contribute significantly to employment and GDP (UN, 2024), their resilience is of critical importance for economic sustainability. In this context, DT appears as a necessary response mechanism to external challenges, affecting the entire organization through changes in management, organizational culture and workforce capabilities (He et al., 2023; Senin et al., 2024). However, DT as a stand-alone concept does not ensure resilience in uncertain environments, which has increased scientific attention to digital resilience - defined as the ability of an organization to use digital technologies to manage disruptions, maintain business continuity and adapt to changes (Boh et al., 2023; Tim and Leidner, 2023).

Despite it is increasingly present in scientific discourse, digital resilience is not yet conceptually developed, with empirical research still in its early stages. As a complex construct, it is often conflated in the literature with the concepts of digital transformation (DT), digital maturity (DM), and organizational resilience (OR). This paper offers a conceptual review to clarify digital resilience (DR) in the SME context and explore factors that support its development in uncertain environments. Accordingly, the following research question is posed: How is digital resilience in small and medium-sized enterprises conceptually differentiated from related constructs, and what factors influence its development?

2 Methodology

This study adopts a qualitative literature review approach to conceptualize the notion of DR and to distinguish it, while also examining its interdependence, from related constructs such as DT, DM and OR within the SME context. To ensure reliability and validity of the research findings, high-quality scientific databases Scopus and Web of Science were used. The search strategy was based on various keyword combinations, including *"digital resilience", "SMEs", "digital transformation", "digital maturity", "organisational resilience", "disruptions"* and *"crisis"*. The inclusion criteria were limited to peer-reviewed articles published in English between 2015 and 2025, as research on business digitalization began to intensify during that period. Identified documents were first screened for thematic relevance (title and abstract), followed by full-text examination. A manual content coding process was applied to tag recurring constructs and identify conceptual patterns related to digital resilience, using a concept-centric approach that facilitates synthesis and identification of theoretical gaps (Webster & Watson, 2002; Snyder, 2019).

While this analysis does not follow the full protocol of a Systematic Literature Review (SLR), it establishes a structured basis for synthesizing existing theoretical insights and identifying conceptual gaps. Accordingly, the study outlines multiple directions for further investigation and contributes to the development of a comprehensive SLR methodology in future stages of this research agenda.

3 Theoretical Framework

In the context of digital challenges and disruptions, understanding organizational resilience requires a clear distinction between several related but distinct concepts that are often used interchangeably in the literature. These include digital transformation (DT), digital maturity (DM), digital resilience (DR), and organizational resilience (OR).

3.1 Digital Transformation

According to Hernández Sánchez & Oskam (2025), DT involves not only the integration of digital tools into business operations, but also a complex process of fundamentally reshaping organizational structures and culture to enable innovation and ensure continuous learning. In contemporary literature, it is described as a strategic imperative that drives the creation of new organisational value and enables development in the face of global pressures and dynamic external influences (Gouveia et al., 2024). These conditions often require internal reorganisation and adaptation of business models. Numerous authors emphasise that DT goes beyond implementing IT solutions, encompassing governance, organisational culture, employee capabilities and business processes (He et al., 2023; Senin et al., 2024; Hernández Sánchez & Oskam, 2025). As noted by Reis & Melão (2023), DT has become a defining factor in reshaping modern business environments, exerting long-term influence on how organisations function and evolve in response to continuous technological and societal change. Thus, although its importance is widely acknowledged and the interest of researchers is global, DT still lacks a unified and universally accepted theoretical definition (Kraus et al., 2021). Variations in how scholars frame DT have produced multiple conceptual perspectives on its scope and meaning.

Succurro & Donati (2025) define DT as a sociocultural transformation supported by digital technologies and driven by a clearly defined digital strategy, aiming to enhance agility, efficiency, and connectivity throughout the organization. Muhammad et al. (2025) also stress the importance of integrating digital initiatives and advanced technologies as key drivers of business performance and competitive advantage. On the other hand, Hinings et al. (2018) highlight the innovative nature of DT, defining it as the result of the interaction among various digital innovations that collectively bring about changes in how business systems operate.

3.2 Digital maturity

DM is often viewed as a multidimensional concept, integrating the level of an organization's DT and its ability to adapt to environmental changes. He et al. (2023) conceptualized DM through two dimensions - digital intensity and transformation management intensity - which together influence organizational performance and employees' optimism. The adaptive role of digitalisation is further illustrated by Sinha et al. (2024), demonstrating how it enabled SMEs to flexibly adjust operations in response to evolving customer and market demands during the pandemic. Rather than a fixed state, DM is an ongoing process of organizational and operational adjustment, assessed at both organizational and process levels (Tubis, 2023).

To effectively harness the potential of digital technologies, organizations must develop the capability for rapid adaptation, including the reconfiguration of existing digital solutions (Boh et al., 2023). Gross et al. (2022) argue that DM is a prerequisite not only for operational but also for financial benefits, while also emphasizing that SMEs still lag behind in the implementation of digital initiatives.

He et al. (2023) suggest that strategic investment in digital technologies may can help strengthen operational control. However, they note that such investments do not necessarily lead to advanced adaptive capabilities, such as recognizing external disruptions or quickly finding effective solutions. Furthermore, theoretical and methodological

inconsistencies exist among DM models, and there is still no unified approach to this concept. The absence of a standardised framework makes it difficult to establish a clear connection between digital maturity and digital resilience, leaving this relationship open for further investigation (Grooss et al., 2022).

3.3 Resilience - Organizational and Digital

In conditions of uncertainty and disruption, an organization's ability to effectively cope with challenges becomes essential for its long-term sustainability. For this reason, the concept of OR has been developed, referring to the capacity of organizations to anticipate threats, respond effectively to adverse events, and adapt to changing conditions (Duchek, 2020), including recovery and stabilization following disruptive events or threatening situations (Chaudhary et al., 2024). This concept is not limited to reaction alone; it also includes proactive efforts aimed at identifying opportunities and enabling strategic adaptation (Duchek, 2020), with its further development relying on the integration of diverse capabilities such as ambidexterity, adaptability, and agility (Hokmabadi et al., 2024).

In more recent literature, resilience is increasingly interpreted as a meta-capability. Duchek (2020) suggests that it involves the ability to anticipate, respond to, and adapt to change. Building on this, Khurana et al. (2022) describe it as a higher-order meta-capability that goes beyond traditional dynamic capabilities like sensing opportunities, taking action, and supporting organizational change and transformation (Augier & Teece, 2009; Teece, 2012, 2016). With the development of digital technologies, the concept of DR has also been developed, which emphasizes the importance of digital solutions in responding to disruptions. Although still relatively new (Tim & Leidner, 2023) and actively evolving only in the post-pandemic period since 2022, the importance of DR for business stability and growth has already been recognized in the literature.

Despite being defined in various ways, there is still no widely accepted conceptual framework or established model of DR, which presents a challenge for further research (Tim & Leidner, 2023). Broadly speaking, DR refers to a set of capabilities that enable organizations not only to absorb the impact of external disruptions and regain stability, but also to adapt and transform by leveraging digital technologies effectively (Boh et al., 2023). Seen as a meta-capability (Duchek, 2020), it includes capacities that enable not only recovery from adverse events, but also the active use of digital technologies for adaptation and fostering innovation. Authors highlight that many digital capabilities developed during earlier DT initiatives, initially aimed at operational efficiency and competitive advantage, have become essential for building DR (Boh et al., 2023). Beyond the aspects already mentioned, the concept of DR also involves the design, implementation, and use of information systems to support business recovery and adaptation following exogenous shocks (Tim & Leidner, 2023). The presented definitions show that that resilience isn't just about technical capabilities, but also involves organizational practices, innovation capacity and change management. These capacities are considered particularly critical and essential for small and medium-sized enterprises, which often face limited resources.

3.4 Interrelation and Distinction of Concepts

Although the concepts of DT, DM, and DR are complementary, their roles in the context of organizational adaptation and stability differ significantly. DT is based on the application of digital technologies in organisational and commercial processes, as well as on organisational transformation involving cultural change and talent development, making it a multidimensional process (Hernández Sánchez & Oskam, 2025). It shapes the new landscape of the modern business environment, which still presents major challenges for organizations. However, despite technological advancement, DT does not guarantee resilience. For example, strategies such as just-in-time supply chain management, while effective under stable conditions, proved vulnerable during the COVID-19 crisis, when optimization had to give way to flexibility and robustness (Boh et al., 2023).

The same applies to DM. Conceptualized through two dimensions (digital intensity and transformation management intensity) it reflects both the level of technology use and the degree to which organizations embed vision, governance, and culture into their digital strategies (He et al., 2023). While this supports innovation and adaptability, it does not necessarily ensure the capacity to respond effectively to disruptions. Tim & Leidner (2023) caution that even digitally advanced organizations may be unprepared for unexpected shocks if they lack internal capabilities for reconfiguring and rapidly adapting their business models. Boh et al. (2023) emphasize that digitally prepared organizations are in a better starting position to implement necessary changes, and similar findings were reported by Robertson et al. (2022), who noted that a higher level of digital maturity in SMEs is often associated with greater OR.

These findings suggest that a high degree of DM can contribute to the development of DR, and consequently, to overall OR. Therefore, while DT and DM may support resilience, they are not synonymous with it. While DM reflects technological readiness, resilience also encompasses strategic, organizational, and human resources required to face crises (Awad & Martín-Rojas, 2024; Robertson et al., 2022). DR, in turn, demands the active and purposeful use of technology for adaptation, innovation, and recovery in

uncertain conditions (Boh et al., 2023; Tim & Leidner, 2023).

4 Digital Resilience as a Strategy for SME Growth and Survival

Small and medium-sized enterprises (SMEs), as key drivers of innovation and economic development, are highly vulnerable to increasingly frequent disruptions in an environment marked by volatility, uncertainty, and complexity (Cho et al., 2025; Boh et al., 2023). This vulnerability is largely due to their limited resources, which necessitate a strategic and cautious approach to digitalisation in order to remain competitive and resilient (Sinha et al., 2024), as well as due to a lack of digital competencies, dependence on technological infrastructure (Zirar et al., 2024), and restricted access to financing (Beck & De La Torre, 2007). In this context, DR becomes a necessity rather than an option.

Although DT, and consequently, DM, enhances efficiency and market relevance (Chan et al., 2023; Kim et al., 2025), they do not guarantee resilience. The COVID-19 pandemic demonstrated that even highly digitalized SMEs had limited capacity to reconfigure their business models, highlighting the gap between technological readiness and OR (Boh et al., 2023). Effectively responding to disruptions requires more than just technology and implies an organisation's capacity for rapid decision-making and innovation under pressure. The acceleration of digitalization and the shift from the Industry 4.0 paradigm toward Industry 5.0, with its emphasis on resilience, sustainability, and a human-centric approach (Latino, 2025), further intensify the need for SMEs to adapt their business models and processes.

Within this framework, DR emerges as a strategic meta-dynamic capability (Duchek, 2020; Khurana et al., 2022) that enables proactive crisis response and the conversion of disruptions into opportunities. Beyond ensuring business continuity, it becomes a factor in growth, internationalization, and long-term competitiveness (Razavi Hajiagha et al., 2024). The contemporary business context is shaped by a multitude of external shocks - from pandemics and climate change to geopolitical and cyber threats (Tim & Leidner, 2023). These exogenous crises represent long-term existential threats (Boh et al., 2023) that demand systemic resilience. The COVID-19 pandemic, as a “black swan event” (Cugno et al., 2022), further emphasized the importance of rapid digital responses and innovative action (Papagiannidis et al., 2020). In response, the concept of “digital antibodies” has been developed - digital solutions that enable SMEs to adapt and maintain business continuity (Ragazou et al., 2023).

However, resilience is not the result of technology alone. Robertson et al. (2022) point out that digital leadership had a stronger impact on resilience than technical preparedness, while Boh et al. (2023) and Awad & Martín-Rojas (2024) highlight the importance of strategically aligned capabilities that go beyond the technical dimension. In this context, DM may be viewed as the foundation for the development of DR, but not as its guarantor. Further research confirms that DT positively influences innovation and resilience in SMEs (Al Omoush et al., 2025), enabling better risk management and the exploitation of market opportunities, with an emphasis on strengthening organizational learning (Awad & Martín-Rojas, 2024), developing frugal innovations as a crisis response (Dubey et al., 2022), facilitating market expansion and improved competitiveness (Yang & Panichakarn, 2025), as well as improving firms' international performance (Razavi Hajiagha et al., 2024).

Digitally advanced SMEs showed greater adaptability and faster recovery during the pandemic, strengthening long-term performance (He et al., 2023; Sinha et al., 2024; Senin et al., 2024), while Saezow & Sukhabot (2024) found that SMEs adopting proactive strategies, DT, and network connections were more likely to survive and adapt effectively. Additional evidence from Mossberger (2024) suggests that SMEs with an established digital presence, such as owning a website, were better prepared to absorb the initial shock of the pandemic. Similar findings come from Chinese SMEs in cross-border e-commerce, which overcame numerous challenges with the help of digital technologies (Fodouop Kouam, 2025). Empirical findings clearly indicate that DR is not a one-time response but a capability for long-term adaptation and transformation in conditions of ongoing uncertainty. SMEs that develop this capability do not perceive disruptions solely as threats, but as opportunities for strategic advancement and growth.

5 Key factors for the development of digital resilience of SMEs

DR development in SMEs depends on technologies and on internal and external factors that enable adjustment and recovery in uncertain conditions (Boh et al., 2023; Tim & Leidner, 2023). Kumar et al. (2024) identified ten key elements of DR for SMEs, highlighting strategic management, human resources management, networking, and monitoring as the most effective drivers of change with low dependence on other factors. This chapter analyzes key DR factors in SMEs, grouped by their nature and activity.

5.1 Technological factors

Access to contemporary digital technologies increases flexibility and speeds up the decision making, with special emphasis on the innovative character of generative artificial intelligence (GAI), including

automation of content generation, predictive analytics (Cho et al., 2025), chatbots (Yun & Park, 2022) and the blockchain technology (Chakraborty et al., 2025). These technologies provide a wider and more equal access to advanced digital tools, which is especially important for SMEs with limited resources because they increase efficiency, reduce costs and facilitate innovation. E.g. the chatbots improve customer support (Yun & Park, 2022), while blockchain contributes to transparency and risk management effectiveness (Chowdhury et al., 2023).

According to Chakraborty et al. (2025), the effective integration of new digital technologies into organizational processes, in addition to technical adoption, requires stimulating organizational conditions, such as an innovation-oriented culture, sustainable investment in human capital, and a strengthened cybersecurity infrastructure. Zhang et al. (2025) point out that technological innovations act as an indirect mechanism in the development of organizational strength, which can also be applied to a digital context. Furthermore, Fodouop Kouam (2025) describes how live streaming, online advertising and contactless delivery emerged as successful technological practices that Chinese SMEs used in cross-border e-commerce during the pandemic lockdowns.

5.2 Organizational factors

Together with the technological aspect, innovation, organizational learning and culture, adaptability of employees, their mutual cooperation and continuous learning also play an important role (Hasayotin et al., 2024; Sony & Mekoth, 2022; Wiesböck & Hess, 2020). Innovation culture, willingness to change and support of creative thinking further strengthens resilience in the digital context. Wiesböck & Hess (2020) emphasize the innovation opportunities (new products, services or business processes) by using digital technologies. Innovations and adaptation of business models are key to improving internal operations (Akpan et al., 2024), while digitalization and innovative strategies are often complementarily implemented (Holl & Rama, 2024).

Innovative companies show higher readiness to DT, and an iterative approach in innovation development has proven to be especially suitable for SMEs with limited resources (Lassen & Larsen, 2024). The adaptability of employees has been emphasized as a meta-ability in overcoming the challenges of DT, with a critical role of continuous learning and knowledge transfer (Sony & Mekoth, 2022). The human factor thus becomes key to the successful integration of social and technical systems aiming to achieving organizational goals. Organizational culture further affects employees' willingness to accept technological changes (Wiesböck & Hess, 2020). In dynamic and disruptive contexts, adaptability and diversification of business have been proved as resilience mechanisms, as confirmed by Hasayotin et al. (2024) in the context of tourist SMEs in Pattayi, where diversification has proven to be key to survival during pandemic.

5.3 Strategic factors

In addition to internal capacities, DR also contributes to networking, which not only opens access to resources, but increases the chances of knowledge exchange and collective learning, i.e., it creates new opportunities for growth (Hasayotin et al., 2024). The indirect contribution to the strengthening of DR is also evident through collaboration with other companies, clusters and institutional bodies such as government and state agencies. For instance, according to De Sa et al. (2023) companies that had access to institutional financial support, showed greater resilience in crises. Although the support of state and government agencies does not directly relate to digitalization, it can indirectly strengthen the capacity for DT and resilience, especially in the case of SMEs who rely on limited resources.

6 Discussion

Recent literature suggests that, although DR has attracted increasing scholarly attention, its conceptual foundations remain underdeveloped. This study seeks to delineate it more clearly from related constructs such as DT, DM, and OR; interrelated but not synonymous concepts. DR is increasingly described as a meta-capability that transcends the technical dimension of DM and encompasses an organization’s ability to respond to disruptions, recover, and innovate through the use of digital technologies (Boh et al., 2023; Tim & Leidner, 2023). Experiences from the COVID-19 crisis have demonstrated that a high level of DM does not guarantee resilience - many technologically advanced organizations struggled to adapt rapidly to change. This highlights the importance of factors such as agility and strategic leadership, yet empirical research on the topic remains limited, especially in transition economies. In accordance with the aforementioned fact, there is a need for develop measurable frameworks and conduct longitudinal research on how DR evolves over time, across industries and economic contexts.

This paper constitutes the starting point of a broader research process, providing a preliminary conceptual basis for examining DR in SMEs. The next stage of research will employ a formal SLR methodology, in order to enable a more comprehensive, transparent, and methodologically rigorous synthesis of existing knowledge. Although the current analysis is constrained by its selective scope and interpretative orientation, the incorporation of validated academic sources and a structured concept-centric approach provides a coherent and credible analytical basis for future theoretical development.

7 Conclusion

The paper offers an insightful review of the recent literature to close identified research gaps and distinguish the concept of DR from interrelated concepts of DT and the DM. The VUCA business environment especially with COVID crisis and growing geopolitical risks due to the arising conflicts and political changes around the world pointed out that the DT and the DM do not guarantee that the companies will also be less vulnerable to external shocks if these are not accompanied by flexibility in strategic, organizational, and human side of companies that enable fast adaptation and innovation. This is particularly important in SMEs as the companies often lack different types of resources (material, human and/or financial). Digital resilience thus represents a very important meta-capability of companies that enables them to sustain and thrive in today's disruptive world.

The key factors that will support building of digital resilience are categorized into three dimensions (technological, organizational and strategic) which offers a holistic approach to tackle modern challenges. The companies must integrate advanced technologies but also accompany these with highly adaptive organizational culture, strong leadership, continuous learning, and robust networking. This together with changes in perspective and looking to challenges as business opportunities offers a vision for how SMEs can not only survive but thrive in an ever more unpredictable business condition.

Acknowledgements

This work has been fully supported by the Croatian Science Foundation under the project IP-2022-10-6703.

References

Akpan, I. J., Effiom, L., & Akpanobong, A. C. (2024). Towards developing a knowledge base for small business survival techniques during COVID-19 and sustainable growth strategies for the post-pandemic era. *Journal of Small Business & Entrepreneurship, 36*(6), 921–943. https://doi.org/10.1080/08276331.2023.2232649

Al Omoush, K., Lassala, C., & Ribeiro-Navarrete, S. (2025). The role of digital business transformation in frugal innovation and SMEs' resilience in emerging markets. *International Journal of Emerging Markets, 20*(1), 366–386. https://doi.org/10.1108/IJOEM-12-2022-1937

Augier, M., & Teece, D. J. (2009). Dynamic capabilities and the role of managers in business strategy and economic performance. *Organization Science, 20*(2), 410–421. https://doi.org/10.1287/orsc.1090.0424

Awad, J. A. R., & Martín-Rojas, R. (2024). Digital transformation influence on organisational resilience through organisational learning and innovation. *Journal of Innovation and Entrepreneurship, 13*(1), 69. https://doi.org/10.1186/s13731-024-00405-4

Beck, T., & De La Torre, A. (2007). The basic analytics of access to financial services. *Financial Markets, Institutions & Instruments, 16*(2), 79–117. https://doi.org/10.1111/j.1468-0416.2007.00120.x

Boh, W., Constantinides, P., Padmanabhan, B., & Viswanathan, S. (2023). Building digital resilience against major shocks. *MIS Quarterly, 47*(1), 343–360. https://misq.umn.edu/misq/downloads/download/editorial/766/

Chakraborty, D., Behl, A., Golgeci, I., & Nazrul, A. (2025). Understanding blockchain adoption in SMEs: A mixed-method study of digital transformation, resilience, and senior leadership support. *IEEE Transactions on Engineering Management*, 1–15. https://doi.org/10.1109/tem.2025.3556371

Chan, S., & Jalaluddin, A. K. (2023). Digital technology as a resilience-enhancing tool for SMEs in earthquake-prone developing countries. *E3S Web of Conferences, 447*, 03002. https://doi.org/10.1051/e3sconf/202344703002

Chaudhary, S., Dhir, A., Meenakshi, N., & Christofi, M. (2024). How small firms build resilience to ward off crises: A paradox perspective. *Entrepreneurship & Regional Development, 36*(1–2), 182–207. https://doi.org/10.1080/08985626.2023.2265327

Cho, Y., Park, J., Yoo, J., Kim, S., & Park, H. (2025). A study on Gen-AI technology development trends to enhance small-medium sized enterprise digital competence and management quality. *Journal of Entrepreneurship in Emerging Economies*. Advance online publication. https://doi.org/10.1108/JEEE-10-2024-0485

Chowdhury, S., Rodriguez-Espindola, O., Dey, P., & Budhwar, P. (2023). Blockchain technology adoption for managing risks in operations and supply chain management: Evidence from the UK. *Annals of Operations Research, 327*(1), 539–574. https://doi.org/10.1007/s10479-021-04487-1

Cugno, M., Castagnoli, R., Büchi, G., & Pini, M. (2022). Industry 4.0 and production recovery in the COVID era. *Technovation, 114*, Article 102443. https://doi.org/10.1016/j.technovation.2021.102443

De Sa, M. M., Prim, A. L., & Birou, L. (2023). With major risks comes great resilience: The COVID-19 effect on SMEs in a developing country. *Operations Management Research, 16*(2), 1043–1055. https://doi.org/10.1007/s12063-023-00362-3

Dubey, R., Bryde, D. J., Dwivedi, Y. K., Graham, G., & Foropon, C. (2022). Impact of artificial intelligence-driven big data analytics culture on agility and resilience in humanitarian supply chain: A practice-

based view. *International Journal of Production Economics, 250*, Article 108618. https://doi.org/10.1016/j.ijpe.2022.108618

Duchek, S. (2020). Organizational resilience: A capability-based conceptualization. *Business Research, 13*(1), 215–246. https://doi.org/10.1007/s40685-019-0085-7

Eggers, F. (2020). Masters of disasters? Challenges and opportunities for SMEs in times of crisis. *Journal of Business Research, 116*, 199–208. https://doi.org/10.1016/j.jbusres.2020.05.025

Fodouop Kouam, A. W. (2025). Toward digital transformation: Insights into Chinese cross-border e-commerce SMEs during the COVID-19 pandemic and the post-pandemic era. *SAGE Open, 15*(1), Article 21582440251318792. https://doi.org/10.1177/21582440251318792

Gouveia, S., De La Iglesia, D. H., Abrantes, J. L., & López Rivero, A. J. (2024). Transforming strategy and value creation through digitalization? *Administrative Sciences, 14*(11), Article 307. https://doi.org/10.3390/admsci14110307

Grooss, O. F., Presser, M., & Tambo, T. (2022). Balancing digital maturity and operational performance: Progressing in a low-digital SME manufacturing setting. *Procedia Computer Science, 200*, 495–504. https://doi.org/10.1016/j.procs.2022.01.247

Hasayotin, K., Setthaji, R., Ratchatakulpat, T., Naburana, W., & Supanut, A. (2024). Empowerment of SMEs and entrepreneurial ecosystems: A qualitative study on diversifying Pattaya's economy. *Revista de Gestão Social e Ambiental, 18*(7), Article e05608. https://doi.org/10.24857/rgsa.v18n7-070

He, Z., Huang, H., Choi, H., & Bilgihan, A. (2023). Building organizational resilience with digital transformation. *Journal of Service Management, 34*(1), 147–171. https://doi.org/10.1108/JOSM-06-2021-0

Hernández Sánchez, N., & Oskam, J. (2025). A "new tourism cycle" on the Canary Islands: Scenarios for digital transformation and resilience of small and medium tourism enterprises. *Journal of Tourism Futures, 11*(1), 6–22. https://doi.org/10.1108/JTF-04-2022-0132

Hinings, B., Gegenhuber, T., & Greenwood, R. (2018). Digital innovation and transformation: An institutional perspective. *Information and Organization, 28*(1), 52–61. https://doi.org/10.1016/j.infoandorg.2018.02.004

Hoang, T. N. H. G., & Teo, H.-H. (2023). The impacts of digital platforms on SMEs' digital resilience: Network effects unveiled. In *Proceedings of the 29th Americas Conference on Information Systems (AMCIS 2023)*. Association for Information Systems. https://aisel.aisnet.org/treos_amcis2023/33

Hokmabadi, H., Rezvani, S. M. H. S., & De Matos, C. A. (2024). Business resilience for small and medium enterprises and startups by digital transformation and the role of marketing capabilities—A systematic review. *Systems, 12*(6), Article 220. https://doi.org/10.3390/systems12060220

Holl, A., & Rama, R. (2024). SME digital transformation and the COVID-19 pandemic: A case study of a hard-hit metropolitan area. *Science and Public Policy, 51*(6), 1212–1226. https://doi.org/10.1093/scipol/scae023

Kawane, T., Adu-Gyamfi, B., Cao, Y., Zhang, Y., Yamazawa, N., He, Z., & Shaw, R. (2024). MSME digitization, informality, and local economy: A study of the tourism sector in Phuket, Thailand. *Sustainability, 16*(4), 1550. https://doi.org/10.3390/su16041550

Khurana, I., Dutta, D. K., & Singh Ghura, A. (2022). SMEs and digital transformation during a crisis: The emergence of resilience as a second-order dynamic capability in an entrepreneurial ecosystem. *Journal of Business Research, 150*, 623–641. https://doi.org/10.1016/j.jbusres.2022.06.048

Kim, C.-B., Yang, H.-J., & Jeong, K.-W. (2025). Identifying and mitigating risks in the digital transformation of cold chain systems: An empirical study on South Korean SMEs. *Asia Pacific Journal of Marketing and Logistics*. Advance online publication. https://doi.org/10.1108/APJML-08-2024-1034

Kraus, S., Jones, P., Kailer, N., Weinmann, A., Chaparro-Banegas, N., & Roig-Tierno, N. (2021). Digital transformation: An overview of the current state of the art of research. *SAGE Open, 11*(3), Article 21582440211047576. https://doi.org/10.1177/21582440211047576

Kumar, V., Sindhwani, R., Behl, A., Kaur, A., & Pereira, V. (2024). Modelling and analysing the enablers of digital resilience for small and medium enterprises. *Journal of Enterprise Information Management, 37*(5), 1677–1708. https://doi.org/10.1108/JEIM-01-2023-0002

Lassen, A. H., & Larsen, M. S. S. (2024). Manufacturing innovation for Industry 4.0: An innovation capability perspective. *Journal of Manufacturing Technology Management, 36*(9), 19–44. https://doi.org/10.1108/JMTM-09-2023-0414

Latino, M. E. (2025). A maturity model for assessing the implementation of Industry 5.0 in manufacturing SMEs: Learning from theory and practice. *Technological Forecasting and Social Change, 214*, Article 124045. https://doi.org/10.1016/j.techfore.2025.124045

Mossberger, K., Martini, N. F., McCullough, M., & Tolbert, C. J. (2023). Digital economic activity and resilience for metros and small businesses during COVID-19. *Small Business Economics, 60*, 1699–1717. https://doi.org/10.1007/s11187-022-00674-x

Muhammad, S. S., Dey, B. L., Kamal, M. M., Samuel, L., & Alzeiby, E. A. (2025). Digital transformation or digital divide? SMEs' use of AI during global crisis.

Technological Forecasting and Social Change, 217, 124184. https://doi.org/10.1016/j.techfore.2025.124184

Papagiannidis, S., Harris, J., & Morton, D. (2020). Who led the digital transformation of your company? A reflection of IT related challenges during the pandemic. *International Journal of Information Management, 55*, Article 102166. https://doi.org/10.1016/j.ijinfomgt.2020.102166

Ragazou, K., Passas, I., & Sklavos, G. (2022). Investigating the strategic role of digital transformation path of SMEs in the era of COVID-19: A bibliometric analysis using R. *Sustainability, 14*(18), Article 11295. https://doi.org/10.3390/su141811295

Razavi Hajiagha, S. H., Alaei, S., Sadraee, A., & Nazmi, P. (2024). A perspective of international performance improvement concentrating on innovation and digital resilience of SMEs: The case of an emerging economy. *Journal of Enterprise Information Management, 37*(5), 1709–1736. https://doi.org/10.1108/JEIM-02-2023-0078

Reis, J. C. G., & Melão, N. (2023). Digital transformation: A meta-review and guidelines for future research. *Heliyon, 9*(1), e12834. https://doi.org/10.1016/j.heliyon.2023.e12834

Robertson, J., Botha, E., Walker, B., Wordsworth, R., & Balzarova, M. (2022). Fortune favours the digitally mature: The impact of digital maturity on the organisational resilience of SME retailers during COVID-19. *International Journal of Retail & Distribution Management, 50*(8/9), 1182–1204. https://doi.org/10.1108/IJRDM-10-2021-0514

Saezow, K., & Sukhabot, S. (2025). Synthesizing the resilience elements of small and medium enterprises (SMEs) amidst the COVID-19 pandemic crisis: A systematic review. *Cogent Business & Management, 12*(1), Article 2512823. https://doi.org/10.1080/23311975.2025.2512823

Senin, S. M., Juhdi, N. H., Che Omar, A. R., & Hashim, N. A. (2024). A systematic review of adaptation of IR 4.0 during COVID-19 pandemic among global SMEs. *Journal of Logistics, Informatics and Service Science, 11*(2), 61–83. https://doi.org/10.33168/JLISS.2024.0205

Sinha, K. K., Raby, S., & Salari, T. (2024). Exploring the scope and depth of digitalisation in times of crisis: Implications for SME resilience. *International Small Business Journal: Researching Entrepreneurship, 43*(3), 219–245. https://doi.org/10.1177/02662426241293000

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Sony, M., & Mekoth, N. (2022). Employee adaptability skills for Industry 4.0 success: A road map. *Production & Manufacturing Research, 10*(1), 24–41. https://doi.org/10.1080/21693277.2022.2035281

Succurro, M., & Donati, F. (2025). Digital transformation and SMEs: Evidence from European enterprises. *International Review of Law & Economics, 71*, 106263. https://doi.org/10.1016/j.irle.2025.106263

Teece, D. J. (2012). Dynamic capabilities: Routines versus entrepreneurial action. *Journal of Management Studies, 49*(8), 1395–1401. https://doi.org/10.1111/j.1467-6486.2012.01080.x

Teece, D. J. (2016). Dynamic capabilities and entrepreneurial management in large organizations: Toward a theory of the (entrepreneurial) firm. *European Economic Review, 86*, 202–216. https://doi.org/10.1016/j.euroecorev.2015.11.006

Tim, Y., & Leidner, D. E. (2023). Digital resilience: A conceptual framework for information systems research. *Journal of the Association for Information Systems, 24*(5), 1184–1198. https://doi.org/10.17705/1jais.00842

Tubis, A. A. (2023). Digital maturity assessment model for the organizational and process dimensions. *Sustainability, 15*(20), Article 15122. https://doi.org/10.3390/su152015122

United Nations. (2024). *Global micro-, small and medium-sized enterprises, Report*. https://www.un.org/sites/un2.un.org/files/globalmsmesreport2024.pdf

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), xiii–xxiii. http://www.jstor.org/stable/4132319

Wiesböck, F., & Hess, T. (2020). Digital innovations: Embedding in organizations. *Electronic Markets, 30*(1), 75–86. https://doi.org/10.1007/s12525-019-00364-9

Yang, F., & Panichakarn, B. (2025). The role of digital transformation and innovation in enhancing resilience and competitiveness of Chinese logistics SMEs. *Decision Science Letters, 14*(3), 793–806. https://doi.org/10.5267/j.dsl.2025.3.002

Yun, J., & Park, J. (2022). The effects of chatbot service recovery with emotion words on customer satisfaction, repurchase intention, and positive word-of-mouth. *Frontiers in Psychology, 13*, Article 922503. https://doi.org/10.3389/fpsyg.2022.922503

Zhang, K., Wang, J., & Wu, Y. (2025). A study of the impact of manufacturing input digitization on firms' organizational resilience: Evidence from China. *Sustainability, 17*(3), Article 897. https://doi.org/10.3390/su17030897

Zirar, A., Jabbar, A., Njoya, E., & Amoozad Mahdiraji, H. (2024). Smart contract challenges and drawbacks for SME digital resilience. *Journal of Enterprise Information Management, 37*(5), 1527–1550. https://doi.org/10.1108/JEIM-02-2023-0082

Potentials and Barriers of Using Big Data in Austrian SMEs

Walter Rath, Berndt Jesenko, Harris Gerzic
CAMPUS 02, University of Applied Sciences
Department IT & Business Informatics
Koerblergasse 126, 8010 Graz, Austria
walter.rath@lv.campus02.at,{berndt.jesenko,harris.gerzic}@campus02.at

Abstract. *This study explores the potentials and barriers of implementing Big Data in Austrian Small and Medium-sized Enterprises (SMEs). Based on expert interviews and qualitative content analysis, it reveals key opportunities such as improved efficiency, data-driven decision-making, and regulatory compliance. However, challenges remain in technological integration, financial resources, legal issues, and organizational change. The study offers practical recommendations and highlights the need for further research, contributing to a better understanding of Big Data as a strategic tool for strengthening SME competitiveness.*

Keywords. Big Data, SMEs, potentials, barriers, qualitative content analysis

1 Introduction

Big Data refers to an ever-growing, diverse, and dynamic volume of data that can hardly be managed using traditional data processing systems (Furht & Villanustre, 2016: 3; Zikopoulos et al., 2012: 10). Originally defined by the three dimensions of volume, variety, and velocity, the dimensions of value and veracity have since been added to better capture the potential benefits and data quality (Meier, 2017: 6–7). The challenges range from processing unstructured data in real time, to integration issues, and legal concerns such as data protection (Gadatsch & Landrock, 2017: 6; Furht & Villanustre, 2016: 5–6). Vossen et al. (2015: 3–4) distinguish four core perspectives for examining Big Data: technological, organizational, legal, and economic dimensions.

In both academic and business practice, Big Data is increasingly regarded as a strategic success factor—particularly with regard to data-driven innovation, automation potential, and improved decision-making. While large enterprises have already explored a variety of application scenarios, SMEs often face particular challenges: limited human, financial, and technical resources meet high demands for data literacy, data protection compliance, and infrastructure. At the same time, SMEs could benefit disproportionately from data-driven approaches—such as through more efficient processes, personalized customer engagement, or new business models.

Against this backdrop, the question of how Big Data can be successfully implemented in SMEs is gaining considerable relevance. Identifying industry-specific potentials and practical barriers is a key prerequisite for developing and sustainably implementing data-driven strategies. Focusing on Austrian SMEs also makes it possible to take country-specific conditions—such as regulatory requirements, digital infrastructure, or funding landscapes—into account. The Research Question reads as follows:

"What potentials and barriers does the implementation of Big Data present for Austrian SMEs?"

This study examines the opportunities and barriers associated with the use of Big Data in Austrian SMEs. The objective is to derive practical recommendations through a systematic analysis that supports the successful adoption and utilization of Big Data. In addition to theoretical insights, practical experiences are incorporated to comprehensively capture the potential of Big Data for SMEs and to outline future strategic directions.

2 Background

The Big Data flow describes the entire process from data collection to the delivery of actionable insights. According to Furht & Villanustre (2015: 5), it comprises six key phases: collection, ingestion, cleansing, integration, analysis, and delivery. This workflow forms the backbone of data-driven applications and supports companies in effectively utilizing their data resources.

Data has become a key strategic resource today (Fasel & Meier, 2016: 13–15). Well-structured data management involves an appropriate data architecture, modern database systems such as NoSQL (e.g., MongoDB, Cassandra), and powerful analytical methods. NoSQL systems are optimized for large, distributed, and flexible data sets (Zomaya & Sakr, 2017: 17), while Hadoop enables distributed

processing across clusters using the MapReduce approach (Klein et al., 2013: 322).

The CAP theorem (Brewer, 2002) outlines a fundamental dilemma of distributed systems: consistency, availability, and partition tolerance cannot all be fully guaranteed at the same time (Meier, 2017: 13–14). For Big Data infrastructures, it is essential to consciously manage these trade-offs.

Big Data Analytics (BDA) extends traditional data analysis by incorporating various analytical approaches:

- Descriptive analytics answers: What has happened?
- Diagnostic analytics explains: Why did it happen?
- Real-time analytics reveals: What is happening right now?
- Predictive analytics forecasts: What could happen?
- Prescriptive analytics recommends: What should be done? (Dorschel, 2015; Schön, 2018)

These approaches enable companies to make data-driven decisions at different levels of complexity and to act more strategically.

BDA offers a wide range of opportunities, especially for SMEs. Due to their agility and flatter hierarchies, they are able to implement data-driven business models more quickly (Maroufkhani et al., 2023). In marketing, for example, customer interactions can be analyzed to create precisely tailored, individualized offers (Berger et al., 2014: 46). In production, the analysis of machine data leads to greater efficiency and quality, while in research and development, innovation cycles can be shortened (BITKOM, 2012: 37). BDA also provides valuable decision-making foundations and forecasting capabilities in logistics and management (Mantri & Mishra, 2023: 11).

Despite the potential, numerous barriers remain. From a technological perspective, poor data quality, limited data access, and interface integration issues complicate implementation (Sharma et al., 2021: 2). On the organizational level, unclear responsibilities, a lack of strategic planning, and low acceptance hinder progress (Alaimo et al., 2020).

Data protection presents a particular challenge in the SME context—the General Data Protection Regulation (GDPR) requires transparent processes, data security, and, where applicable, the anonymization of sensitive data (Karaboga, 2022: 277). Another significant obstacle is the shortage of skilled professionals in data analysis and IT—especially for smaller companies with limited resources (Maroufkhani et al., 2023).

Emerging technological advancements offer new opportunities for SMEs:

- Artificial Intelligence enhances analysis speed and decision-making (Aggarwal et al., 2021).
- Edge Computing brings data processing closer to the source—ideal for IoT and manufacturing (Satyanarayanan, 2017).
- Blockchain ensures data security and traceability (Hellani et al., 2021).
- Quantum Computing will enable the resolution of highly complex problems in the long term (Montanaro, 2016).
- Privacy-Enhancing Technologies (PETs) such as Federated Learning or Differential Privacy allow for data protection-compliant analyses (Shokri & Shmatikov, 2015).

These trends illustrate that BDA is not just a tool for process optimization in SMEs, but remains a key driver of innovation and competitiveness.

3 SMEs in Austria in the Context of Big Data

According to the EU definition, SMEs are classified as enterprises with fewer than 250 employees and either an annual turnover of up to 50 million euros or a balance sheet total of up to 43 million euros (EU Commission 2003/361/EC). This definition is also applied in Austria and, to a large extent, in Switzerland. Typical characteristics of SMEs include close personal management, low formalization, high flexibility, and strong customer proximity (Mugler, 2008: 31–32).

Compared to large enterprises, SMEs differ not only in size and revenue but also in structure, management, financing, and organization (Pfohl, 2013). SMEs often operate with flat hierarchies, short decision-making paths, and personalized market approaches—factors that provide agility on the one hand, but also imply limited resources on the other.

Economic Influence. SMEs account for approximately 99% of all businesses in the DACH region and are responsible for over 60% of total employment. In Austria, they contribute around two-thirds of total employment and play a crucial role in economic performance and innovation capacity (Pfohl, 2013: 49; Mugler, 2008: 43).

Areas of Application for Big Data. Big Data offers SMEs opportunities for increased efficiency, customization, and process improvement across many areas (Dorschel, 2015: 104). In marketing, it enables precise customer segmentation, targeted campaigns, and personalized offerings (BITKOM, 2012: 36). In research and development, Big Data shortens innovation cycles through faster data analysis and idea validation (BITKOM, 2012: 37). In production and logistics, it supports process optimization and predictive maintenance. In the financial sector, Big Data enables scenario planning, fraud detection, and real-time analysis. In IT, it enhances security, performance, and resource efficiency.

Use in SMEs. Despite its great potential, the use of Big Data in SMEs remains limited. Reasons include the lack of IT infrastructure, limited human resources, and the assumption that Big Data is only relevant for large enterprises (Wang et al., 2016: 11). Many SMEs also

underestimate the value of their own data. Nevertheless, it is evident that data-driven solutions can significantly enhance efficiency, innovation capacity, and competitive position even for smaller companies.

Status Quo. According to Deloitte (2014), 87% of the surveyed SMEs expect a significant increase in data volume. Nevertheless, according to Seufert (2014), only 15% have gained practical experience with Big Data. The Data4KMU study (2021) shows that while many companies recognize the importance of Big Data, they have yet to develop concrete strategies. There is a clear need to catch up in the implementation of data-driven approaches.

Opportunities and Potential. Big Data can enhance the innovative capacity, productivity, and customer loyalty of SMEs (Hung, 2016; Markl et al., 2013). Integrating it into existing systems such as ERP solutions or leveraging external data sources (e.g., social media) can lead to new insights and business models (Gadatsch & Landrock, 2017: 14). Predictive analytics enables early market responses, and data-driven decisions improve the quality of strategic actions (BITKOM, 2012: 14).

Barriers. Despite these opportunities, SMEs face numerous obstacles: lack of technical know-how, high costs, missing concepts, and a low sense of urgency are cited as the main barriers (Seufert, 2014; Data4KMU, 2021). Technological challenges such as data silos, data quality, or data protection, as well as organizational issues like change management, complicate implementation (Markl et al., 2013; Vossen et al., 2015).

Big Data and Privacy. Handling personal data is legally particularly sensitive. The EU General Data Protection Regulation (GDPR) requires transparent, secure, and legally compliant data processing (Article 6 GDPR). Companies must uphold user rights, protect sensitive data, and provide a lawful basis for data processing. Violations can result in reputational damage and substantial fines.

4 Methodology

To answer the research question, a qualitative, exploratory approach was chosen. The core basis consisted of semi-structured expert interviews guided by a predefined framework based on Big Data potentials and barriers identified in the literature. The four analytical categories—technical, organizational, economic, and legal—structured both the interview guide and the subsequent evaluation.

Interviews enable in-depth insights into individual experiences and company-specific conditions (Dresing & Pehl, 2018). The chosen semi-structured format allows for flexibility in the flow of conversation while ensuring comparability of responses (Bortz & Döring, 2006: 218). The 10 interviewees represent a diverse group of professionals in leadership and digitalization roles across various industries. They include Chief Digital Officers, managing directors, and IT executives from sectors such as IT services, insurance, CAD software development, civil engineering, advertising, construction, water treatment, fitness, waste management, and the church sector. This range ensures insights from both traditional and digitally advanced industries.

The qualitative content analysis according to Mayring (2015) serves as the methodological framework for the systematic evaluation of the interviews. The content is categorized deductively along the four main dimensions, complemented by inductively developed subcategories. The approach includes defining units of analysis, developing a category system, and gradually reducing the material to extract key statements. The aim is a theory-driven and transparent interpretation of the interview data.

Based on the developed set of criteria, key potentials such as strategic advantages, process optimization, or customer retention were considered, along with typical barriers (e.g., data protection, lack of resources, or expertise) (Markl et al., 2013; King, 2014).

In addition, practical recommendations were derived from the literature. These include partnerships with Big Data providers, the use of existing process data, targeted training, and the standardization of legal templates to ensure data protection compliance (Gadatsch & Landrock, 2017). These recommendations were incorporated into the development of the interview guide and the discussion of results.

Finally, the author reflects on three key strategic questions from an SME perspective:

(1) Where can Big Data be usefully applied within the company?

(2) What internal prerequisites must be met?

(3) What data is available, and how can it be processed?

5 Results

The qualitative content analysis reveals that Austrian SMEs exhibit widely varying levels of experience and development paths in dealing with Big Data. It becomes clear that the implementation and use of data-driven technologies are highly dependent on the specific industry, business objectives, and the available technical and human resources. While some companies are already strategically relying on data-driven decision-making processes, others are still in an exploratory or preparatory phase. The following sections summarize the key findings according to the categories “Industry-specific expertise,” “Company profile,” “Areas of application,” and “Potentials,”.

Industry-Specific Expertise. The industry-specific nature of how Big Data is handled became particularly evident in the interviews. For example, in a church

organization, the entry into data processing began with the digitization of historical parish registers. The aim was not only to make administrative processes more efficient but also to create a basis for academic analysis. Although it is a rather traditional institution, it became clear that Big Data is also seen there as a tool for structural modernization.

In contrast, in the field of medical technology, Big Data technologies are already being used for advanced diagnostic processes. One company reported the targeted use of artificial intelligence for pattern recognition in medical data, which led to a significant improvement in diagnostic quality. Here, Big Data is not merely considered a supporting element but a central component of product development.

In the construction industry, the application is structured differently. A property developer described how external data sources are analyzed to better understand buyer profiles and to tailor real estate projects to specific target groups. The analysis helps to avoid poor investments and anticipate market developments at an early stage. In the advertising industry, on the other hand, an operational perspective dominates. Real-time analytics are used to immediately evaluate and control advertising campaigns. This ability for ongoing optimization is seen as essential for precisely addressing customer needs and minimizing waste coverage.

Concrete applications were also evident in the waste management sector. By integrating sensors to monitor container fill levels, more efficient route planning was achieved. This not only led to cost savings but also to improved customer satisfaction. Overall, it is clear that the level of expertise, industry-specific challenges, and objectives play a key role in determining how extensively Big Data has already been integrated.

Company Profile. The statements of the interviewees show that the use of Big Data varies not only by industry but also by individual company. In the real estate sector, for example, the strategic analysis of external data sources for better targeting was described as a key competitive advantage. By linking this data with existing customer information, it becomes possible to develop new projects more precisely and to improve market access efficiency.

In the advertising industry, the importance of real-time data was once again emphasized. The ability to immediately evaluate and optimize ongoing campaigns was, according to one respondent, critical to economic success. This involves not only technical performance but also the interpretation and evaluation of the collected data.

A company in the waste management sector highlighted how data-driven decisions contribute to resource efficiency. Automated systems have reduced the effort required for route planning while also improving the company's environmental footprint. In the medical technology sector, a strong focus on process optimization was also evident. AI-supported applications enable continuous quality assurance in diagnostics and, according to the interviewee, are considered pioneering for the entire business model.

These examples demonstrate that Big Data contributes in various ways to improving efficiency, quality, or the achievement of strategic objectives—depending on the company's level of maturity and strategic focus.

Areas of Application

The range of possible Big Data applications in SMEs is as diverse as it is practical. This became particularly evident in the waste management sector, where sensor technology is used to monitor container fill levels in order to optimize route planning. This not only results in time and fuel savings but also improves service quality and the company's environmental footprint.

In the medical technology sector, the focus lies on analyzing medical data to enhance diagnostic accuracy. By identifying specific patterns, error rates were reduced and diagnoses became more precise, ultimately benefiting patient care. One interviewee emphasized that this technology will also play a key role in preventive approaches and personalized medicine in the future.

Even institutions such as churches are now recognizing the potential of data-driven strategies. By digitizing historical records, more targeted fundraising campaigns can be developed. The deliberate use of digital communication tools—based on existing data—significantly increases the effectiveness of such initiatives.

In the real estate sector, data-driven planning processes have also become established. The evaluation of external data sources makes it possible to identify demographic trends and market developments early on. As a result, projects can be tailored more precisely to target groups, which not only offers economic advantages but also enables more efficient use of resources.

In the advertising industry, real-time optimization is at the forefront. The ability to react immediately to changes in customer behavior not only improves the efficiency of marketing activities but also significantly strengthens customer loyalty.

The interviews demonstrate that Big Data is no longer merely a technical tool but is increasingly being seen as a strategic asset that can sustainably influence the competitiveness of small and medium-sized enterprises across almost every sector title.

6 Conclusion

This paper set out to explore the question: "What potentials and barriers does the implementation of Big Data present for Austrian SMEs?" The findings reveal that Big Data offers a wide range of strategic and operational opportunities for SMEs—such as improved efficiency, real-time decision-making, and

innovation potential—while also confronting them with substantial challenges in terms of infrastructure, data protection, and organizational readiness. The analysis of expert interviews across various industries showed that the degree of Big Data adoption varies significantly depending on sector, company size, and available resources. These insights formed the basis for the concrete technological, legal, and organizational recommendations presented in the previous chapter, which constitute the main contribution of this study to both research and practice.

From a technological perspective, SMEs are gaining new room to maneuver—through AI, cloud computing, or automated analytics. These technologies enable faster processing of large volumes of data and open up space for strategic decision-making. The advertising sector, for instance, uses tools like BigQuery for real-time campaign analysis, while construction companies perform external data comparisons for targeted audience planning. Economically, data-driven optimizations can make investments more efficient and demonstrably increase return on investment. From a legal standpoint, automated documentation, encrypted data storage, and GDPR-compliant systems offer new possibilities for ensuring regulatory standards. However, the legal framework remains a major challenge for many SMEs. In particularly sensitive fields such as medical technology or marketing, comprehensive compliance measures and regular audits are required. Finally, from an organizational perspective, it is evident that data-driven processes demand profound changes in corporate culture—both at the leadership level and among employees. Acceptance, training, and participative change management are essential success factors in this regard.

This paper provides valuable insights into the potentials and barriers of Big Data implementation in Austrian SMEs. However, the results must be viewed in light of certain methodological and conceptual limitations, which also offer starting points for further research and practical development. The qualitative content analysis is based on ten semi-structured interviews with experts from various industries. Despite the broad industry distribution, the generalizability of the findings is limited. Subjective assessments, varying interview styles, and the small sample size may influence the validity of the results. Additionally, the interviewees' statements are strongly shaped by individual experiences and organizational contexts—and thus cannot be easily transferred to all SMEs. Moreover, Big Data is a highly dynamic field that is subject to constant technological and regulatory changes. The tools, challenges, and solutions discussed in this paper reflect a specific point in time and require continuous updating. The focus on the Austrian context—with its specific legal, economic, and structural conditions—also limits the transferability of the findings to other countries or regions.

Despite these limitations, the paper opens up a variety of promising approaches for future research. A broader quantitative study with a larger sample could systematically capture patterns and correlations—for example, between Big Data usage and business success. Industry-specific analyses could also help develop more differentiated recommendations, such as for the healthcare, construction, or logistics sectors. Furthermore, longitudinal studies would be useful to observe technological and organizational developments over time and to identify sustainable success factors. Another line of research concerns the integration of new technologies such as AI, cloud, or edge computing into SMEs. Understanding how these tools can be adapted, scaled, and used economically is of high practical relevance. Interdisciplinary perspectives are also essential: the interplay between technological implementation, legal requirements, and human acceptance offers rich potential for economic, legal, and social science approaches. For practice, the paper leads to a clear conclusion: Big Data is not a passing technological trend but a strategic tool that can significantly shape the future viability of SMEs. The potentials range from more efficient resource utilization and optimized market analysis to compliance with regulatory requirements. At the same time, successful implementation requires structural adjustments, cultural change, investments in know-how, and a professional approach to data protection. SMEs should not view Big Data as an isolated IT project, but as part of a broader transformation process. Building technical infrastructure, establishing data governance, and training employees are not one-off tasks, but part of continuous strategic development. Only if technological, economic, legal, and cultural requirements are aligned can Big Data fully unfold its potential—as a key technology for innovation, efficiency, and competitiveness in the SME sector.

References

Aggarwal, C. C. (2021). Machine Learning for Text, 7, 9.

Alaimo, C., Kallinikos, J., & Aaltonen, A. (2020). Data and value. In Handbook of digital innovation (pp. 162-178). Edward Elgar Publishing.

Berger, H., Dittenbach, M., Haas, M., Bierig, R., Hanbury, A., Lupu, M. & Piroi, F. (2014). Conquering Data in Austria: Technologie-Roadmap für das Programm IKT der Zukunft: Daten durchdringen - Intelligente Systeme. Wien.

BITKOM. (2012). Big Data im Praxiseinsatz: Szenarien, Beispiele, Effekte. Berlin.

Bortz, J. & Döring, N. (2006). Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler (4. Aufl.). Springer.

Data4KMU. (2021). Data Science für KMU leicht gemacht: Aktuelle Erkenntnisse und Lösungen. Zürich. Abgerufen am 19. April 2024, von bzi40.eu/informationen/publikationen/studien/382 abschlussbericht-data4kmu/file

Deloitte. (2014). Data Analytics im Mittelstand: Die Evolution der Entscheidungsfindung.

Dorschel, J. (Hrsg.). (2015). Praxishandbuch Big Data: Wirtschaft - Recht - Technik. Springer Gabler.

Empfehlung der Kommission vom 6. Mai 2003 betreffend die Definition der Kleinstunternehmen sowie der kleinen und mittleren Unternehmen. In Amtsblatt der Europäischen Union, L 124/36, 20.5.2003.

Europäische Kommission. (2021a). Digital Economy and Society Index (DESI) 2021: Austria In:Europäische Kommission. Abgerufen am 19. März 2024. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/desi-austria.

Europäische Kommission. (2021b). 2021 SME Country Fact Sheet Austria In: Europäische Kommission. Abgerufen am 19. März 2024, von https://ec.europa.eu/growth/smes/smestrategy/sme-performance-review_en.

Fasel, D. & Meier, A. (Hg.). (2016). Edition HMD. Big Data: Grundlagen, Systeme und Nutzungspotenziale (1. Aufl.). Springer Vieweg.

Furht, B. & Villanustre, F. (2016). Big Data Technologies And Applications (1. Aufl.). Springer International Publishing.

Gadatsch, A. & Landrock, H. (2017). Big Data für Entscheider: Entwicklung und Umsetzung datengetriebener Geschäftsmodelle. essentials. Springer Vieweg.

Hellani, H., Sliman, L., Samhat, A. E., & Exposito, E. (2021). On blockchain integration with supply chain: Overview on data transparency. Logistics, 5(3), 46.

Huber, M. & Köhler, M. (2014). Best Practice für Big Data Projekte. Wien.

Karaboga, M. (2022). Datenschutzrechtliche Gestaltungsmöglichkeiten jenseits der Ermächtigung des Individuums: Die Multi-Stakeholder-Datenschutz-Folgenabschätzung. In Selbstbestimmung, Privatheit und Datenschutz: Gestaltungsoptionen für einen europäischen Weg (pp. 275-302). Wiesbaden: Springer Fachmedien Wiesbaden.

King, S. (2014). Big Data. Springer Fachmedien Wiesbaden.

KMU Forschung Austria. (2021). KMU im Fokus 2020: Bericht über die Situation und Entwicklung kleiner und mittlerer Unternehmen der österreichischen Wirtschaft. Retrieved from https://kmu.unisg.ch/de/forschung-und-publikationen/kmu-zahlen.

Klein, D., Tran-Gia, P. & Hartmann, M. (2013). Big Data. Informatik-Spektrum, 36(3), 319–323. https://doi.org/10.1007/s00287-013-0702-3

Kuckartz, U., Dresing, T., Rädiker, S. & Stefer, C. (2022). Qualitative Evaluation: Der Einstieg in die Praxis (2., aktualisierte Auflage). VS Verlag für Sozialwissenschaften.

Markl, V., Hoeren, T. & Krcmar, H. (2013). Innovationspotenzialanalyse für die neuen Technologien für das Verwalten und Analysieren von großen Datenmengen. Berlin. MAXQDA. (2021, 30. November).

Mayring, P. (2015). Qualitative Inhaltsanalyse: Grundlagen und Techniken (12. Aufl.). Beltz Verlag.

Mayring, P., & Fenzl, T. (2016). Qualitative content analysis program qcamap–an open access text analysis software. In 15th Biennial EARLI Conference for Research on Learning and Instruction, Munich, Germany.

Maroufkhani, P., Iranmanesh, M., & Ghobakhloo, M. (2023). Determinants of big data analytics adoption in small and medium-sized enterprises (SMEs). Industrial Management & Data Systems, 123(1), 278-301.

Mantri, A., & Mishra, R. (2023). Empowering small businesses with the force of big data analytics and AI: A technological integration for enhanced business management. The Journal of High Technology Management Research, 34(2), 100476.

Meier, A. (2017). Werkzeuge der digitalen Wirtschaft: Big Data, NoSQL & Co. : Eine Einführung in relationale und nicht-relationale Datenbanken (1. Aufl.). essentials. Springer Vieweg.

Microsoft. (2024, 24. April). Übersicht über Online Analytical Processing (OLAP). Abgerufen am 24. April 2024, von https://support.microsoft.com/de-de/office/%C3%BCbersicht-%C3%BCberdie-analytische-onlineverarbeitung-olap-15d2cdde-f70b-4277-b009-ed732b75fdd6.

Montanaro, A. (2016). Quantum algorithms: an overview. npj Quantum Information, 2(1), 1-8.

Mugler, J. (2008). Grundlagen der BWL der Klein- und Mittelbetriebe. Manual. WUV-Univ.- Verl. https://permalink.obvsg.at/fwg/AC04698461.

Pfohl, H.-C. (2013). Betriebswirtschaftslehre der Mittel- und Kleinbetriebe: Größenspezifische Probleme und Möglichkeiten zu ihrer Lösung (5. Aufl.). Management und Wirtschaft Praxis: Bd. 44. Schmidt.

Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30-39.

Seufert, A. (2014). Entwicklungsstand, Potenziale und zukünftige Barrieren von Big Data – Ergebnisse einer empirischen Studie. HMD Praxis der Wirtschaftsinformatik, 51(4), 412–423. https://doi.org/10.1365/s40702-014-0039-7

Schön, D. (2018). Planung und Reporting im BI-gestützten Controlling: Grundlagen, Business Intelligence, Mobile BI und Big-Data-Analytics (3., erweiterte Auflage). Springer Gabler.

Sharma, K., Mahesh, T. R., & Bhuvana, J. (2021, August). Big data technology for developing learning resources. In Journal of Physics: Conference Series (Vol. 1979, No. 1, p. 012019). IOP Publishing.

Shokri, R., & Shmatikov, V. (2015, October). Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC conference on computer and communications security (pp. 1310-1321).

Statistik Austria. (2025). Anteil der Online-Käufer an der österreichischenBevölkerung von 2003 bis 2021. Retrieved from https://de.statista.com/statistik/daten/studie/298302/umfrage/nutzung-von-online- shopping-inoesterreich/.

Vossen, G., Lechtenbörger, J. & Fekete, D. (2015). Big Data in kleinen und mittleren Unterneh- men: Eine empirische Bestandsaufnahme. Arbeitsberichte des Instituts für Wirtschaftsinfor- matik (Nr. 135). Münster. http://hdl.handle.net/10419/112719

Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International journal of production economics, 176, 98-110.

Zikopoulos, P., Eaton, C., deRoos, D., Deutsch, T. & Lapis, G. (2012). Understanding big data: Analytics for enterprise class Hadoop and streaming data. McGraw-Hill.

Zomaya, A. Y. & Sakr, S. (2017). Handbook of big data technologies. Springer.

Cultural Factors that Inhibit or Facilitate Innovation in Family Businesses: A Systematic Literature Review

Idoia de Paz Gómez

UIC Barcelona (Universitat Internacional de Catalunya)

Doctoral Program in Economics and Law

Carrer de la Immaculada, 22, Sarrià-Sant Gervasi, 08017 Barcelona, Spain

idoiadepaz@uic.es

Abstract. *This systematic literature review rigorously examines published research focused on organizational culture, innovation, and family business through analysis of 47 academic publications and 12 foundational theoretical works spanning 1950-2024. The review addresses the conceptual evolution of organizational culture and its link with innovation processes to identify cultural factors that act as facilitators or inhibitors of innovation, with particular attention to family business contexts. These organizations present distinctive characteristics in terms of power structures, value systems, and relational dynamics, making it necessary to examine how such cultural characteristics influence their capacity to innovate.*

The study employs qualitative thematic synthesis using the Competing Values Framework and established innovation-culture taxonomies as analytical foundation. Rather than conducting an exhaustive analysis of each concept, this review identifies the main theoretical frameworks and empirical findings that enable understanding of culture-innovation relationships in family businesses.

Key findings include: (1) development of a comprehensive taxonomy distinguishing facilitating factors (leadership support, risk tolerance, open communication, long-term vision) from inhibiting factors (resistance to change, risk aversion, centralized decision-making, succession uncertainty); (2) identification of family-specific cultural dynamics including socioemotional wealth preservation effects and generational transition impacts; (3) evidence of conceptual evolution from general organizational culture theory toward family business specialization; (4) recognition of both universal organizational factors and context-specific dimensions that require differentiated management approaches.

This systematic synthesis contributes to family business and innovation management literature by providing a structured understanding of cultural factors affecting innovation, establishing theoretical foundations that can inform future empirical research and organizational practice in family business innovation management.

Keywords. culture, organizational culture, innovation, and family business

1 Introduction

Since its emergence and popularization in the 1980s, the concept of organizational culture has become one of the most influential terms in management research and practice. The concept refers to the set of shared values, beliefs, behaviors, and practices that influence how members of an organization interact with each other and with the outside world. These cultural elements can largely determine how decisions are made, how leadership is exercised, and how challenges are faced within a company, including innovation processes.

The literature highlights that organizational culture is one of the main organizational elements that influence innovation, being able to favor or hinder it (Naranjo-Valencia, 2011). Despite the broad recognition of organizational culture as a key factor for innovation, historically most empirical studies have been theoretical and geographically limited, focused primarily on the United States (Naranjo-Valencia et al., 2012; Ahmed, 1998; Martins & Terblanche, 2003). This geographical limitation has been identified by various authors as a significant gap in the literature (Calabrò et al., 2019; De Massis et al., 2015)

In the realm of family business, this relationship becomes even more complex. Family businesses often possess organizational cultures intensely influenced by foundational values and family dynamics. While this specificity can promote strong internal cohesion and commitment to long-term vision, it can also limit openness to change if tradition is excessively privileged (Zahra et al., 2004).

Family businesses demonstrate significant capabilities for innovation, especially those with larger size and professionalized orientation; therefore, their potential for improvement in this area is considerable (EY & IE Center for Families in Business, 2022; Duran et al., 2016). However, they present heterogeneous innovation patterns according to their size, sector, and

organizational configuration (Zellweger, 2017; Calabrò et al., 2019)

This systematic literature review aims to examine the evolution of the concept of organizational culture and its link with innovation to identify cultural factors that act as facilitators or inhibitors of innovation in family organizations.

The aim of this paper is to analyse how the academic literature has addressed the evolution of the concepts of culture, organisational culture and innovation, as well as the relationship between them, in order to determine which cultural factors have been identified by different researchers as facilitators or inhibitors of innovation in family organisations.

This study poses the following research questions:

Q1. How have the concepts of culture, organizational culture, and innovation evolved and how do they relate to each other in a business context?

Q2. What cultural factors or characteristics are facilitators and which are inhibitors of innovation in organizations?

2 Methodology

2.1 Methodological justification

A systematic literature review was selected based on three specific criteria: (1) the research questions require comprehensive synthesis of existing theoretical and empirical knowledge across organizational behavior, innovation management, and family business studies; (2) systematic methodology ensures transparency and replicability in literature identification, selection, and analysis processes; (3) this approach enables development of a comprehensive taxonomy of cultural factors essential for understanding family business innovation dynamics.

This methodology follows established protocols for systematic reviews in management research (Codina, 2018) and is appropriate for exploratory studies examining interdisciplinary relationships such as culture-innovation dynamics in family business contexts.

2.2 Data collection

The following research procedures were implemented during this study

2.2.1 Search Strategy and Information Sources

Literature search was conducted across recognized academic databases ensuring comprehensive coverage:

- **Primary databases**: Scopus, Web of Science, Google Scholar
- **Supplementary databases**: Dialnet, Sage Journals
- **Institutional sources**: EY Global Family Business Survey, Instituto de la Empresa Familiar reports
- **Search terms**: " culture, organizational culture, innovation, and family business
- **Temporal scope**: 1950-2024, enabling comprehensive analysis of conceptual evolution from foundational works to contemporary research
- **Language criteria**: English and Spanish publications

2.2.2 Selection Criteria

Inclusion and exclusion criteria were systematically applied

Inclusion Criteria:

- Studies examining organizational culture-innovation relationships
- Peer-reviewed articles with empirical or theoretical contributions
- Full-text accessibility
- Publications in English or Spanish

Exclusion Criteria:

- Opinion pieces without empirical support
- Inaccessible full-text publications
- Individual-level creativity studies without organizational context
- Non-peer-reviewed conference abstracts

2.2.3 Search Results and Study Selection

- **Literature Search Outcomes:** The systematic search across multiple databases yielded over 200 potentially relevant publications stored in the Mendeley reference management system. This extensive collection encompassed studies on organizational culture, innovation management, and family business research from the specified temporal range (1950-2024).
- **Selection and Refinement Process:** From this comprehensive collection, a focused selection process was implemented to identify the most relevant and high-quality studies for systematic analysis. The selection prioritized studies that: explicitly examined culture-innovation relationships, provided empirical or theoretical contributions, demonstrated methodological rigor and conceptual clarity and offered novel insights into cultural factors affecting innovation.
- **Final Corpus Development:** Through iterative review and refinement, 47

contemporary studies (2010-2024) were selected as the primary analytical corpus, supplemented by 12 foundational theoretical works essential for conceptual grounding. This selection represents the most relevant and influential contributions to understanding culture-innovation dynamics in family business contexts.

2.2.4 Data Extraction and Analysis

Analytical Approach. Rather than mechanical frequency counting, this study employed qualitative thematic synthesis to identify patterns and relationships across the selected literature. This approach is particularly appropriate for exploratory research examining complex interdisciplinary relationships (Tranfield et al., 2003).

Theoretical Framework Application. Cultural factors were identified and categorized using established theoretical frameworks:

- **Primary framework**: Competing Values Framework (Cameron & Quinn, 1999)
- **Supporting taxonomies**: Naranjo-Valencia et al. (2010, 2012) innovation-culture classifications
- **Validation reference**: Martins & Terblanche (2003) creativity-supporting culture dimensions

Factor Synthesis Process:

- Systematic extraction of cultural dimensions from each selected study
- Thematic categorization based on reported innovation outcomes
- Binary classification into facilitating and inhibiting factors
- Cross-validation with established theoretical frameworks
- Integration of findings to develop comprehensive taxonomy

Quality and Relevance Assessment:

Selection Criteria Application. Studies included in the final analysis demonstrated:

- Theoretical rigor: Clear conceptual frameworks and literature integration
- Methodological adequacy: Appropriate research design and analytical approach
- Empirical contribution: Novel insights or significant findings
- Relevance: Direct connection to culture-innovation relationships

Focus on Quality over Quantity. The selection emphasized depth and relevance rather than exhaustive inclusion. The 47-study corpus represents the most significant and methodologically sound contributions to the research domain, providing sufficient foundation for comprehensive theoretical synthesis.

2.2.5 Methodological Limitations

- **Selection subjectivity**: Single-researcher selection process may introduce bias
- **Scope limitations**: Focus on most influential studies may exclude emerging perspectives
- **Geographic bias**: Predominance of Western research contexts limits global generalizability
- **Language restrictions**: English and Spanish publications only

3 Literature Review

3.1 Evolution of Organizational Culture

3.1.1 Historical Foundations of Culture Studies

The concept of culture constitutes a fundamental pillar in social sciences, with definitions that have evolved considerably across disciplines and theoretical perspectives. The systematic examination of cultural definitions was pioneered by Kroeber and Kluckhohn (1952), whose seminal work *Culture: A critical review of concepts and definitions* analyzed 164 definitions from anthropological, sociological, and psychological literature. Their comparative analysis characterized culture as socially acquired knowledge, habits, and dispositions, emphasizing its learned, collective, and transmissible nature. This foundational work established culture as a universal human phenomenon manifesting through contextually specific variations (Kroeber & Kluckhohn, 1952).

Contemporary scholarship conceptualizes culture as a system of shared meanings that are socially constructed and contextually situated, structuring both practices and representations of human groups (Geertz, 1973).

3.1.2 Emergence of Organizational Culture Studies

Interest in workplace cultural dimensions emerged throughout the 20th century as researchers explored symbolic and relational aspects of work environments (Pettigrew, 1979). Organizational culture became "one of the main themes in academic research and education, both in organization theory and management practice" (Alvesson, 2013, p. 15).

The 1980s marked organizational culture's consolidation as a central axis in management studies. This period witnessed unprecedented growth in cultural research, with publications increasing from 281 works between 1942-1986 to over 2,550 between 1980-1990 (Alvesson & Olofsson, 1992). This expansion reflected growing recognition of cultural factors as key determinants of business effectiveness (Beer, 1988).

Schein (1985) attributed this renewed interest to Japanese companies' success during this period. Their management model, emphasizing cultural cohesion, teamwork, and continuous improvement, contrasted sharply with declining North American competitive performance. Sackmann (1997) argued that organizational culture studies responded to the need for deeper understanding of organizational internal dynamics and their relationship with collective performance.

3.1.3 Theoretical Frameworks and Models

Schein's Three-Level Model. Schein's (1985) theoretical contribution constitutes one of the most systematic approaches to organizational culture analysis. His model articulates culture through three hierarchical levels: artifacts (observable behaviors and structures), espoused values (stated beliefs and ideologies), and basic underlying assumptions (unconscious beliefs and perceptions). Leadership plays a crucial role in cultural formation and transmission, with organizational change occurring when new values are introduced through managerial behavior (Schein, 1988, 2004). However, these values achieve deep integration only when organizational members perceive tangible benefits from their adoption.

Hofstede's Cultural Dimensions Theory. Hofstede's (1980) empirical research, based on longitudinal studies with IBM employees across 70+ countries, established a systematic framework for examining cultural differences. His dimensional approach—including power distance, individualism-collectivism, uncertainty avoidance, and masculinity-femininity—provided analytical tools for understanding how cultural variations influence leadership, motivation, decision-making, and communication patterns in organizational contexts (Hofstede, 2001).

Values-Based Approaches. Peters and Waterman (1982) emphasized the central role of values in defining organizational culture through their analysis of high-performing companies. Their research demonstrated that successful organizations share firmly internalized value systems among employees, suggesting that cultural "strength" resides in collective commitment intensity toward shared principles, facilitating internal coherence and strategic orientation.

Competing Values Framework. The Competing Values Framework (CVF), developed by Quinn and Cameron (1999), represents one of the most influential conceptual frameworks for organizational culture analysis and development. This model has been extensively utilized in organizational research, including studies examined in this systematic review (Naranjo-Valencia et al., 2010, 2012). The CVF categorizes organizational cultures into four typologies based on two conceptual axes: flexibility versus control (stability) and internal versus external orientation.

1. Clan Culture combines flexibility with internal orientation, emphasizing cohesion, commitment, participation, and community sense. These organizations develop collaborative, family-like environments where leadership is facilitative and decision-making values tradition, loyalty, and consensus (Cameron & Quinn, 2011).

2. Adhocracy Culture associates flexibility with external orientation, characterized by dynamism, innovation, and change readiness. Adhocratic organizations prioritize creativity, experimentation, and proactivity, typical of knowledge-intensive and technology sectors (Quinn & Rohrbaugh, 1983).

3. Market Culture combines stability with external orientation, prioritizing results, productivity, competitiveness, and goal achievement. Success measurement focuses on market share, profitability, and performance metrics within a transactional efficiency logic (Cameron & Quinn, 2011).

4. Hierarchical Culture represents stability with internal focus, valuing order, predictability, control, and formalization. These organizations adopt vertical structures and strict regulations, where leadership derives from formal authority and performance measurement emphasizes process adherence (Quinn & Rohrbaugh, 1983).

3.1.4 Theoretical Frameworks and Models

Since 2000, organizational culture research has continued evolving with increasing attention to how globalization, technology, and generational dynamics affect organizational practices and values. Foundational theorists like Schein have continued contributing to the field through updated editions of their seminal works. In the fourth edition of *Organizational Culture and Leadership* (2016), Schein incorporates new perspectives on how leadership influences and is influenced by culture, particularly in changing global environments, emphasizing the importance of adapting leadership and organizational cultural practices to contemporary challenges including diversity, technology, and innovation.

Scott (2014) expands organizational culture analysis through an institutional approach, situating culture within external institutional pressures. His three-pillar model—regulative, normative, and cultural-cognitive—enables understanding how social norms, formal structures, and shared cognitive frameworks configure organizational practices. This approach proves especially relevant for comprehending how culture is constructed both internally and through institutional environments.

Van den Berg and Wilderom (2004) propose conceptualizing organizational culture through shared

perceptions of work practices. This operational perspective structures culture around five dimensions: autonomy, external orientation, interdepartmental coordination, human resources orientation, and continuous improvement orientation. By linking this perspective with resource-based theory, these authors position culture as a strategic resource for competitive advantage.

Groysberg et al. (2018) analyzed cultures of over 230 companies alongside leadership styles and values of more than 1,300 executives across different sectors, regions, and organizational types. Their diagnostic utilized responses from approximately 25,000 employees through online surveys and executive interviews, aiming to understand how organizational cultures within different company types and sectors influence leadership styles and organizational performance outcomes.

Recent studies highlight how globalization, digitalization, and generational diversity have forced organizations to reconfigure their value systems and cultural structures (García & López, 2020; Martínez et al., 2021; Hernández, 2019). These works argue that organizational culture emerges as a dynamic process resulting from interaction between inherited organizational traditions and new competitive environment requirements.

Current research reaffirms the need to understand organizational culture as an adaptable system (Caligiuri, 2021; Fernandes et al., 2023; Riemer et al., 2025). These authors address dimensions such as cultural agility, sustainability, and organizational citizenship, positioning culture as a strategic element capable of facilitating innovation, resilience, and transformation in complex scenarios.

Contemporary organizational culture literature is characterized by increasing multidisciplinarity and incorporation of both qualitative and quantitative analytical approaches (Alvesson & Sveningsson, 2015; Gutterman, 2023). Recent research has adopted mixed methodologies to explore how cultural elements influence innovation, digital transformation, sustainability, and inclusion across different organizational types. This evolution reflects a shift from normative, universalist models toward more interpretative visions sensitive to diversity and constant transformation of organizational environments (Hatch, 2018; Hofstede et al., 2010).

3.2 Innovation Concept Evolution

3.2.1 Classical Foundations and Early Conceptualizations

From classical economics, Joseph Schumpeter is recognized as the author who established the theoretical foundations of innovation as an engine of economic development. In his seminal work *The Theory of Economic Development* (1934), he introduced the concept of "creative destruction," whereby new combinations of productive factors replace existing structures, generating economic growth. These combinations materialize in five forms: new products, new production methods, opening of new markets, new sources of raw materials, and new forms of organization. Schumpeter distinguished between radical innovations, which profoundly transform sectors, and incremental innovations, which represent gradual improvements, while emphasizing the entrepreneur's role as an agent of change (Schumpeter, 1934).

Following Schumpeter, various authors sought to categorize innovation. Knight (1967) proposed a typology encompassing products, processes, organizational structures, and people. Evan (1966) and Damanpour (1991, 1996) differentiated between technical and administrative innovations. Zaltman, Duncan, and Holbek (1973), and subsequently Kimberly and Evanisko (1981), introduced the sequential invention-innovation-diffusion model, which proved highly influential in organizational theory.

3.2.2 Organizational Innovation Theory Development

The linear approach was challenged by neo-Schumpeterian theory, particularly by Nelson and Winter (1982), who proposed an evolutionary vision based on organizational routines and continuous learning. Within this framework, innovation is not an isolated phase but a recurrent process of knowledge generation, adaptation, and diffusion.

Damanpour (1991) conceptualized innovation as "the adoption of an idea or behavior new to the organization that adopts it," emphasizing the relative character of novelty. Dewar and Dutton (1986) highlighted perception of novelty as an essential criterion. Amabile et al. (1996) contributed a creative dimension by considering innovation as "the successful implementation of creative ideas within an organization."

During the first decade of the 21st century, efforts were made to integrate previous approaches. Baregheh, Rowley, and Sambrook (2009) proposed a multidisciplinary definition describing innovation as a "multi-stage process whereby organizations transform ideas into new or improved products, services, or processes to compete and differentiate." Crossan and Apaydin (2010) defined innovation as "the production or adoption, assimilation, and exploitation of a value-added novelty in economic and social spheres," encompassing innovation as both process and outcome.

3.2.3 Standardization and Measurement Frameworks

A key reference point for innovation definition and measurement is the Oslo Manual, prepared by the OECD and Eurostat. Since its first edition in 1992, it has evolved through multiple editions to include service sectors and non-technological innovations. In

its fourth edition, it defines innovation as "a new or improved business product or process (or a combination of both) that differs significantly from previous ones and has been introduced to the market or put into use" (OECD/Eurostat, 2018, p.20). This definition emphasizes three key elements: relative novelty, effective implementation, and applicability to products and processes. The Manual distinguishes four types of innovation: product, process, organizational, and marketing; and introduces levels of novelty (company, market, world). Additionally, it differentiates between incremental and radical innovations. This framework has become a methodological reference for empirical research, international comparisons, and public policy design.

3.2.4 Contemporary Theoretical Developments

The open innovation theory, proposed by Chesbrough (2003), argues that companies no longer depend exclusively on internal knowledge generation but benefit from external flows that enrich their processes. This represents an evolution from closed innovation models traditionally addressed by organizations, relying mainly on their internal R&D laboratories protecting their developments.

The resource-based view (RBV) and dynamic capabilities have consolidated their theoretical relevance. Teece, Pisano, and Shuen (1997) argue that innovation requires capabilities to integrate, build, and reconfigure resources. Lawson and Samson (2001) consider innovation capability as the competence to transform knowledge into products or processes.

Organizational learning theory has linked innovation with knowledge creation, transfer, and application (Nonaka & Takeuchi, 1995; Argyris & Schön, 1978). Network approaches and social capital perspectives (Powell et al., 1996) highlight that position in inter-organizational networks facilitates access to novel resources. The distinction between exploratory and exploitative innovation (March, 1991) and the concept of organizational ambidexterity (Tushman & O'Reilly, 1996) have influenced understanding of organizations' capacity to balance exploration of new opportunities with efficient implementation.

3.2.5 Current Paradigms and Future Directions

In the last decade, innovation has been closely linked to contemporary challenges such as digitalization, sustainability, and organizational resilience. Nambisan et al. (2017) have redefined digital innovation, incorporating emerging technologies such as artificial intelligence, big data, and the Internet of Things. These technologies have transformed not only products but also processes, business models, and organizational structures, introducing new challenges in innovation management. Studies by Adams et al. (2016) and George et al. (2021) address sustainable innovation, highlighting the economic, social, and environmental impact of innovative practices. This perspective aligns with the Sustainable Development Goals (SDGs) and positions innovation as a tool to face global challenges such as climate change and inequality.

Recent literature has also recognized innovation's role in crisis contexts, demonstrating that innovative organizations show greater capacity for adaptation and recovery in volatile environments, reinforcing the need to promote innovative cultures and transformational leadership approaches.

3.3. Family Business

Family businesses represent the predominant form of business organization globally (Gersick et al., 1997; Miller & Le Breton-Miller, 2005). They constitute "the essence of the market economy" (Echezárraga & Martínez, 2010) and "the backbone of economic development" (Gómez-Betancourt, 2006). In the Spanish context, these organizations represent approximately 89% of the business fabric and generate more than 67% of private employment (Instituto de la Empresa Familiar, 2022).

Despite their quantitative relevance, family businesses have been subject to multiple definitions that have evolved as researchers have deepened their understanding of their unique characteristics (Chua et al., 1999; Astrachan et al., 2002). One of the most influential conceptualizations is that proposed by Chua, Chrisman, and Sharma (1999), who define the family business as "a business governed and/or managed with the intention that a dominant coalition controlled by members of the same family (or few families) shape and pursue the business vision in a potentially sustainable way across generations."

Given the challenge of establishing a universal delimitation, some authors have opted to evaluate family influence through scales such as the F-PEC index (Family Power, Experience, Culture) developed by Astrachan, Klein, and Smyrnios (2002). This index measures family involvement in three fundamental dimensions: (1) power (percentage of ownership and presence in governing bodies), (2) experience (generations involved), and (3) culture (family values and beliefs reflected in the company). Family businesses do not constitute a homogeneous group but present notable diversity in terms of size, ownership structure, degree of professionalization, and strategic orientation (Chua et al., 1999; Zellweger, 2017). This heterogeneity is particularly relevant for understanding their organizational behavior and, specifically, their relationship with innovation processes.

Nordqvist and Melin (2010) establish a differentiated typology based on temporal orientation and governance structures. They identify that family businesses with more formalized governance structures and long-term orientation tend to show greater

propensity toward innovative behaviors. Contrary to the common perception that these organizations are intrinsically conservative, studies such as that by EY & IE Center for Families in Business (2022) demonstrate that many family businesses lead in innovation within their respective sectors.

Zellweger et al. (2012) introduce the concept of "patient capital" as a distinctive characteristic of family businesses. They argue that this capacity to maintain long-term investments can constitute both a competitive advantage and a potential limitation, depending on how the specific organizational culture is configured.

Academic literature has identified five key dimensions that structure family business functioning: (1) family ownership and control, (2) generational succession, (3) family values and organizational identity, (4) specific organizational culture, and (5) family governance structures (Miller & Le Breton-Miller, 2005; Zellweger, 2017; Nason et al., 2018). These dimensions explain both the strengths and inherent challenges of family businesses, particularly their relationship with innovation processes.

During the 1980s and 1990s, family business studies focused primarily on issues such as succession planning and survival beyond the founder (Lansberg, 1988; Gersick et al., 1997). Subsequently, the field expanded toward broader areas such as professionalization, competitive strategy, and value creation. Research such as that of Sharma (2004) and Gómez-Mejía et al. (2007) introduced the Socioemotional Wealth theory, which highlights the importance of non-economic objectives in family decision-making.

3.4. Cultural Factors that Inhibit and Facilitate Innovation

Internal culture constitutes a key determinant for achieving sustainable competitive advantages, especially regarding product innovation (Naranjo-Valencia, Jiménez-Jiménez & Sanz-Valle, 2010). Innovation does not depend solely on technological or structural resources, but on shared values, organizational norms, and predominant attitudes that legitimize change and favor experimentation.

The culture-innovation relationship has been extensively documented in organizational theory. Schein (1985) demonstrated that cultural assumptions influence how organizations perceive and respond to change, directly affecting innovative capacity. Organizations with cultures valuing learning, experimentation, and calculated risk-taking demonstrate superior innovation performance compared to those emphasizing control and stability (Ahmed, 1998).

Damanpour's (1991) meta-analysis established organizational culture as a significant predictor of innovation adoption, with cultural characteristics accounting for substantial variance in innovation outcomes. His findings revealed that cultural factors often outweigh structural variables in determining innovation success.

The Competing Values Framework provides analytical structure for understanding cultural influences on innovation. Cameron and Quinn (1999) demonstrated that adhocracy cultures consistently outperform hierarchical cultures in innovation metrics, while clan cultures excel in incremental innovation and market cultures in commercialization.

Empirical Evidence. Martins and Terblanche (2003) identified cultural dimensions correlating with innovative behavior: individual freedom, open communication, and learning orientation. Their analysis revealed these factors create environments where innovative ideas emerge and develop effectively. Naranjo-Valencia et al. (2010, 2012) provided quantitative evidence of measurable relationships between cultural typologies and innovation performance. Organizations with strong adhocracy orientations achieved significantly higher rates of both incremental and radical innovation adoption compared to hierarchical counterparts.

Family Business Cultural Dynamics. Family businesses present distinctive cultural characteristics requiring specialized analysis. The intersection of family and business systems creates unique innovation patterns absent in non-family organizations (Zahra et al., 2004). Family values, succession planning, and socioemotional wealth considerations introduce cultural factors that can accelerate or constrain innovation processes. De Massis et al. (2015) found that family businesses with entrepreneurial traditions and values alignment between family and innovation objectives demonstrate superior innovation performance. Conversely, emphasis on tradition preservation and risk avoidance creates cultural barriers to innovation adoption (Gómez-Mejía et al., 2007).

4 Results

4.1 Systematic Review Overview

Publications Analyzed. This systematic literature review analyzed 47 academic publications spanning 1950-2024, with primary focus on contemporary research examining culture-innovation relationships in family business contexts. The corpus includes peer-reviewed journal articles, theoretical contributions, and empirical studies with solid methodological foundation.

Geographic Distribution. The analyzed publications demonstrate significant geographic concentration:

- North America: 28 studies (60%) - predominantly United States research
- Europe: 15 studies (32%) - Spain, Germany, Italy, United Kingdom contexts
- Asia-Pacific: 3 studies (6%) - Australia, China perspectives
- Latin America: 1 study (2%) - Brazilian family business analysis

This distribution reflects the predominance of Western research contexts previously identified as a limitation in family business innovation literature (Calabrò et al., 2019; De Massis et al., 2015).

4.2 Cultural Factors Identification

Based on systematic thematic analysis of 47 academic publications (1950-2024), cultural factors were identified and classified into facilitators and inhibitors of innovation through qualitative synthesis employing established theoretical frameworks as analytical foundation.

Cultural factors were extracted from empirical studies examining culture-innovation relationships and categorized using the Competing Values Framework (Cameron & Quinn, 1999) and innovation-culture taxonomies (Naranjo-Valencia et al., 2010, 2012) as analytical frameworks. Thematic categorization was based on reported innovation outcomes across the reviewed literature, with binary classification into facilitating and inhibiting factors.

4.2.1 Cultural Factors that Facilitate Innovation

The systematic analysis identified eleven facilitating factors demonstrating positive correlation with innovation outcomes across the reviewed studies: creativity, initiative and entrepreneurial spirit; freedom and autonomy; teamwork and collaboration; risk tolerance; resource sufficiency and access; participatory decision making; leadership support; long-term vision; continuous learning; flexibility and adaptability; and open communication and trust. Each factor is supported by specific empirical evidence and key references from the reviewed literature, as detailed in the following taxonomy (Table1).

4.2.2 Cultural Factors that Inhibit Innovation

Eight inhibiting factors demonstrating negative correlation with innovation performance were identified through the systematic review: rigid traditionalism and resistance to change; risk aversion; poor or closed communication; internal family conflicts; extreme financial conservatism; excessive focus on preserving socioemotional wealth; excessive decision centralization; and lack of professionalization. Each inhibiting factor is grounded in empirical findings from multiple studies within the analyzed corpus (Table 2).

Table 1. Cultural Factors that Facilitate Innovation in Family Businesses
Source: Own elaboration based on systematic literature review

Cultural Factor	Mechanism of Influence	Key References
Creativity, Initiative and Entrepreneurial Spirit	Fundamental predictor of innovative behavior	Naranjo-Valencia et al. (2012); Calabrò et al. (2019)
Freedom and Autonomy	Significant correlation with idea generation	De Massis et al. (2015); Nason et al. (2018)
Teamwork and Collaboration	Facilitation of collaborative innovation	Cruz & Nordqvist (2012); Miller & Le Breton-Miller (2005)
Risk Tolerance	Positive correlation with radical innovation	Calabrò et al. (2019); Zellweger (2017)
Resource Sufficiency and Access	Critical enabler of innovative projects	Duran et al. (2016); Block et al. (2013)
Participatory Decision Making	Increase in diversity of perspectives	Nordqvist & Melin (2010); Brush et al. (2019)
Leadership Support	Mediation in culture-innovation relationship	Miller & Le Breton-Miller (2005); Zahra et al. (2004)
Long-term Vision	Facilitation of patient capital and sustained investment	Zellweger (2017); Gómez-Mejía et al. (2007)
Continuous Learning	Impact on knowledge absorption capacity	Nason et al. (2018); Cruz & Nordqvist (2012)
Flexibility and Adaptability	Effective response to environmental changes	Calabrò et al. (2019); Nordqvist & Melin (2010)
Open Communication and Trust	Facilitation of idea flow and collaboration	De Massis et al. (2015); Vivanco & Brand (2019)

Table 2. Cultural Factors that Inhibit Innovation in Family Businesses
Source: Own elaboration based on systematic literature review

Cultural Factor	Mechanism of Influence	Key References
Rigid Traditionalism and Resistance to Change	Main barrier to disruptive innovation	Zahra et al. (2004); Gómez & Núñez-Cacho Utrilla (2012)
Risk Aversion	Limitation of opportunity exploration	Getz & Carlsen (2005); Morrison et al. (2008)
Poor or Closed Communication	Restriction of information flow	Vivanco & Brand (2019); Schulze et al. (2001)
Internal Family Conflicts	Distraction from innovation objectives	Calabrò et al. (2019); Chrisman et al. (2002)
Extreme Financial Conservatism	Limitation of R&D investment	Block et al. (2013); Duran et al. (2016)
Excessive Focus on Preserving Socioemotional Wealth	Prioritization of non-economic objectives over innovation	Gómez-Mejía et al. (2007); Berrone et al. (2012)
Excessive Decision Centralization	Limitation of innovation participation	Schulze et al. (2001); Nordqvist & Melin (2010)
Lack of Professionalization	Deficiencies in innovative process management	Nason et al. (2018); Calabrò et al. (2019)

4.3 Family Business Particularities

The cultural factors identified demonstrate differential manifestations within family business contexts compared to non-family organizations. Universal factors such as leadership support, risk tolerance, creativity and entrepreneurial spirit, and open communication operate through modified mechanisms due to family ownership and governance structures (Zahra et al., 2004; De Massis et al., 2015).

Family businesses exhibit specific cultural factors absent in non-family enterprises. Internal family conflicts emerge as a distinctive inhibiting factor affecting innovation processes (Calabrò et al., 2019; Chrisman et al., 2002). Excessive focus on preserving socioemotional wealth represents a family-specific cultural barrier that prioritizes non-economic objectives over innovation investment (Gómez-Mejía et al., 2007; Berrone et al., 2012). Rigid traditionalism and resistance to change constitute additional family-specific cultural inhibitors linked to preservation of family legacy (Zahra et al., 2004; Gómez & Núñez-Cacho Utrilla, 2012). Lack of professionalization emerges as a distinctive cultural deficit affecting innovation process management (Nason et al., 2018; Calabrò et al., 2019).

Conversely, long-term vision orientation represents a distinctive cultural facilitator enabling patient capital approaches and sustained innovation investment (Zellweger, 2017; Gómez-Mejía et al., 2007). Universal factors operate through family-mediated mechanisms where risk tolerance reflects family risk appetite, leadership support manifests through family governance dynamics, and communication patterns are influenced by family relationships (Miller & Le Breton-Miller, 2005; Vivanco & Brand, 2019).

5 Discussion

5.1 Theoretical Implications

This systematic review contributes to organizational culture and family business literature by providing comprehensive synthesis of culture-innovation relationships within family business contexts. The findings confirm that organizational culture constitutes a determining element in companies' innovative capacity, while revealing additional complexity introduced by family business characteristics.

The taxonomy developed demonstrates coexistence of universal organizational antecedents of innovation with family-specific cultural dimensions. This theoretical distinction enhances understanding of how generalizable organizational factors interact with contextual characteristics to generate distinctive innovation patterns in family businesses.

Evolution from General to Specialized Theory Evidence of conceptual evolution from general organizational culture theory toward family business specialization reflects maturation of the research field. This progression indicates recognition that family businesses require specialized theoretical frameworks rather than simple application of general organizational theories.

5.2 Practical Implications for Family Business Management

Cultural Factor Management. The identified taxonomy provides family businesses with specific cultural dimensions to assess and develop in their innovation enhancement efforts. Rather than treating culture as an abstract concept, organizations can focus on concrete factors such as leadership support, communication patterns, and risk tolerance development.

Family-Specific Consideration. Family businesses must address unique cultural challenges including socioemotional wealth preservation effects, succession planning impacts, and family governance influences on innovation processes. Recognition of these family-specific factors enables development of appropriate management strategies.
Professionalization and Innovation. The consistent identification of professionalization deficits as an innovation inhibitor suggests that family businesses should prioritize management professionalization as a cultural development strategy for innovation enhancement.

5.3 Methodological Contributions

Systematic Synthesis Approach This study demonstrates the value of systematic literature review methodology for understanding complex interdisciplinary relationships. The comprehensive temporal scope (1950-2024) enables tracing conceptual evolution while the systematic approach ensures transparent and replicable analysis.
Taxonomic Development The binary classification of cultural factors into facilitators and inhibitors provides clear analytical structure while maintaining theoretical grounding in established frameworks. This approach enables practical application while preserving academic rigor.

5.4 Limitations and Future Research Directions

Geographic and Cultural Limitations. The predominance of Western research contexts limits generalizability to emerging economies and non-Western cultural contexts. Future research should address this geographic bias through comparative studies across diverse cultural environments.
Methodological Considerations. The single-researcher approach to literature selection and analysis may introduce bias. Future systematic reviews could benefit from multi-researcher teams and inter-coder reliability assessment.
Empirical Validation Needs. While this review identifies cultural factors based on existing literature, empirical validation through primary research remains necessary. Longitudinal studies examining cultural change and innovation outcomes would provide valuable insights into causal mechanisms.
Measurement and Assessment. Development of validated instruments for measuring family business cultural factors represents an important future research direction. Such instruments would enable empirical testing of the relationships identified in this review.

6 Conclusions

This systematic review identified cultural factors that facilitate or inhibit innovation in family businesses through analysis of 47 academic publications spanning 1950-2024. The findings provide theoretical and practical contributions to family business and innovation management literature.

6.1 Main contributions

Comprehensive taxonomy of cultural factors affecting innovation. The study developed a systematic classification distinguishing facilitating factors from inhibiting factors
Evidence of cultural complexity in family business innovation. The analysis demonstrates that family businesses exhibit both universal organizational factors and family-specific cultural dimensions that operate through distinctive mechanisms absent in non-family enterprises.
Conceptual evolution documentation from general organizational culture theory to family business specialization. The review traces theoretical development from foundational culture concepts (1950s) through organizational culture theory emergence (1980s) to contemporary family business innovation research (2000s-present).

6.2 Practical implications

Family businesses seeking innovation enhancement must develop systematic approaches to cultural factor management. Critical areas include leadership support mechanisms, open communication systems, and organizational learning processes. Management professionalization emerges as essential for innovation success, enabling integration of external expertise while maintaining family strategic involvement.

Organizations should prioritize cultural assessment focusing on identified facilitating and inhibiting factors, with particular attention to family-specific dynamics including socioemotional wealth considerations and succession planning effects on innovation

6.3 Future research lines

- Empirical validation of identified cultural factors through quantitative studies measuring relationships between specific cultural dimensions and innovation outcomes in family business samples.
- Longitudinal research examining cultural factor evolution and innovation performance relationships

over time, particularly during generational transitions and organizational development phases.
- Cross-cultural comparative analysis addressing geographic limitations through studies across diverse national and cultural contexts.
- Measurement instrument development for family business cultural assessment, enabling practical application of identified factor taxonomy.
- Causal mechanism analysis investigating specific processes through which cultural factors influence innovation outcomes, moving beyond correlational relationships to explanatory understanding.

Acknowledgments

The author would like to express sincere gratitude to Dr. Marta Mas, Dean of UIC Barcelona, and Dr. Xavier Ferrás, Professor at Universitat Ramon Llull (ESADE Business School), thesis directors, for their invaluable guidance, constructive feedback, and continuous support throughout this research. Special thanks are extended to the Doctoral Program in Economics and Law at UIC Barcelona for providing the academic environment and resources necessary for conducting this systematic literature review. The author also acknowledges the valuable feedback received from anonymous reviewers, which significantly improved the quality and rigor of this

References

Adams, R., Jeanrenaud, S., Bessant, J., Denyer, D., & Overy, P. (2016). Sustainability-oriented innovation: A systematic review. *International Journal of Management Reviews*, 18(2), 180-205.

Ahmed, P. K. (1998). Culture and climate for innovation. *European Journal of Innovation Management*, 1(1), 30-43.

Alvesson, M. (2013). *Understanding organizational culture*. Sage Publications.

Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron, M. (1996). Assessing the work environment for creativity. *Academy of Management Journal*, 39(5), 1154-1184.

Arad, S., Hanson, M. A., & Schneider, R. J. (1997). A framework for the study of relationships between organizational characteristics and organizational innovation. *Journal of Creative Behavior*, 31(1), 42-58.

Argyris, C., & Schön, D. A. (1978). Organizational learning: A theory of action perspective. Addison-Wesley.

Astrachan, J. H., Klein, S. B., & Smyrnios, K. X. (2002). The F-PEC scale of family influence: A proposal for solving the family business definition problem. Family Business Review, 15(1), 45-58.

Baregheh, A., Rowley, J., & Sambrook, S. (2009). Towards a multidisciplinary definition of innovation. Management Decision, 47(8), 1323-1339.

Beer, M. (1988). Toward a redefinition of OD: A critique of research focus and method. Academy of Management OD Newsletter, Winter, 4-6.

Berrone, P., Cruz, C., & Gomez-Mejia, L. R. (2012). Socioemotional wealth in family firms: Theoretical dimensions, assessment approaches, and agenda for future research. *Family Business Review*, 25(3), 258-279.

Block, J., Miller, D., Jaskiewicz, P., & Spiegel, F. (2013). Economic crisis and the resilience of family firms: The role of diversification. *Journal of Business Venturing*, 28(6), 773-786.

Brush, C. G., Edelman, L. F., Manolova, T. S., & Welter, F. (2019). A gendered look at entrepreneurship ecosystems. *Small Business Economics*, 53(2), 393-408.

Calabrò, A., Vecchiarini, M., Gast, J., Campopiano, G., De Massis, A., & Kraus, S. (2019). Innovation in family firms: A systematic literature review and guidance for future research. *International Journal of Management Reviews*, 21(3), 317-355.

Cameron, K. S., & Quinn, R. E. (1999). *Diagnosing and changing organizational culture: Based on the competing values framework*. Addison-Wesley.

Chesbrough, H. W. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business School Press.

Chrisman, J. J., Chua, J. H., & Steier, L. P. (2002). The influence of national culture and family involvement on entrepreneurial perceptions and performance at the state level. *Entrepreneurship Theory and Practice*, 26(4), 113-130.

Codina, L. (2018). Revisiones sistemáticas y metaanálisis: bases conceptuales e interpretación. *Revista Española de Cardiología*, 71(10), 790-797.

Cruz, C., & Nordqvist, M. (2012). Entrepreneurial orientation in family firms: A generational perspective. *Small Business Economics*, 38(1), 33-49.

Damanpour, F. (1991). Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of Management Journal*, 34(3), 555-590.

Damanpour, F. (1996). Organizational complexity and innovation: Developing and testing multiple contingency models. Management Science, 42(5), 693-716.

De Massis, A., Di Minin, A., & Frattini, F. (2015). Family-driven innovation: Resolving the paradox in family firms. *California Management Review*, 58(1), 5-19.

Duran, P., Kammerlander, N., Van Essen, M., & Zellweger, T. (2016). Doing more with less: Innovation input and output in family firms. *Academy of Management Journal*, 59(4), 1224-1264.

EY & IE Center for Families in Business. (2022). *Global Family Business Survey 2021: The next generation of family business – Innovation and technology*. Ernst & Young Global Limited.

Instituto de la Empresa Familiar. (2022). *Factores de competitividad y análisis financiero de la empresa familiar española.* Madrid: Instituto de la Empresa Familiar.

García, L., & López, S. (2020). Transformación digital y cultura organizacional: nuevos paradigmas. *Revista de Innovación Empresarial*, 12(4), 45-62.

George, G., Howard-Grenville, J., Joshi, A., & Tihanyi, L. (2021). Understanding and tackling societal grand challenges through management research. *Academy of Management Journal*, 59(6), 1880-1895.

Geertz, C. (1973). The interpretation of cultures: Selected essays. Basic Books.

Getz, D., & Carlsen, J. (2005). Family business research: The pace of change and time to consolidate. *Family Business Review*, 18(2), 101-111.

Gómez, D., & Núñez-Cacho Utrilla, P. (2012). Inmovilismo estratégico en la empresa familiar. *Harvard Deusto Business Review*, 207, 56-63.

Gómez-Mejía, L. R., Haynes, K. T., Núñez-Nickel, M., Jacobson, K. J., & Moyano-Fuentes, J. (2007). Socioemotional wealth and business risks in family-controlled firms. *Administrative Science Quarterly*, 52(1), 106-137.

Groysberg, B., Lee, J., Price, J., & Cheng, J. (2018). The leader's guide to corporate culture. *Harvard Business Review*, 96(1), 44-52.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Sage Publications.

Hofstede, G. (2001). Culture's consequences: Comparing values, behaviors, institutions and organizations across nations (2nd ed.). Sage Publications.

Kroeber, A. L., & Kluckhohn, C. (1952). *Culture: A critical review of concepts and definitions*. Harvard University Press.

March, J. G. (1991). Exploration and exploitation in organizational learning. Organization Science, 2(1), 71-87.

Martins, E. C., & Terblanche, F. (2003). Building organisational culture that stimulates creativity and innovation. *European Journal of Innovation Management*, 6(1), 64-74.

Miller, D., & Le Breton-Miller, I. (2005). *Managing for the long run: Lessons in competitive advantage from great family businesses*. Harvard Business Review Press.

Naranjo-Valencia, J. C. (2011). *Organizational culture and innovation performance: A study in Spanish firms* [Tesis doctoral]. Universidad de Valencia.

Naranjo-Valencia, J. C., Jiménez-Jiménez, D., & Sanz-Valle, R. (2010). Innovation or imitation? The role of organizational culture. *Management Decision*, 49(1), 55-72.

Nason, R. S., Mazzelli, A., & Carney, M. (2018). The ties that bind: Examining the relationship between family involvement and firm innovation. *Family Business Review*, 31(3), 385-410.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change.* Harvard University Press.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation.* Oxford University Press.

Nordqvist, M., & Melin, L. (2010). Entrepreneurial families and family firms. *Entrepreneurship & Regional Development*, 22(3-4), 211-239.

OECD/Eurostat. (2018). Oslo Manual 2018: Guidelines for collecting, reporting and using data on innovation (4th ed.). OECD Publishing.

Peters, T. J., & Waterman, R. H. (1982). *In search of excellence: Lessons from America's best-run companies*. Harper & Row.

Schein, E. H. (1985). *Organizational culture and leadership*. Jossey-Bass.

Schein, E. H. (2004). *Organizational culture and leadership* (3rd ed.). Jossey-Bass.

Schumpeter, J. A. (1934). *The theory of economic development*. Harvard University Press.

Schulze, W. S., Lubatkin, M. H., Dino, R. N., & Buchholtz, A. K. (2001). Agency relationships in family firms: Theory and evidence. *Organization Science*, 12(2), 99-116.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. Strategic Management Journal, 18(7), 509-533.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. British Journal of Management, 14(3), 207-222.

Vivanco, S., & Brand, M. (2019). Family governance and innovation in family businesses. *Journal of Family Business Management*, 9(2), 148-165.

Zahra, S. A., Hayton, J. C., & Salvato, C. (2004). Entrepreneurship in family vs. non-family firms. *Entrepreneurship Theory and Practice*, 28(4), 363-381.

Zellweger, T. M. (2017). *Managing the family business: Theory and practice*. Edward Elgar Publishing.

Digital Transformation of HR Practices: A Systematic Review and Bibliometric Analysis of e-HRM Literature

Digitalna transformacija prakse LJP-a: Sustavni pregled i bibliometrijska analiza literature o e-ULJP-u

Luka Juričinec, Marina Klačmer Čalopa
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
ljuricine24@student.foi.hr, marina.klacmer@foi.hr

Abstract. *The digital transformation of Human Resource Management (HRM) has accelerated the adoption of electronic Human Resource Management (e-HRM), integrating advanced technologies such as artificial intelligence (AI) into HR components. This paper presents a systematic review and bibliometric analysis of the literature on e-HRM from 2015 to 2025. Special attention is paid to the role of artificial intelligence in shaping modern HR practices. In addition to quantitative bibliometric insights, a qualitative content analysis of eight of the most relevant papers dealing with artificial intelligence in e-HRM was conducted to explore emerging topics, challenges, and research gaps. The study contributes to the academic discourse by mapping the intellectual structure of the e-HRM field and suggesting future research directions in the era of intelligent HR systems.*

Keywords. electronic human resource management, e-HRM, digital transformation, artificial intelligence, systematic literature review, bibliometric analysis

Sažetak. *Digitalna transformacija upravljanja ljudskim potencijalima (ULJP) ubrzala je usvajanje elektroničkih sustava upravljanja ljudskim potencijalima (e-ULJP), integrirajući napredne tehnologije poput umjetne inteligencije (UI) u komponente LJP-a. Ovaj rad predstavlja sustavni pregled i bibliometrijsku analizu literature o e-ULJP-u od 2015. do 2025. godine. Posebna pozornost se posvećuje ulozi umjetne inteligencije u oblikovanju modernih praksi LJP-a. Uz kvantitativne bibliometrijske uvide, provedena je kvalitativna analiza sadržaja osam najrelevantnijih radova koji se bave umjetnom inteligencijom u e-ULJP-u kako bi se istražile nove teme, izazovi i istraživački nedostaci. Studija doprinosi akademskom diskursu mapiranjem intelektualne strukture područja e-ULJP-a i predlaganjem budućih istraživačkih smjerova u eri inteligentnih sustava LJP-a.*

Ključne riječi. elektroničko upravljanje ljudskim potencijalima, e-ULJP, digitalna transformacija, umjetna inteligencija, sustavni pregled literature, bibliometrijska analiza

1 Introduction

The digital transformation of HRM marks one of the most significant changes in organizational practices in recent decades. Faced with rapid change, organizations across sectors are recognizing the need to integrate digital technologies into HRM, which has become crucial for achieving efficiency and agility. This change includes the application of cloud computing, data analytics, artificial intelligence, and various digital tools that are reshaping the way HR services are delivered and decisions about employees are made.

1 Uvod

Digitalna transformacija LJP-a označava jednu od najvažnijih promjena u organizacijskim praksama u proteklim desetljećima. Suočene s brzim promjenama, organizacije u različitim sektorima prepoznaju potrebu za integracijom digitalnih tehnologija u LJP-u, što je postalo presudno za postizanje efikasnosti i agilnosti. Ova promjena uključuje primjenu računarstva u oblaku, analize podataka, umjetne inteligencije i raznih digitalnih alata koji preoblikuju način na koji se pružaju usluge LJP-a i donose odluke o zaposlenicima.

One of the main elements of this transformation is electronic Human Resources Management (e-HRM), which involves the use of various technologies to support and implement HR functions (Bondarouk & Brewster, 2016; Strohmeier, 2020). From an initial focus on the digitalization of administrative processes, e-HRM has evolved into a strategic system that impacts workforce planning, talent acquisition, performance management, learning and development, and employee engagement (Marler & Parry, 2015). Recent advances, especially in the application of artificial intelligence within HRM, have expanded the reach of e-HRM by enabling predictive analytics, algorithm-based decision-making, and automation of routine tasks (Nyathi & Kekwaletswe, 2023). The use of generative AI, such as ChatGPT, enables new research within HRM. This research encompasses the automation of communication, ethical implications, potential changes in work roles, and the need for transparency and human oversight (Budhwar et al., 2023). Empirical research has shown that the use of artificial intelligence in e-HRM improves efficiency, helps decision-making and reduces costs, but there are significant challenges such as lack of skills, fear of privacy and resistance to change (Mwita & Kitole, 2025).

Considering the increasing challenges and strategic importance of e-HRM, there is a growing need for a systematic investigation of previous scientific research in this area. Bibliometric analysis offers a useful method for insight into the structure, dynamics and development of scientific literature. It enables the identification of authors, topics, institutions and publishing trends, as well as finding gaps in knowledge and potential new research areas.

The aim of this paper is to create a comprehensive overview of e-HRM research in the past ten years (2015–2025) through a combination of systematic review, bibliometric mapping and qualitative content analysis. The research particularly emphasizes the importance of artificial intelligence within the literature on e-HRM, with the aim of investigating the impact of artificial intelligence on the transformation of the HR function.

The research is based on the following research questions:

1. How has the scope and focus of e-HRM research evolved over the last decade?
2. Who are the authors, topics, and institutions that have shaped this field?
3. What is the role of artificial intelligence in the context of e-HRM?

Jedan od glavnih elemenata ove transformacije jest elektroničko upravljanje ljudskim potencijalima (e-ULJP), koje podrazumijeva korištenje različitih tehnologija za podršku i implementaciju funkcija LJP-a (Bondarouk & Brewster, 2016; Strohmeier, 2020). Od početnog fokusiranja na digitalizaciju administrativnih procesa, e-ULJP je evoluirao u strateški sustav koji utječe na planiranje radne snage, privlačenje talenata, upravljanje radnim učinkom, obrazovanje i razvoj, kao i angažman zaposlenika (Marler & Parry, 2015). Nedavni napredak, posebno u primjeni umjetne inteligencije unutar ULJP-a, proširio je domet e-ULJP-a omogućujući prediktivnu analizu, donošenje odluka na osnovi algoritama i automatizaciju rutinskih zadataka (Nyathi & Kekwaletswe, 2023). Upotreba generativne umjetne inteligencije, kao što je ChatGPT, omogućuje nova istraživanja unutar ULJP-a. Ova istraživanja obuhvaćaju automatizaciju komunikacije, etičke posljedice, potencijalne promjene u radnim ulogama te potrebe za transparentnošću i ljudskim nadzorom (Budhwar et al., 2023). Empirijska istraživanja su pokazala da korištenje umjetne inteligencije u e-ULJP-u poboljšava efikasnost, pomaže u donošenju odluka i smanjuje troškove, ali postoje značajni izazovi poput nedostatka vještina, straha za privatnost i otpornosti ka promjenama (Mwita & Kitole, 2025).

S obzirom na sve veće izazove i stratešku važnost e-ULJP-a, postoji rastuća potreba za sustavnim istraživanjem dosadašnjih znanstvenih istraživanja u ovom području. Bibliometrijska analiza nudi korisnu metodu za uvid u strukturu, dinamiku i razvoj znanstvene literature. Ona omogućava identifikaciju autora, tema, institucija i trendova u objavljivanju, kao i pronalaženje praznina u znanju i potencijalnih novih istraživačkih područja.

Cilj ovog rada je izraditi sveobuhvatan prikaz istraživanja e-ULJP-a u proteklih deset godina (2015.–2025.) kroz kombinaciju sustavnog pregleda, bibliometrijskog mapiranja i kvalitativne analize sadržaja. Istraživanje posebno naglašava važnost umjetne inteligencije unutar literature o e-ULJP-u, s ciljem istraživanja utjecaja umjetne inteligencije na transformaciju funkcije LJP-a.

Istraživanje se temelji na sljedećim istraživačkim pitanjima:

1. Kako je obujam i fokus istraživanja e-ULJP-a evoluirao u posljednjem desetljeću?
2. Tko su autori, teme i institucije koje su oblikovale ovo područje?
3. Kakva je uloga umjetne inteligencije u kontekstu e-ULJP-a?

2 Theoretical framework and conceptual definition

Electronic Human Resources Management, known as e-HRM, encompasses a wide range of activities, including planning, application, and use of information technology. Its goal is to network and enable cooperation among actors in the joint implementation of activities related to Human Resources Management (Strohmeier, 2007).

Digital Human Resource Management enables individualized employee experiences, real-time workforce analytics, and agile HR operations. In this way, digital HRM can be considered a subcategory of e-HRM that focuses more on innovation and transformation (Bondarouk & Brewster, 2016).

The digital transformation of Human Resources Management brings significant opportunities, but also serious challenges. In terms of capabilities, digital HR tools can ease administrative tasks, improve the employee experience, offer useful workforce insights, and aid in data-driven decision-making. Organizations have the ability to use artificial intelligence to predict labor trends, adjust development programs, and improve the quality and speed of talent acquisition (Strohmeier, 2020).

On the other hand, digital transformation also brings with it certain challenges, including technical integration, acceptance of change, cybersecurity risks, as well as ethical aspects related to data privacy and algorithmic bias. Digital HRM also requires new competencies from HR professionals, who must be skilled in data analysis, system implementation, and digital strategy (Marler & Fisher 2013).

Academic interest in e-HRM began in the early 2000s, when it primarily focused on the adoption and technical use of information systems within HR functions. As time went on, researchers expanded their attention to the impact of e-HRM on organizational performance, employee experience, and strategic decision-making (Ruël & Bondarouk & Looise 2004).

Several bibliometric studies have explored the broader picture of HRM, but few have specifically addressed e-HRM or placed a strong emphasis on artificial intelligence. Previous systematic reviews have demonstrated the fragmentation of the field and the need for integrated frameworks that connect technology, human behavior, and organizational outcomes.

2 Teorijski okvir i pojmovno određenje

Elektroničko upravljanje ljudskim potencijalima, poznato kao e-ULJP, obuhvaća široki spektar aktivnosti, a uključuje planiranje, primjenu i korištenje informacijske tehnologije. Cilj mu je umrežiti i omogućiti suradnju među akterima u zajedničkom provođenju aktivnosti vezanih uz upravljanje ljudskim potencijalima (Strohmeier, 2007).

Digitalno upravljanje ljudskim potencijalima omogućuje individualizirana iskustva zaposlenika, analitiku radne snage u stvarnom vremenu te agilnost operacija LJP-a. Na taj način, digitalno ULJP se može smatrati podkategorijom e-ULJP-a koja se više fokusira na inovacije i transformaciju (Bondarouk & Brewster, 2016).

Digitalna transformacija upravljanja ljudskim potencijalima donosi značajne mogućnosti, ali i ozbiljne izazove. U vezi s mogućnostima, alati za digitalno upravljanje ljudskim potencijalima mogu olakšati administrativne zadatke, unaprijediti iskustvo zaposlenika, ponuditi korisne uvide o radnoj snazi i pomoći kod donošenja odluka koje se oslanjaju na podatke. Organizacije imaju mogućnost korištenja umjetne inteligencije za predviđanje radnih trendova, prilagodbu programa za razvoj i poboljšanje kvalitete i brzine pronalaženja talenata (Strohmeier, 2020).

S druge strane, digitalna transformacija nosi sa sobom i određene izazove, uključujući tehničku integraciju, prihvaćanje promjena, rizike vezane uz kibernetičku sigurnost, kao i etičke aspekte vezane uz zaštitu privatnosti podatka i pristranost u algoritmima. Također, digitalno upravljanje ljudskim potencijalima traži nove kompetencije od stručnjaka za LJP, koji moraju biti vješti u analizi podataka, implementaciji sustava i digitalnoj strategiji (Marler & Fisher 2013).

Akademski interes za e-ULJP započeo je na početku 2000-ih, kada se prvenstveno fokusirao na usvajanje i tehničku upotrebu informacijskih sustava unutar funkcija LJP-a. Kako je vrijeme prolazilo, istraživači su proširili svoju pažnju na utjecaj e-ULJP-a na učinkovitost organizacije, doživljaj zaposlenika i strateško donošenje odluka (Ruël & Bondarouk & Looise 2004).

Nekoliko bibliometrijskih istraživanja istražilo je širu sliku ULJP-a, no malo njih se konkretno bavilo e-ULJP-om ili je stavilo snažan naglasak na umjetnu inteligenciju. Prethodni sustavni pregledi pokazali su raštrkanost ovog područja i potrebu za integriranim okvirima koji povezuju tehnologiju, ljudsko ponašanje i organizacijske rezultate.

This theoretical framework sets the stage for future analysis that explores how these concepts and themes have been reflected and developed in the academic literature over the past decade.

3 Methodology

In order to achieve comprehensive coverage in the research, two recognized databases were used: Web of Science (WoS) and Scopus. Their selection is based on defined criteria for the inclusion of publications, which ensures a certain level of scientific relevance and reliability of sources. WoS and Scopus include journals with international peer review and clearly defined impact factors, which enables the analysis of not only the quantity but also the quality of published works. In addition, although partially overlapping, these bases also have complementary indexes, which further increases the breadth of coverage. These databases provide high-quality bibliographic information as well as advanced filtering tools needed for bibliometric and content analyses. The inclusion of these sources provides broad access to relevant peer-reviewed journal articles covering all disciplines related to human resource management and information systems.

The search was limited in time to the period from 2015 to 2025 in order to cover the latest trends in the field of e-HRM. Only journal articles were included in the analysis. An extensive list of key terms has been created to ensure broad coverage of the relevant literature. The following terms were used during the search: e-HRM, e-HR, Electronic HRM, Electronic HR, Electronic Human Resource, Digital HRM, Digital HR, Digital Human Resource, HRIS, Human Resource Information System, HRIT, Human Resource Information Technology; with the integration of advanced technologies such as artificial intelligence. The search was conducted separately in both databases, and the results were downloaded in .xls and .csv format. Manual and automatic removal of duplicates based on identical titles and authors was performed in Microsoft Excel.

In order to additionally ensure the objectivity and scientific foundation of the selection of papers for detailed analysis, additional selection criteria were defined. Each paper was scored according to the number of citations (citations), which attempted to quantify the scientific relevance and impact of the work within the research community. In parallel, the content of the titles was also analyzed – priority was given to papers that explicitly contained key terms relevant to the research area in the title, such as "e-HRM", "Digital HR", "AI in HRM", "HRIS" and related terms.

Based on the combination of these criteria – citations and semantic relevance of the title – a

Ovaj teorijski okvir postavlja osnovu za buduću analizu koja istražuje kako se ti koncepti i teme reflektiraju i razvijaju u akademskoj literaturi tokom proteklog desetljeća.

3 Metodologija

Da bi se postigla sveobuhvatna pokrivenost u istraživanju, korištene su dvije priznate baze podataka: Web of Science (WoS) i Scopus. Njihov odabir temelji se na definiranim kriterijima za uključivanje publikacija, što osigurava određenu razinu znanstvene relevantnosti i pouzdanosti izvora. WoS i Scopus uključuju časopise s međunarodnom recenzijom i jasno definiranim faktorima utjecaja, čime se omogućava analiza ne samo kvantitete nego i kvalitete objavljenih radova. Osim toga, iako se djelomično preklapaju, ove baze imaju i komplementarne indekse, što dodatno povećava širinu pokrivenosti. Ove baze podataka pružaju visoko kvalitetne bibliografske informacije, kao i napredne alate za filtraciju potrebne za bibliometrijske i analize sadržaja. Uključivanje ovih izvora osigurava širok pristup relevantnim recenziranim člancima iz časopisa koji pokrivaju sve discipline povezane s upravljanjem ljudskim potencijalima i informacijskim sustavima.

Pretraživanje je bilo vremenski ograničeno na razdoblje od 2015. do 2025. godine kako bi se obuhvatili najnoviji trendovi u području e-ULJP-a. U analizu su uključeni samo članci iz časopisa. Napravljena je opsežna lista ključnih izraza kako bi se osigurala široka pokrivenost važne literature. Tokom pretraživanja korišteni su sljedeći pojmovi: e-HRM, e-HR, Electronic HRM, Electronic HR, Electronic Human Resource, Digital HRM, Digital HR, Digital Human Resource, HRIS, Human Resource Information System, HRIT, Human Resource Information Technology; uz integraciju naprednih tehnologija kao što je umjetna inteligencija. Pretraga je provedena odvojeno u obje baze, a rezultati su preuzeti u .xls i .csv formatu. U Microsoft Excelu provedeno je ručno i automatsko uklanjanje duplikata na temelju identičnih naslova i autora.

Kako bi se dodatno osigurala objektivnost i znanstvena utemeljenost odabira radova za detaljnu analizu, definirani su dodatni kriteriji selekcije. Svaki rad je bodovan prema broju citata (citiranosti), čime se nastojala kvantificirati znanstvena relevantnost i utjecaj rada unutar istraživačke zajednice. Paralelno s tim, analiziran je i sadržaj naslova – prioritet su dobivali radovi koji u naslovu eksplicitno sadrže ključne izraze relevantne za područje istraživanja, poput "e-HRM", "Digital HR", "AI in HRM", "HRIS" i srodnih termina.

Na temelju kombinacije ovih kriterija – citiranosti i semantičke relevantnosti naslova –

narrower group of papers was formed. The final selection included the 8 most relevant scientific papers, which were taken into a detailed qualitative analysis. The aim of this phase was to understand how individual authors conceptualize and operationalize e-HRM and its connection with digital technologies, especially artificial intelligence. This approach ensured methodological transparency and objectivity in the selection of literature that forms the basis for further research conclusions.

4 Results of bibliometric analysis

In the initial phase of the search, 56 works were identified in the Scopus database and 44 works in the WoS database, which made a total of 100 works. After exporting data to Excel and implementing manual and automatic removal of duplicates, the number of papers was reduced to 74 unique scientific articles. The obtained set of 74 papers forms the basis for further bibliometric analysis. After the implementation of the bibliometric analysis, a qualitative selection of papers was additionally carried out for a deeper analysis. For this purpose, additional criteria were applied: the number of citations of an individual paper and the presence of key terms in the title of the paper, which enabled the identification of the most relevant publications. On the basis of these criteria, 8 works were singled out, on which a detailed content analysis was carried out. This ensures a balance between the breadth of coverage and in-depth processing of the relevant scientific literature. Other papers were used to identify thematic areas, key authors, collaborative networks and time trends.

The largest number of publications was recorded in 2024, indicating an increased interest in the topic of e-HRM and its connection with the application of artificial intelligence in modern HR practices. The growth in scientific production began in 2018, in a period when the importance of digital transformation is increasingly emphasized globally, especially in the context of adapting businesses to new technologies, as shown in the figure.

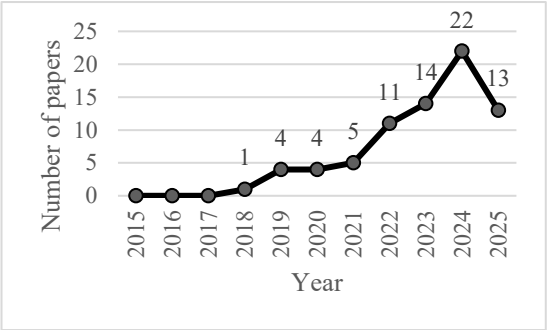


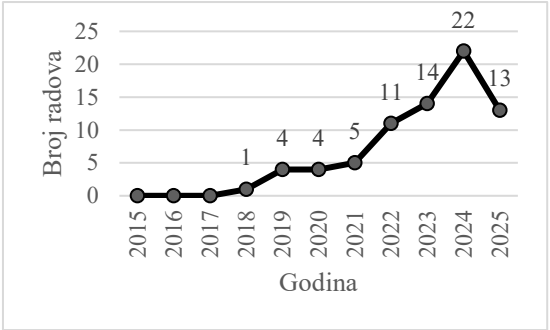
Figure 1. Overview of the number of published papers by year in the period from 2015 to 2025

formirana je uža skupina radova. Konačni odabir uključivao je 8 najrelevantnijih znanstvenih radova, koji su uzeti u detaljnu kvalitativnu analizu. Cilj ove faze bio je razumjeti kako pojedini autori konceptualiziraju i operacionaliziraju e-ULJP i njegovu povezanost s digitalnim tehnologijama, osobito umjetnom inteligencijom. Ovim pristupom osigurana je metodološka transparentnost i objektivnost u izboru literature koja čini temelj daljnjih istraživačkih zaključaka.

4 Rezultati bibliometrijske analize

U početnoj fazi pretrage identificirano je 56 radova u bazi Scopus i 44 rada u bazi WoS, što je ukupno činilo 100 radova. Nakon izvoza podataka u Excel i provedbe ručnog te automatskog uklanjanja duplikata, broj radova smanjen je na 74 jedinstvena znanstvena članka. Dobiveni skup od 74 rada čini osnova je za daljnju bibliometrijsku analizu. Nakon provedbe bibliometrijske analize, dodatno je provedena kvalitativna selekcija radova za dublju analizu. U tu svrhu primijenjeni su dodatni kriteriji: broj citata pojedinog rada te prisutnost ključnih pojmova u naslovu rada, što je omogućilo identifikaciju najrelevantnijih publikacija. Na temelju tih kriterija izdvojeno je 8 radova na kojima je provedena detaljna sadržajna analiza. Time je osigurana ravnoteža između širine obuhvata i dubinske obrade relevantne znanstvene literature. Ostali radovi korišteni su za identifikaciju tematskih područja, ključnih autora, kolaborativnih mreža i vremenskih trendova.

Najveći broj publikacija zabilježen je 2024. godine, što ukazuje na pojačan interes za temu e-ULJP-a i njezinu vezu s primjenom umjetne inteligencije u suvremenim praksama LJP-a. Rast u znanstvenoj produkciji započinje od 2018. godine, u razdoblju kada se globalno sve više ističe važnost digitalne transformacije, posebno u kontekstu prilagodbe poslovanja novim tehnologijama, što je prikazano slikom.



Slika 1. Prikaz broja objavljenih radova po godinama u razdoblju od 2015. do 2025. godine

The increase in announcements during the COVID-19 pandemic can be linked to the need for quick adaptations and the implementation of digital solutions in human resources management, especially in the context of remote work, digital communication, and automation of administrative HR processes.

The above confirms that e-HRM has become an important and rapidly developing research area. The growth trend has been observed since 2018, when discussions on the digital transformation of the HR function intensified, especially through the implementation of artificial intelligence, predictive analytics, chatbots, performance monitoring systems and digital talent management platforms.

Among the most important topics that stand out in recent literature are digitization of the recruitment and selection process, personalization of employee development through analytics, digital evaluation of work performance, and ethical challenges and employee data privacy in digital HR systems. In this way, the literature not only reflects technological changes, but also provides insight into the transformation of HRM from an administrative function to a strategic role, with increasing support for digital tools and data-driven HR management.

5 Overview of the most important works according to the chosen methodology

The digital transformation of HR practice in the literature is most often thematized through several key areas: (1) recruitment, selection and process automation, (2) learning, development and user experience, (3) analytics and decision-making, and (4) strategic role and digital transformation of HR.

Votto et al. (2021) prepared the paper "Artificial Intelligence in Tactical Human Resource Management: A Systematic Literature Review" which presents a systematic literature review focused on the application of artificial intelligence in tactical HRM. Their analysis demonstrates that AI plays an important role in processes such as recruitment, education and performance management. However, the authors emphasize the need for better data, standardization and practical integration of AI technologies into HR processes. Similarly, Sithambaram & Tajudeen (2023), in a survey based on interviews in 12 Malay companies, confirm that recruitment is one of the key application areas of AI, along with talent management and learning. They conclude that AI does not replace workers, but rather complements them, encouraging the development of skills and enabling focus on strategic activities. Johnson, Stone & Lukaszewski (2020), in the paper „The benefits of e-HRM and AI

Porast objava tijekom pandemije COVID-19 može se povezati s potrebom za brzim prilagodbama i implementacijom digitalnih rješenja u upravljanju ljudskim potencijalima, osobito u kontekstu rada na daljinu, digitalne komunikacije i automatizacije administrativnih procesa LJP-a.

Navedeno potvrđuje da je e-ULJP postao važno istraživačko područje koje se brzo razvija. Trend rasta uočava se već od 2018. godine, kada se intenziviraju rasprave o digitalnoj transformaciji funkcije LJP-a, osobito kroz implementaciju umjetne inteligencije, prediktivne analitike, chatbotova, sustava za praćenje učinka i digitalnih platformi za upravljanje talentima.

Među najvažnijim temama koje se izdvajaju u recentnoj literaturi su: digitalizacija procesa regrutacije i selekcije, personalizacija razvoja zaposlenika putem analitike, digitalna evaluacija radnog učinka te etički izazovi i privatnost podataka zaposlenika u digitalnim sustavima LJP-a. Na taj način literatura ne samo da odražava tehnološke promjene, već i pruža uvid u transformaciju ULJP-a iz administrativne funkcije prema strateškoj ulozi, uz sve veću potporu digitalnim alatima i podatkovno vođenom upravljanju LJP-a.

5 Pregled najznačajnijih radova prema odabranoj metodologiji

Digitalna transformacija prakse LJP-a u literaturi se najčešće tematizira kroz nekoliko ključnih područja: (1) regrutacija, selekcija i automatizacija procesa, (2) učenje, razvoj i korisničko iskustvo, (3) analitika i donošenje odluka, te (4) strateški uloga i digitalna transformacija LJP-a.

Votto i sur. (2021) pripremili su rad "*Artificial Intelligence in Tactical Human Resource Management: A Systematic Literature Review*" koji predstavlja sustavni pregled literature usmjeren na primjenu umjetne inteligencije u taktičkom ULJP-u. Njihova analiza pokazuje da UI ima važnu ulogu u procesima kao što su regrutacija, edukacija i upravljanje učinkom. Međutim, autori ističu potrebu za boljim podacima, standardizacijom i praktičnom integracijom UI tehnologija u procese LJP-a. Slično tome, Sithambaram & Tajudeen (2023), u istraživanju temeljenom na intervjuima u 12 malajskih tvrtki, potvrđuju da je regrutacija jedno od ključnih područja primjene UI, zajedno s upravljanjem talentima i učenjem. Zaključuju kako UI ne zamjenjuje radnike, već ih nadopunjuje, potičući razvoj vještina i omogućujući fokusiranje na strateške aktivnosti. Johnson, Stone & Lukaszewski (2020), u radu *„The benefits of e-HRM and AI for talent acquisition“*, povezuju e-ULJP i UI s potrebama sektora turizma i ugostiteljstva. Ističu da

for talent acquisition“, connect e-HRM and AI with the needs of the tourism and hospitality sector. They point out that AI improves the finding, selection and retention of qualified candidates, which contributes to reducing the costs and time of replacing employees.

Stanley & Aggarwal (2019) analyze how disruptive technologies, including AI, robots and networked systems, are transforming the recruitment, training and legal support functions. Robots, according to their findings, play a role in education and conducting interviews, while network systems improve connectivity within the organization. Majumder & Mondal (2021) focus on chatbots, emphasizing their role in increasing operational efficiency and automating routine interactions. However, they point out that in order to have a structural impact on HR, organizations must develop AI systems that can understand and respond to the emotional aspects of communication, thereby building trust.

Radonjić, Duarte & Pereira (2024) in the work „Artificial intelligence and HRM: HR managers' perspective on decisiveness and challenges“ show that AI improves decision-making in HR through faster data processing. The authors emphasize that HR managers are willing to accept AI as a data management tool, but also that there are challenges associated with strategic alignment and organizational structure.

Sharma et al. (2022), based on analyzes and interviews in the Indian energy sector, highlight three strategic priorities: digital HR and employee performance, implementation of digital HR, and digital HR as a competitive advantage. They conclude that AI can support the development of more efficient and personalized HR strategies, which contributes to overall organizational efficiency. Kim et al. (2021) provide a long-term perspective of technology development in HRM, analyzing changes in 154 papers over 60 years. They identify key milestones (computers, internet) and conclude that technology, including AI, is continuously redefining HR functions such as recruitment, training and assessment.

6 Conclusion and future scope

In conclusion, it is possible to answer all 3 key research questions. The first research question How has the scope and focus of e-HRM research evolved in the last decade? demonstrates a significant increase in scientific papers in the period from 2015 to 2025. The peak was in 2024 when the largest number of publications was directed towards the digital transformation of the HR function and the applications of artificial intelligence. This trend indicates a growing interest among scientists in technologically assisted HR practices, especially in

UI unaprjeđuje pronalaženje, selekciju i zadržavanje kvalificiranih kandidata, čime doprinosi smanjenju troškova i vremena zamjene zaposlenika.

Stanley & Aggarwal (2019) analiziraju kako disruptivne tehnologije, uključujući UI, robote i mrežne sustave, transformiraju funkcije zapošljavanja, obuke i pravne podrške. Roboti, prema njihovim nalazima, imaju ulogu u edukaciji i vođenju intervjua, dok mrežni sustavi poboljšavaju povezanost unutar organizacije. Majumder & Mondal (2021) fokusiraju se na chatbotove, naglašavajući njihovu ulogu u povećanju operativne učinkovitosti i automatizaciji rutinskih interakcija. Ipak, ističu da za ostvarenje strukturnog utjecaja na LJP, organizacije moraju razviti UI sustave koji mogu razumjeti i odgovarati na emocionalne aspekte komunikacije, čime se gradi povjerenje.

Radonjić, Duarte & Pereira (2024) u radu „*Artificial intelligence and HRM: HR managers' perspective on decisiveness and challenges“* pokazuju da UI unaprjeđuje donošenje odluka u LJP-u kroz bržu obradu podataka. Autori naglašavaju kako je prisutna spremnost menadžera LJP-a na prihvaćanje UI kao alata za upravljanje podacima, ali i da postoje izazovi povezani s strateškim usklađivanjem i organizacijskom strukturom.

Sharma i sur. (2022), temeljem analiza i intervjua u indijskom energetskom sektoru, ističu tri strateška prioriteta: digitalni LJP i učinak zaposlenih, implementacija digitalnog LJP-a te digitalni LJP kao konkurentska prednost. Zaključuju da UI može podržati razvoj efikasnijih i personaliziranih strategija LJP-a, što doprinosi ukupnoj organizacijskoj učinkovitosti. Kim i sur. (2021) daju dugoročnu perspektivu razvoja tehnologije u ULJP-u, analizirajući promjene u 154 rada kroz 60 godina. Prepoznaju ključne prekretnice (računala, internet) i zaključuju da tehnologija, uključujući UI, kontinuirano redefinira funkcije LJP-a poput regrutacije, obuke i procjene.

6 Zaključak i budući istraživački ciljevi

Zaključno, moguće je odgovoriti na sva 3 ključna istraživačka pitanja. Prvo istraživačko pitanje Kako je obujam i fokus istraživanja e-ULJP-a evoluirao u posljednjem desetljeću? pokazuje značajan porast znanstvenih radova u razdoblju od 2015. do 2025. godine. Vrhunac je bio 2024. godine kada je najveći broj objava bio usmjeren prema digitalnoj transformaciji funkcije LJP-a i primjenama umjetne inteligencije. Ovaj trend ukazuje na rastući interes

the context of dynamic changes in the work environment and digital adaptation of organizations. The second research question Who are the authors, topics and institutions that have shaped this field? identified scientists who are recognized through the citation and relevance of their works, which is shown through an overview of the most significant research to date in Chapter 5. The analysis identified key thematic groups that encompass technological implementation in HR (such as process automation, the use of digital platforms and learning systems), as well as a strategic approach to HR management through the use of advanced analytics and digital support in HR management. The third research question What is the role of artificial intelligence in the context of e-HRM? she identified several areas of application of artificial intelligence. First of all, artificial intelligence is used to predict patterns in employee behavior, to automate the selection process, as well as to make informed decisions, which significantly contributes to improving the strategic role of the HR function. In the context of the growing need for agile and personalized HR solutions, AI stands out as a key driver of changes in the HR approach.

An analysis of eight key scientific papers demonstrates that digital technologies, especially artificial intelligence, are increasingly being used in modern human resources management. The authors agree that AI has the potential to significantly improve selection, employee development, decision-making and user experience, using tools such as chatbots and predictive analytics. At the same time, challenges such as ethical dilemmas, lack of transparency in algorithmic decisions and the possibility of discrimination are highlighted. A deeper understanding of the impact of technologies on employees, their perception of fairness and trust in digital systems is needed.

Recommended directions for future research include the development of frameworks and tools for the ethical application of AI in HR management, as well as empirical research on the functionality of digital tools in different organizations (e.g. small vs. large companies, public vs. private sector). It is also recommended to conduct longitudinal studies to track the effects of digital transformation of the HR function over time, and comparative analyses between sectors with different levels of digital maturity.

For the purpose of a deeper analysis of this topic, future research should include additional criteria in literature selection, such as methodological approach, research context (industry, country) and evaluation of organizational effects, which would enable a more precise mapping of the impact of digital transformation on HRM practice.

znanstvenika za tehnološki potpomognute prakse LJP-a, posebno u kontekstu dinamičnih promjena u radnom okruženju i digitalne prilagodbe organizacija. Drugo istraživačko pitanje Tko su autori, teme i institucije koje su oblikovale ovo područje? identificiralo je znanstvenike koji su prepoznati kroz citiranost i relevantnost radova što je i prikazano kroz pregled najznačajnijih dosadašnjih istraživanja u poglavlju 5. Analizom su utvrđene ključne tematske skupine koje obuhvaćaju tehnološku implementaciju u LJP-u (poput automatizacije procesa, korištenja digitalnih platformi i sustava za učenje), kao i strateški pristup upravljanju LJP-a kroz upotrebu napredne analitike i digitalne podrške u upravljanju LJP-a. Treće istraživačko pitanje Kakva je uloga umjetne inteligencije u kontekstu e-ULJP-a? identificirala je nekoliko područja primjene umjetne inteligencije. Prije svega umjetna inteligencija se koristi za predviđanje obrazaca u ponašanju zaposlenika, automatizaciji procesa selekcije, kao i za donošenje utemeljenih odluka što značajno doprinosi unaprjeđenju strateške uloge funkcije LJP-a. U kontekstu rastuće potrebe za agilnim i personaliziranim rješenjima LJP-a, UI se ističe kao ključni pokretač promjena u pristupu LJP-a.

Analiza osam ključnih znanstvenih radova pokazuje da se digitalne tehnologije, osobito umjetna inteligencija, sve više koriste u suvremenom upravljanju ljudskim potencijalima. Autori se slažu da UI ima potencijal za značajno poboljšanje selekcije, razvoja radnika, donošenja odluka i korisničkog iskustva, koristeći alate kao što su chatbotovi i prediktivna analitika. Istovremeno se ističu izazovi poput etičkih dilema, nedostatka transparentnosti u algoritamskim odlukama i mogućnosti diskriminacije. Potrebno je dublje razumijevanje utjecaja tehnologija na zaposlenike, njihovu percepciju pravednosti i povjerenje u digitalne sustave.

Preporučeni smjerovi za buduća istraživanja uključuju razvoj okvira i alata za etičku primjenu umjetne inteligencije u upravljanju LJP-a, kao i empirijska istraživanja funkcionalnosti digitalnih alata u različitim organizacijama (npr. mala vs. velika poduzeća, javni vs. privatni sektor). Preporuka je i provođenje longitudinalnih studija koje bi pratile učinke digitalne transformacije funkcije LJP-a tijekom vremena te komparativne analize između sektora s različitim razinama digitalne zrelosti.

U svrhu dublje analize ove teme, buduća istraživanja trebala bi uključiti dodatne kriterije pri selekciji literature, kao što su metodološki pristup, kontekst istraživanja (industrija, zemlja) i evaluacija organizacijskih učinaka, čime bi se omogućilo preciznije mapiranje utjecaja digitalne transformacije na praksu ULJP-a.

Acknowledgments

This work has been fully supported by the Croatian Science Foundation under the project IP-2022-10-6703.

References

Bondarouk, T., & Brewster, C. (2016). Conceptualising the future of HRM and technology research. *The International Journal of Human Resource Management*, 27(21), 2652-2671.

Budhwar, P. et al. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*, 33(3), 606–659.

Johnson, R.D., Stone, D.L., Lukaszewski, K.M. (2020). The benefits of eHRM and AI for talent acquisition. *Journal of Tourism Futures*. 7(1), 40-52.

Kim, S., Wang, Y., Boon, C. (2021). Sixty years of research on technology and human resource management: Looking back and looking forward. *Special Issue: 60th Anniversary Issue of HRM*, 60(1), 229-247.

Majumder, S., Mondal, A. (2021). Are chatbots really useful for human resource management? International Journal of Speech Technology, 24, 969–977.

Marler, J. & Fisher, S. (2013). An evidence-based review of E-HRM and strategic human re-source management. *Human Resource Management Review*, 23(1), 18–36.

Marler, J. & Parry, E. (2015). Human resource management, strategic involvement and e-HRM technology. *The International Journal of Human Resource Management*, 27(19), 1-21.

Mwita, K. & Kitole, F. (2025). Potential benefits and challenges of artificial intelligence in human resource management in public institutions. Discover Global Society. 3(35), 1-19

Nyathi, M. & Kekwaletswe, R. (2022). Realizing employee and organizational performance gains through electronic human resource management use in developing countries. *African Journal of Economic and Management Studies*, 14(1), 121-134.

Radonjić, A., Duarte, H., Pereira, N. (2024). Artificial intelligence and HRM: HR managers'

Zahvale

Ovaj rad je u potpunosti podržan od strane Hrvatske zaklade za znanost u okviru projekta IP-2022-10-6703.

Reference

Bondarouk, T., & Brewster, C. (2016). Conceptualising the future of HRM and technology research. The International Journal of Human Resource Management, 27(21), 2652-2671.

Budhwar, P. et al. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. Human Resource Management Journal, 33(3), 606–659.

Johnson, R.D., Stone, D.L., Lukaszewski, K.M. (2020). The benefits of eHRM and AI for talent acquisition. Journal of Tourism Futures. 7(1), 40-52.

Kim, S., Wang, Y., Boon, C. (2021). Sixty years of research on technology and human resource management: Looking back and looking forward. Special Issue: 60th Anniversary Issue of HRM, 60(1), 229-247.

Majumder, S., Mondal, A. (2021). Are chatbots really useful for human resource management? International Journal of Speech Technology, 24, 969–977.

Marler, J. & Fisher, S. (2013). An evidence-based review of E-HRM and strategic human re-source management. Human Resource Management Review, 23(1), 18–36.

Marler, J. & Parry, E. (2015). Human resource management, strategic involvement and e-HRM technology. The International Journal of Human Resource Management, 27(19), 1-21.

Mwita, K. & Kitole, F. (2025). Potential benefits and challenges of artificial intelligence in human resource management in public institutions. Discover Global Society. 3(35), 1-19

Nyathi, M. & Kekwaletswe, R. (2022). Realizing employee and organizational performance gains through electronic human resource management use in developing countries. African Journal of Economic and Management Studies, 14(1), 121-134.

Radonjić, A., Duarte, H., Pereira, N. (2024). Artificial intelligence and HRM: HR managers'

perspective on decisiveness and challenges. *European Management Journal*, 42(1), 57-66.

Ruël, H.J.M. & Bondarouk, T. & Looise, J. K. (2004). E-HRM: Innovation or Irritation. An Explorative Empirical Study in Five Large Companies on Web-based HRM. Management Revue - The international Review of Management Studies, 15(3).

Sharma, A. et al. (2022). Review on Digitalisation and Artificial Intelligence in Human Resource Function of Energy Sector. Water and Energy International, 65r(2), 38-46.

Sithambaram, R.A., Tajudeen, F.P. (2023). Impact of artificial intelligence in human resource management: a qualitative study in the Malaysian context. Asia Pacific Journal of Human Resources, 61(4), 821-844.

Stanley, D.S., Aggarwal, V. (2019). Impact of disruptive technology on human resource management practices. International Journal of Business Continuity and Risk Management, 9(4), 350-361.

Strohmeier, S. (2020). Digital human resource management: A conceptual clarification. German Journal of Human Resource Management: Zeitschrift für Personalforschung, 34(3), 345-365.

Strohmeier, S. (2007). Research in E-HRM: Review and Implications. Human Resource Management Review, 17, 19-37.

Votto, A.M. et al. (2021). Artificial Intelligence in Tactical Human Resource Management: A Systematic Literature Review. International Journal of Information Management Data Insights, 1(2), 2667-0968.

perspective on decisiveness and challenges. European Management Journal, 42(1), 57-66.

Ruël, H.J.M. & Bondarouk, T. & Looise, J. K. (2004). E-HRM: Innovation or Irritation. An Explorative Empirical Study in Five Large Companies on Web-based HRM. Management Revue - The international Review of Management Studies, 15(3).

Sharma, A. et al. (2022). Review on Digitalisation and Artificial Intelligence in Human Resource Function of Energy Sector. Water and Energy International, 65r(2), 38-46.

Sithambaram, R.A., Tajudeen, F.P. (2023). Impact of artificial intelligence in human resource management: a qualitative study in the Malaysian context. Asia Pacific Journal of Human Resources, 61(4), 821-844.

Stanley, D.S., Aggarwal, V. (2019). Impact of disruptive technology on human resource management practices. International Journal of Business Continuity and Risk Management, 9(4), 350-361.

Strohmeier, S. (2020). Digital human resource management: A conceptual clarification. German Journal of Human Resource Management: Zeitschrift für Personalforschung, 34(3), 345-365.

Strohmeier, S. (2007). Research in E-HRM: Review and Implications. Human Resource Management Review, 17, 19-37.

Votto, A.M. et al. (2021). Artificial Intelligence in Tactical Human Resource Management: A Systematic Literature Review. International Journal of Information Management Data Insights, 1(2), 2667-0968.

IT project management: Does sustainability matter?

Upravljanje IT projektima: Je li održivost važna?

Marija Vugrinec
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
mvugrinec20@student.foi.hr

Katarina Pažur Aničić
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
kpazur@foi.hr

Abstract. *Sustainability is important in many aspects of human life today, and this paper focuses on the sustainability of IT projects. IT projects face a dynamic environment and numerous challenges, raising the question of to what extent and how sustainability principles are considered.*
The paper reviews different types of sustainability in the IT sector, with a focus on IT projects. It analyzes the presence of sustainability topics in the literature, proposes their integration across the project life cycle, and presents an example of sustainability in an IT project. The paper lays the groundwork for further research on sustainability in IT projects.

Keywords. Sustainability, IT projects

Sažetak. *Održivost je pojam koji je u današnje vrijeme važan u raznim aspektima ljudskog života, a u ovom radu naglasak je na održivosti IT projekata. IT projekte karakterizira dinamično okruženje te brojni izazovi. Postavlja se pitanje u kojoj se mjeri i na koji način u IT projektima uvažavaju načela održivosti.*
Ovaj rad donosi pregled vrsta održivosti u IT sektoru općenito, s naglaskom na IT projekte. Rad obuhvaća analizu zastupljenosti tema održivosti u literaturi, donosi prijedlog održivosti kroz faze životnog ciklusa projekta te opisuje konkretan primjer održivosti u IT projektu. Rad pruža podlogu za daljnje istraživanje održivosti u konkretnim primjerima IT projekata.

Ključne riječi. Održivost, IT projekti

1 Introduction

The concept of sustainability and sustainable development is increasingly being mentioned in various areas of everyday life, including the IT sector. Although there is no universally accepted definition of sustainability, one of the most well-known definitions comes from the World Commission on Environment and Development, published in the 1987 report *Our Common Future*, which defines sustainability as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (World Commission on Environment and Development, 1987). Sustainability can be divided into several dimensions. For example, the World Commission on Environment and Development identifies three dimensions of sustainability: people, profit, and planet. These are also referred to as the economic, environmental, and social dimensions of

1 Uvod

Pojam održivosti i održivog razvoja sve se češće spominje u različitim područjima svakodnevnog života, pa tako i u IT sektoru. Iako ne postoji jedinstvena definicija održivosti, jedna od najpoznatijih definicija je ona Svjetske komisije za okoliš i razvoj, objavljena 1987. godine u izvještaju Our Common Future, prema kojoj je održivost „razvoj koji zadovoljava potrebe današnjice, a pritom ne ugrožava potrebe budućih generacija“ (Svjetska komisija za okoliš i razvoj, 1987). Održivost se može podijeliti na više dimenzija. Primjerice, Svjetska komisija za okoliš i razvoj dijeli održivost na tri dimenzije: ljudi, profit i planet. Te dimenzije još se nazivaju ekonomska, ekološka i društvena (socijalna) dimenzija. Pojedini autori uvode i dodatne dimenzije, kao što su tehnička, individualna, i kulturna održivost (Amri i Bellamine Ben Saoud, 2014; Moreira et al, 2023; Soini i Dessein, 2016; Swacha, 2022).

U ovom radu fokus je na održivosti u IT projektima. Upravljanje IT projektima samo je po sebi vrlo

sustainability. Some authors introduce additional dimensions, such as technical, individual, and cultural sustainability (Amri and Bellamine Ben Saoud, 2014; Moreira et al., 2023; Soini and Dessein, 2016; Swacha, 2022).

This paper focuses on sustainability in IT projects. Managing IT projects is, in itself, a highly dynamic and challenging field. In general, PMI defines project as "a temporary endeavor undertaken to create a unique product, service, or result" (Project Management Institute, 2013, p. 3), while IT projects "involve the use of hardware, software, and networks to create a product, service, or result" (Schwalbe, 2019, p. 3). By their nature, IT projects most often involve software development, in which case we refer to software project management (Peters, 2024); however, there are also infrastructure IT projects, consulting projects, upgrade projects, and others. One of the most frequently cited statistics related to IT projects is the percentage of projects that fail — meaning those that exceed their planned time and budget — which has remained consistently high since The Chaos Report (Standish Group, 1995), showing that only 16.2% of software projects were completed on time and within budget. The reason for this likely lies in the complexity of IT projects, which is reflected in the numerous elements that comprise the project ecosystem. These include project goals, scope, team size, number and interests of stakeholders, organizational and technological complexity, interdependencies among project elements, and many other factors (Morcov et al., 2021; Vidal and Marle, 2008). Given the rapid pace of technological advancement, IT projects can be considered a prime example of a VUCA (Volatility, Uncertainty, Complexity, and Ambiguity) environment, whose characteristics significantly impact project (un)success (Moura et al., 2023). Considering the complexity of IT projects and the environment in which they operate, as well as the numerous factors influencing project success, it is not surprising that sustainability is often not a primary focus for project managers (Clinning and Marnewick, 2017).

Therefore, the main goal of this paper is to provide a more detailed analysis of current research in the field of sustainability related to IT projects and to propose various types of sustainable actions throughout the phases of the project life cycle.

2 Types of sustainability

To better understand sustainability in the context of IT projects, this chapter first describes the fundamental dimensions of sustainability.

The economic dimension of sustainability refers to the sustainability of an organization's business operations. Economically sustainable business

dinamično i izazovno područje. Općenito, projekt se definira kao „privremeni pothvat poduzet radi stvaranja jedinstvenog proizvoda, usluge ili rezultata" (Project Management Institute, 2013, p. 3), dok IT projekti „uključuju korištenje hardvera, softvera i mreža za stvaranje proizvoda, usluge ili rezultata" (Schwalbe, 2019, p.3). Po svojoj prirodi, IT projekti najčešće uključuju razvoj softvera, pri čemu se govori o upravljanju projektima softverskih proizvoda (engl. *software project management*) (Peters, 2024), no također postoje i infrastrukturni IT projekti, konzultantski projekti, projekti nadogradnje itd. Jedna od najčešćih statistika koja se veže uz IT projekte jest postotak neuspješno završenih projekata, odnosno projekata koji su završili prekoračenjem planiranog vremena i budžeta, a koji je kontinuirano vrlo visok, još od The Chaos Reporta (Standish Group, 1995) koji pokazuje da samo je samo 16.2% softverskih projekata završeno na vrijeme i u okviru budžeta. Razlog tome sigurno leži u kompleksnosti IT projekata koja se ogleda u raznim elementima koji čine ekosustav projekta, a koji obuhvaćaju ciljeve projekta, opseg projekta, veličinu projektnog tima, broj i interese dionika, organizacijsku kompleksnost, tehnološku kompleksnost, međuovisnosti između projektnih elemenata i mnoge druge faktore (Morcov i sur., 2021; Vidal i Marle, 2008). Uzme li se u obzir brzi napredak tehnologije, može se zaključiti da IT projekti predstavljaju pravi primjer VUCA (engl. *Volatility, Uncertainty, Complexity, and Ambiguity*) okruženja koje karakteriziraju volatilnost, nesigurnost, kompleksnost i dvoznačnost, pri čemu karakteristike takvog okruženja zasigurno utječu na (ne)uspjeh projekta (Moura i sur., 2023). Uzevši u obzir kompleksnost IT projekata i okruženja u kojem djeluju, te brojne faktore koji utječu na uspjeh projekt, nije neobično da održivost nije u fokusu voditelja upravljanja projektima (Clinning i Marnewick, 2017).

Stoga je glavni cilj ovog rada detaljnije analizirati trenutna istraživanja u području održivosti vezane uz IT projekte te predložiti različite vrste održivih akcija kroz faze životnog ciklusa projekta.

2 Vrste održivosti

Radi boljeg razumijevanja održivosti u kontekstu IT projekata, ovo poglavlje najprije donosi opis osnovnih dimenzija održivosti.

Ekonomska dimenzija održivosti odnosi se na održivost poslovanja organizacije. Ekonomski održivo poslovanje znači ostvarivanje prihoda koji su veći od troškova, odnosno poslovanje s profitom. Ova dimenzija održivosti također uključuje sve ono što doprinosi povećanju prihoda organizacije, kao što je povećanje kvalitete proizvoda i usluga. Clinning i Marnewick (2017) ističu kako je upravo ekonomska dimenzija još uvijek najvažnija dimenzija održivosti, osobito za profitne organizacije. Naime, organizacija koja zanemari ekološku ili socijalnu dimenziju i dalje

means generating revenue that exceeds costs — in other words, operating at a profit. This dimension of sustainability also encompasses everything that contributes to increasing an organisation's revenue, such as enhancing the quality of products and services. Clinning and Marnewick (2017) emphasize that the economic dimension remains the most important aspect of sustainability, especially for for-profit organizations. An organization that neglects its environmental or social dimension may continue to operate, but a profit-oriented organization that operates at a loss cannot survive (Clinning and Marnewick, 2017). Economic sustainability is reflected in the rational management of financial resources, investment in quality to generate higher profits, and cost reduction (Doane and MacGillivray, 2001).

Environmental sustainability is often the first association that comes to mind when the term "sustainability" is mentioned. It refers to the protection of the environment to minimise negative environmental impacts. Among other things, environmental sustainability includes the rational use of resources, efficient energy consumption, the use of renewable energy sources, reduction of CO2 emissions, extending product life cycles, repairing instead of replacing products, reducing waste, recycling, and disposing of waste in an environmentally friendly manner (Sutton, 2024).

Social sustainability refers to the well-being of all community members in which an individual or organisation operates. Social sustainability encompasses communication, safety, transparency, mutual honesty, and trust between individuals and groups, as well as the inclusion of diverse social groups in decision-making processes, the acceptance of diversity, and equal opportunities for all (Clinning and Marnewick, 2017; Moreira et al., 2023).

Technical sustainability refers to the properties and performance of a technological product. It encompasses the durability, adaptability, resilience, modularity, and safety of hardware and software. A technically sustainable product should be functional (performing its intended purpose) and operate without errors. It should also be safe, reliable, and adaptable to the user's needs. A technically sustainable product should be modular, meaning that one component of the product can be replaced without the need to purchase an entirely new product and discard the old one (Moreira et al., 2023). In the context of IT projects, the concept of technical debt is well known. According to Gartner, technical debt represents "a deviation of a system from any of its non-functional requirements" (Williams, 2025). As such, technical debt negatively impacts sustainability by reducing the long-term adaptability of systems to new technologies, increasing maintenance costs, and

može poslovati, ali profitna organizacija koja posluje s gubitkom ne može opstati (Clinning i Marnewick, 2017). Ekonomska održivost očituje se u racionalnom upravljanju financijskim sredstvima, ulaganju u kvalitetu kako bi se ostvario veći profit te smanjenju troškova (Doane i MacGillivray, 2001).

Ekološka održivost mnogima je prva asocijacija na spomen održivosti. Označava zaštitu okoliša uz nastojanje da se smanji negativan utjecaj na okoliš. Između ostalog, ekološka održivost uključuje racionalno korištenje resursa, racionalnu potrošnju energije, korištenje energije iz obnovljivih izvora, smanjenje emisija CO2, produljenje životnog vijeka proizvoda, popravak proizvoda umjesto zamjene, smanjenje količine otpada, recikliranje te zbrinjavanje otpada na ekološki prihvatljiv način (Sutton, 2024).
Društvena održivost odnosi se na dobrobit svih pripadnika zajednice u kojoj djeluje pojedinac ili organizacija. Društvena održivost uključuje komunikaciju, sigurnost, transparentnost, međusobnu iskrenost i povjerenje između pojedinaca i skupina, uključivanje različitih društvenih skupina u donošenje odluka, prihvaćanje različitosti i jednake prilike za sve (Clinning i Marnewick, 2017; Moreira i sur., 2023).

Tehnička održivost odnosi se na svojstva i performanse tehnološkog proizvoda. Označava dugotrajnost, prilagodljivost, otpornost, modularnost i sigurnost korištenja hardvera ili softvera. Tehnički održiv proizvod trebao bi biti funkcionalan (raditi ono čemu je namijenjen) te raditi bez grešaka. Također, trebao bi biti siguran za korištenje, pouzdan i prilagodljiv prema potrebi korisnika. Tehnički održiv proizvod trebao bi biti modularan, što znači da se jedna komponenta proizvoda može zamijeniti bez potrebe da se nabavlja cijeli novi proizvod, a stari se baca (Moreira i sur., 2023). U kontekstu IT projekata poznat je pojam tzv. tehničkog duga (engl. technical debt) koji, prema Gartneru, predstavlja „odstupanje sustava od bilo kojeg od njegovih nefunkcionalnih zahtjeva" (Williams, 2025). Kao takav, tehnički dug negativno utječe na održivost obzirom da smanjuje dugoročnu prilagodljivost sustava novih tehnologijama, povećava troškove održavanja i općenito uključuje negativne prakse koje utječu na održivost.

Neki autori dodatno navode individualnu i kulturnu dimenziju održivosti. Obje su usko vezane uz društvenu održivost. Individualna održivost više je fokusirana na dobrobit svakog pojedinca, za razliku od društvene održivost koja označava dobrobit društva u cjelini. Primjer individualne održivosti je dobrobit svakog zaposlenika u nekoj organizaciji – količina posla, primjerenost radnih zadataka, plaća, radni uvjeti, potrebna znanja i vještine (Amri i Bellamine Ben Saoud, 2014). Kulturna održivost se odnosi na očuvanje kulturne baštine, običaja i tradicije (Soini i Dessein, 2016).

U Tablici 1 sažeto su navedene i opisane dimenzije održivosti te autori koji spominju ove dimenzije u svojim radovima.

generally involving practices that harm sustainability.

Some authors additionally mention individual and cultural dimensions of sustainability. Both are closely related to social sustainability. Individual sustainability focuses more on the well-being of each individual, as opposed to social sustainability, which refers to the overall well-being of society. An example of individual sustainability is the well-being of each employee in an organization — workload, appropriateness of tasks, salary, working conditions, required knowledge, and skills (Amri and Bellamine Ben Saoud, 2014). Cultural sustainability refers to the preservation of cultural heritage, customs, and traditions (Soini and Dessein, 2016).

Table 1 summarizes and describes the dimensions of sustainability as well as the authors who mention these dimensions in their works.

Table 1. Sustainability dimensions (source: authors)

Sustainability dimension	Description	Authors
Economic sustainability	Refers to the sustainability of an organization's business operations. It includes increasing revenue, reducing costs, and operating profitably.	Amri & Bellamine Ben Saoud (2014), Clinning & Marnewick (2017), Doane & MacGillivray (2001), Santarius & Wagner (2023)
Environmental sustainability	Refers to reducing negative impacts on the environment. It involves rational energy consumption, ecological waste management, and the use of environmentally friendly materials.	Amri & Bellamine Ben Saoud (2014), Clinning & Marnewick (2017), Santarius & Wagner (2023), Sutton (2024)
Social sustainability	Refers to the well-being of all members of society. It includes communication, transparency, cooperation, and various social aspects.	Amri & Bellamine Ben Saoud (2014), Clinning & Marnewick (2017), Moreira et al. (2023), Santarius & Wagner (2023)

Tablica 1. Dimenzije održivosti (izvor: autori)

Dimenzija održivosti	Opis	Autori
Ekonomska održivost	Odnosi se na održivost poslovanja organizacije. Uključuje povećanje prihoda, smanjenje troškova te poslovanje s profitom.	Amri i Bellamine Ben Saoud (2014), Clinning i Marnewick (2017), Doane i MacGillivray (2001), Santarius i Wagner (2023)
Ekološka održivost	Odnosi se na smanjenje negativnog utjecaja na okoliš. Uključuje racionalnu potrošnju energije, ekološko zbrinjavanje otpada i korištenje ekološki prihvatljivih materijala	Amri i Bellamine Ben Saoud (2014), Clinning i Marnewick (2017), Santarius i Wagner (2023), Sutton (2024)
Društvena (socijalna) održivost	Odnosi se na dobrobit svih pripadnika društva. Uključuje komunikaciju, transparentnost, suradnju i različite društvene aspekte.	Amri i Bellamine Ben Saoud (2014), Clinning i Marnewick (2017), Moreira et al. (2023), Santarius i Wagner (2023)
Tehnička održivost	Odnosi se na karakteristike tehnološkog proizvoda.	Moreira et al. (2023), Swacha (2022)
Individualna održivost	Usko vezana uz društvenu održivost. Odnosi se na dobrobit pojedinca.	Amri i Bellamine Ben Saoud (2014), Moreira et al. (2023), Swacha (2022)
Kulturna održivost	Odnosi se na održavanje određene kulture, baštine i običaja.	Soini i Dessein (2016),

3 Održivost u IT sektoru

Kad se govori o održivosti u IT projektima i IT sektoru općenito, potrebno je razlikovati „održivi IT“ i „IT za održivost“. „Održivi IT“ označava informacijsko-komunikacijske tehnologije (IKT) koje su same po sebi održive. To može značiti da su ekološki prihvatljivije od drugih tehnologija – primjerice, troše manje energije ili su napravljene od materijala koji se može reciklirati (Swacha, 2022). S druge strane, „IT za održivost“ predstavlja tehnologije koje doprinose postizanju ciljeva održivosti u širem smislu To mogu biti

Technical sustainability	Refers to the characteristics of a technological product.	Moreira et al. (2023), Swacha (2022)
Individual sustainability	Closely linked to social sustainability. Refers to the well-being of the individual.	Amri & Bellamine Ben Saoud (2014), Moreira et al. (2023), Swacha (2022)
Cultural sustainability	Refers to the preservation of a specific culture, heritage, and customs.	Soini & Dessein (2016),

3 Sustainability in the IT Sector

When discussing sustainability in IT projects and the IT sector in general, it is important to distinguish between "sustainable IT" and "IT for sustainability." "Sustainable IT" refers to information and communication technologies (ICT) that are sustainable in themselves. This can mean that they are more environmentally friendly than other technologies—for example, they consume less energy or are made from recyclable materials (Swacha, 2022). On the other hand, "IT for sustainability" represents technologies that contribute to achieving sustainability goals in a broader sense. These can be technologies that digitise processes previously performed on paper, thereby enabling a reduction in paper waste. Another example is the use of technologies that optimise production processes, resulting in resource savings (Santarius and Wagner, 2023). It is important to acknowledge that, in addition to their positive impact, ICT can also have negative effects on sustainability, particularly in the environmental dimension. This is most evident in electricity consumption (both in the production and use of devices) and the disposal of devices that are no longer in use (Santarius and Wagner, 2023).

IT projects can also be divided into "sustainable projects" and "projects for sustainability." A "sustainable project" is one in which sustainable practices are implemented in all phases of the project execution. In contrast, a "project for sustainability" is a project whose outcome contributes in some way to one of the dimensions of sustainability (Soares, Fernandes, and Santos, 2024).

An organization striving for sustainability should introduce as many sustainable practices as possible into its projects. However, this is not always easy, as there are contradictions between project management and sustainability. While tehnologije koje digitaliziraju određene procese za koje se prije koristio papir te na taj način omogućavaju smanjenje količina papirnatog otpada. Drugi primjer su tehnologije za optimizaciju procesa u proizvodnji, koje doprinose uštedi resursa (Santarius i Wagner, 2023). Važno je napomenuti da IKT osim pozitivnog, mogu imati i negativni utjecaj na održivost, osobito ekološku dimenziju. To se najviše očituje u potrošnji električne energije (kod proizvodnje i korištenja uređaja) te zbrinjavanju uređaja koji se više ne koriste (Santarius i Wagner, 2023).

IT projekti također se mogu podijeliti na „održive projekte“ i „projekte za održivost“. „Održivi projekt“ je onaj u kojem su održive prakse implementirane u sve faze provođenja projekta, dok je „projekt za održivost“ onaj projekt čiji rezultat na neki način doprinosi nekoj od dimenzija održivosti (Soares, Fernandes i Santos, 2024).

Organizacija koja teži održivosti idealno bi trebala uvesti što više održivih praksi u svoje projekte. Međutim, to nije uvijek jednostavno, jer postoje kontradikcije između upravljanja projektima i održivosti. Dok su projekti jedinstveni, s jasno određenim trajanjem, početkom i krajem, održivi razvoj nema početak ni kraj. Održivost podrazumijeva pogled u budućnost, koja može biti bliža i dalja, i traženje dugoročnih rješenja. U projektu je cilj provesti sve planirane aktivnosti i doći do rezultata, nakon čega projekt završava. Održivi razvoj nema kraja – ako je cilj da nešto bude održivo, cilj je da traje što je dulje moguće. Projekt ima definiran tim, voditelja, ciljnu skupinu i druge zainteresirane skupine – u projekt je obično uključen manji broj ljudi. S druge strane, održivost se odnosi na društvo u cjelini, i to ne samo na sadašnje, već i buduće generacije. Nije moguće jednim projektom ostvariti sve ciljeve čovječanstva vezane uz održivost, no zato se manjim projektima mogu ostvariti specifični ciljevi manjih zajednica (Clinning i Marnewick, 2017).

4 Održivost u IT projektima u literaturi

Kako bi se dobio detaljniji uvid u zastupljenost održivosti u kontekstu IT projekata u znanstvenoj literaturi, napravljena je jednostavnija bibliometrijska analiza u bazi Web of Science (WoS). Upit na bazu TI=((IT OR "Information technology") AND project AND manag*) rezultirao je s ukupno 502 rada, a analiza prema ključnim riječima napravljena je korištenjem alata VOSWiever (2025). Identificirano je ukupno 1433 ključnih riječi, od kojih njih 75 ima više od 5 pojavljivanja, međutim među njima se nijednom ne pojavljuje ključna riječ „sustainability“.

Dodavanje ključne riječi „sustain*“ u temu rada (TI=((IT OR "Information technology") AND project AND manag*) AND TS=sustain*) rezultiralo je sa samo 13 radova. Od njih, samo su tri rada zapravo

projects are unique, with clearly defined durations, beginnings, and endings, sustainable development has no start or finish. Sustainability implies a forward-looking perspective, which can be near-term or long-term, and seeks solutions that address long-term needs. The goal of a project is to carry out all planned activities and achieve the desired results, after which the project concludes. Sustainable development, on the other hand, has no end — if the goal is for something to be sustainable, the aim is for it to last as long as possible. A project has a defined team, a project manager, a target group, and other stakeholders — usually involving a smaller number of people. Sustainability, on the other hand, refers to society as a whole, including not only current but also future generations. It is not possible to achieve all humanity's sustainability goals with a single project; however, smaller projects can achieve specific goals for smaller communities (Clinning and Marnewick, 2017).

4 Sustainability in IT projects in the literature

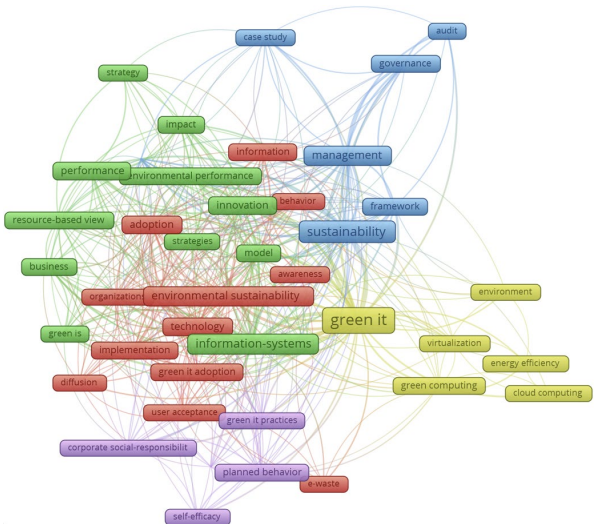
In order to get a more detailed insight into the representation of sustainability in the context of IT projects in the scientific literature, a simpler bibliometric analysis was made in the Web of Science (WoS) database. The query to the database TI=((IT OR "Information technology") AND project AND manag*) resulted in a total of 502 works, and the analysis according to keywords was done using the VOSWiever tool (2025). A total of 1433 keywords were identified, of which 75 have more than 5 occurrences, but the keyword "sustainability" does not appear among them.

Adding the keyword "sustain*" to the topic of the paper (TI=((IT OR "Information technology") AND project AND manag*) AND TS=sustain*) resulted in only 13 papers. Of these, only three papers are actually focused on sustainability in IT projects. Pantović et al. (2024) emphasize that ensuring sustainability in IT project management is an ongoing challenge and propose a new approach focused on data-driven strategies to improve sustainability and excellence in IT project management. Woźniak (2021) proposes and verifies the concept of a sustainable approach to IT project management by involving the client in the selection of project management methodology. Another sustainable approach in terms of the role of stakeholders in project management, and especially in IT service management projects, was presented by Zaleski and Michalski (2021), who tried to identify the success factors of an IT service project, including both traditional and agile approaches and taking into account sustainable development.

As mentioned in the introduction, IT projects are most often software development projects, so an additional search was made in order to investigate usmjerena na održivost u IT projektima. Pantović i sur. (2024.) naglašavaju da je osiguranje održivosti u upravljanju IT projektima stalni izazov i predlažu novi pristup usmjeren na strategije utemeljene na podacima kako bi se poboljšala održivost i izvrsnost u upravljanju IT projektima. Woźniak (2021.) predlaže i provjerava koncept održivog pristupa upravljanju IT projektima uključivanjem klijenta u odabir metodologije upravljanja projektima. Drugi održivi pristup u smislu uloge dionika u upravljanju projektima, a posebice u projektima upravljanja IT uslugama, predstavili su Zaleski i Michalski (2021.), koji su pokušali identificirati čimbenike uspjeha projekta IT usluga, uključujući i tradicionalne i agilne pristupe te uzimajući u obzir održivi razvoj.

Kao što je uvodno bilo spomenuto, IT projekti najčešće su projekti razvoj softvera pa je stoga napravljena dodatna pretraga kako bi se detaljnije istražilo spominje li se i u kojoj mjeri održivost vezana konkretnije uz razvoj softvera. Pojmovi koji se spominju u literaturi su „green IT" i „green software engineering", stoga je upit uključio navedene ključne riječi: TI=("green software" OR "green IT") i rezultirao s 302 rada, među kojima su detektirane 52 ključne riječi s više od 5 pojavljivanja i 22 ključne riječi s više od 10 pojavljivanja.

Slika 1. prikazuje najučestalije ključne riječi s više od 5 pojavljivanja, pri čemu se može zaključiti da tri najveća klastera obuhvaćaju teme koje se odnose na 1) energetsku učinkovitost i održivi razvoj softvera, 2) teme koje obuhvaćaju inovacije i prihvaćanje „green IT" te 3) svjesnost, ponašanje i ekološku održivost. Iako bi za konkretnije zaključke trebalo provesti detaljnju analizu detektiranih radova, iz navedenog se može zaključiti da raznici oblici održivosti kroz faze životnog ciklusa projekta u dosadašnjoj literaturi nisu zastupljeni u većoj mjeri.



Slika 1. Najučestalije ključne riječi u radovima u bazi WoS vezanima uz "Green software" i "Green IT", izvor: autori, korištenjem WoS baze i VOSviewer alata

in more detail whether and to what extent sustainability is mentioned more specifically related to software development. The terms mentioned in the literature are "green IT" and "green software engineering", so the query included the specified keywords: TI=("green software" OR "green IT") and resulted in 302 papers, among which 52 keywords with more than 5 occurrences and 22 keywords with more than 10 occurrences were detected.

Fig. 1 shows the most frequent keywords with more than 5 occurrences, whereby it can be concluded that the three largest clusters include topics related to 1) energy efficiency and sustainable software development, 2) topics covering innovation and adoption of "green IT", and 3) awareness, behaviour and environmental sustainability. Although more specific conclusions would require a detailed analysis of the analysed papers, it can be concluded from the above that various forms of sustainability across the project life cycle phases are not represented to a greater extent in the literature to date.

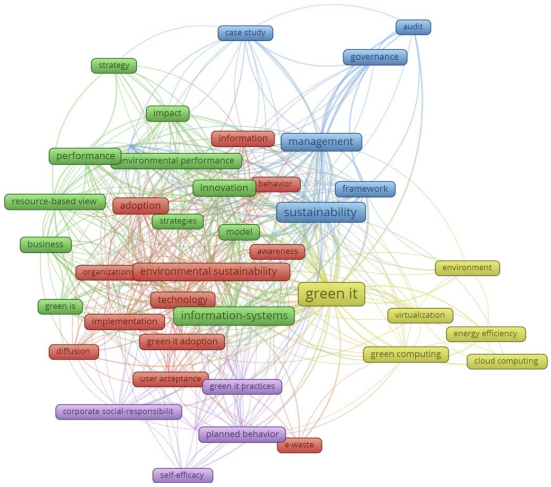


Figure 1. Most frequent keywords in papers in the WoS database related to “Green software” and “Green IT”, source: authors, using WoS database and VOSviewer

The importance of sustainability research in IT projects is supported by the fact that a conference called *ICT for Sustainability (ICT4S)* has been held once a year for more than 10 years. It brings together ICT experts, industry representatives and public administration representatives (ICT4S, 2025). According to Rugeviciute, Courboulay and Hilty (2023), the mapping of research communities according to the ICTS4S model indicates various aspects of sustainability in IT that are intertwined - *Green ICT, Green IS* and *Environmental Informatics* study the impact of ICT on the environment; *Sustainable Software Engineering (SSE)* and *Sustainable Human-Computer Interaction* study the impact of ICT on the

U prilog važnosti istraživanja održivosti u IT projektima ide činjenica da da se već više od 10 godina jednom godišnje održava konferencija naziva *ICT for Sustainability (ICT4S).* Ona okuplja stručnjake iz područja IKT-a, predstavnike industrije te predstavnike javne uprave (ICT4S, 2025). Prema Rugeviciute, Courboulay i Hilty (2023), mapiranje istraživačkih zajednica prema ICTS4S modelu ukazuje na razne aspekte održivosti u IT-ju koji se međusobno isprepliću - *Green ICT, Green IS* i *Environmental Informatics* proučavaju utjecaj IKT na okoliš; *Sustainable Software Engineering (SSE)* i *Sustainable Human-Computer Interactio*n proučavaju utjecaj IKT na okoliš i društvo; *ICT for Development* također istražuje utjecaj tehnologije na društvo, s fokusom na zemlje u razvoju; zajednica *Corporate Digital Sustainability* bavi se odgovornim donošenjem odluka u organizacijama, a *ICT & Ethics* pitanjima etike u IT sektoru.

Autori Santarius i Wagner (2023) analizirali su ukupno 211 radova s ICT4S konferencija, objavljenih u razdoblju od 2013. do 2019. godine. Podijelili su radove u tri razine, prema načinu na koji pristupaju održivosti. Autori definiraju razine na sljedeći način:

- Prva razina odnosi se na održivost tehnološkog proizvoda - materijal od kojeg je sastavljen, energiju potrošenu za proizvodnju, energiju koju troši proizvod dok se koristi te resurse potrebne za recikliranje i zbrinjavanje nakon korištenja proizvoda
- Druga razina odnosi se na doprinos tehnologije održivosti
- Treća razina odnosi se na strukturalne promjene koje nastaju od utjecajem tehnologije, a vezane su uz održivost (npr. promjene zakona ili društvenih normi)

Dodatno, isti autori utvrdili su da se većina radova s ICT4S konferencija, njih 111 od 211, bavi drugom razinom održivosti. Nešto manje radova, njih 66, bavi se prvom razinom. Tek 34 rada bave se trećom razinom održivosti (Santarius i Wagner, 2023). To pokazuje da su istraživanja u području održivosti u IT-ju uglavnom fokusirana na načine na koje digitalne tehnologije mogu doprinijeti održivosti u drugim područjima.

5 Održivost kroz faze IT projekta

Sukladno životnom ciklusu projekta (Project Management Institute, 2013), i IT projekte karakterizira 5 osnovnih faza životnog ciklusa: pokretanje projekta (inicijacija), planiranje projekta, provedba, zatvaranje te kontinuirano nadgledanje i kontrola. Prethodna istraživanja već ukazuju na održive prakse prema fazama upravljanja projektima (Soares et al., 2024), a u nastavku ovog rada slijedi prijedlog održivih akcija te pitanja o kojima je važno promisliti u kontekstu životnog ciklusa IT projekta.

environment and society; *ICT for Development* also explores the impact of technology on society, with a focus on developing countries; the *Corporate Digital Sustainability* community deals with responsible decision-making in organizations, and *ICT & Ethics* deals with ethical issues in the IT sector.

Authors Santarius and Wagner (2023) analysed a total of 211 papers from ICT4S conferences, published between 2013 and 2019. They divided the papers into three levels based on their approach to sustainability. The authors define the levels as follows:

- The first level refers to the sustainability of the technological product - the material from which it is composed, the energy consumed in production, the energy consumed by the product while in use, and the resources needed for recycling and disposal after use of the product
- The second level refers to the contribution of technology to sustainability
- The third level refers to the structural changes that arise from the impact of technology, and are related to sustainability (e.g. changes in laws or social norms).

Additionaly, based on their analysis, Santarius and Wagner (2023) found that the majority of papers from ICT4S conferences, 111 out of 211, deal with the second level of sustainability. Slightly fewer papers, 66, deal with the first level. Only 34 papers deal with the third level of sustainability. This indicates that research in the field of sustainability in IT primarily focuses on how digital technologies can contribute to sustainability in other areas.

5 Sustainability through IT project phases

According to the project life cycle (Project Management Institute, 2013), IT projects are also characterized by 5 basic phases of the life cycle: project initiation, project planning, implementation, closure, and continuous monitoring and control. Previous research has already proposed sustainable practices according to project management phases (Soares et al., 2024). In continuation of this work, a proposal of sustainable actions and questions that are important to consider in the context of an IT project's life cycle follows.

5.1 Project initiation

Sustainability can be an integral part of a project from the very first phase, project initiation. Before implementing a project, its financial viability is usually examined, which is an element of the

5.1 Pokretanje projekta

Održivost može biti sastavni dio projekta već od prve faze, pokretanja projekta. Prije provođenja projekta obično se ispituje njegova financijska isplativost, što je element ekonomske dimenzije održivosti (Soares et al., 2024). Prilikom ispitivanja izvodljivosti projekta na isti način mogu se razmotriti ostale dimenzije održivosti. Primjerice, može se ispitati kakav će utjecaj projekt imati na okoliš. Ovo je posebno važno ako se u sklopu projekta razvija hardver, odnosno neki fizički proizvod. Ako organizacija želi da taj proizvod bude ekološki održiv, može razmisliti o korištenju ekološki prihvatljivih materijala u proizvodnji te o ekološkom zbrinjavanju proizvoda na kraju njegova životnog vijeka. Također, treba sagledati utjecaj projekta na društvo. Tim koji pokreće projekt treba se zapitati koje će pozitivne i negativne učinke projekt imati prvo na sve koji su direktno uključeni, a zatim i na širu zajednicu. Tehnička održivost osobito je važna u IT projektima i već u fazi pokretanja dobro je razmisliti o tome hoće li finalni proizvod (bilo softver ili hardver) biti održiv.

5.2 Planiranje projekta

U sljedećoj fazi, planiranju projekta, već se mogu planirati konkretni načini na koje će projekt biti održiv. Ekonomska održivost najviše se očituje u planiranju budžeta, koji treba biti takav da omogući sve planirane aktivnosti, ali da ne stvara nepotrebne dodatne troškove za organizaciju. Ekološka održivost može se uključiti na način da se maksimalno iskoristi postojeća oprema (računala, mobilni uređaji itd.) umjesto nabavljanja nove (što doprinosi i ekonomskoj održivosti). Primjerice, ako se razvoja softver za virtualnu stvarnost (vengl. Virtual reality - VR) naočale, organizacija može kupiti naočale za potrebe testiranja, dok za pisanje koda može koristiti postojeća računala. Soares et al. (2024) navode da se prilikom nabave za potrebe projekta može napraviti evaluacija dobavljača i odabrati one dobavljače koji posluju održivo. Ugljični otisak projekta može se smanjiti tako da se sastanci projektnog tima održavaju online (ako se članovi tima nalaze na različitim lokacijama), da se izbjegnu putovanja. Projektna dokumentacija i ostale datoteke mogu se dijeliti online umjesto ispisivanja na papir.

Društvena održivost očituje se u uključivanju različitih skupina dionika u planiranje projekta, primjerice provedbom istraživanja s klijentima i budućim krajnjim korisnicima. Prije provedbe projekta dobro se zapitati koja će znanja biti potrebna krajnjim korisnicima za korištenje sustava i planirati edukacije. Clinning i Marnewick (2017) navode da je jednako važno osigurati da članovi projektnog tima imaju potrebna znanja te osigurati edukacije za njih. Komunikacija s članovima projektnog tima, osiguravanje potrebnih sredstava za rad, pružanje dodatnih edukacija i pravedna raspodjela posla ulaze u

economic dimension of sustainability (Soares et al., 2024). When examining the feasibility of a project, other dimensions of sustainability can be considered in the same way. For example, the impact of the project on the environment can be examined. This is especially important if the project develops hardware, or a physical product. If the organization wants the product to be environmentally sustainable, it can consider using environmentally friendly materials in production and environmentally friendly disposal of the product at the end of its life. The impact of the project on society should also be considered. The team initiating the project should ask itself what positive and negative effects the project will have, first on everyone directly involved, and then on the wider community. Technical sustainability is particularly important in IT projects, and it is a good idea to consider whether the final product (whether software or hardware) will be sustainable already at the initiation stage.

5.2 Project planning

In the next phase, project planning, specific ways in which the project will be sustainable can already be planned. Economic sustainability is most evident in budget planning, which should enable all planned activities without creating unnecessary additional costs for the organisation. Environmental sustainability can be included by maximising the use of existing equipment (computers, mobile devices, etc.) instead of purchasing new ones (which also contributes to economic sustainability). For example, if software for virtual reality (VR) glasses is being developed, the organization can purchase glasses for testing purposes, while existing computers can be used for writing code. Soares et al. (2024) state that when procuring resources for a project, an evaluation of suppliers can be conducted to select those that operate sustainably. The carbon footprint of the project can be reduced by holding project team meetings online (if team members are in different locations), to avoid travel. Project documentation and other files can be shared online instead of being printed on paper.

Social sustainability is reflected in the inclusion of different stakeholder groups in project planning, for example by conducting research with clients and future end users. Before implementing the project, the organization should ask itself what knowledge end users will need to use the system and plan training. Clinning and Marnewick (2017) emphasize that it is equally important to ensure that project team members have the necessary knowledge and to provide training for them as well. Communication with project team members, securing the necessary resources for work, providing additional training and fair distribution of

društvenu i individualnu dimenziju održivosti. Prilikom planiranja svakako treba razmisliti o karakteristikama proizvoda koji se razvija, što ulazi u sferu tehničke održivosti. Ako se razvija softver, treba razmisliti o odgovorima na sljedeća pitanja: Na kojim uređajima će on raditi? Hoće li raditi i na starijim uređajima? Hoće li raditi na različitim operacijskim sustavima? Hoće li sustav biti modularan i omogućiti da se zamijeni jedna njegova komponenta, bez da se mijenja cijeli sustav? Kako će se sustav ažurirati, i hoće li nakon provedbe ažuriranja raditi na istom hardveru? Kakva će biti interakcija krajnjih korisnika sa sustavom? Hoće li moći sami raditi prilagodbe i promjene? Kako će prijavljivati i rješavati probleme?

Već u fazi planiranja može se razmišljati o održivosti rezultata projekta nakon završetka projekta. Koji je očekivani životni vijek proizvoda? Kako će se pružati podrška klijentu i krajnjim korisnicima? Hoće li se proizvod dalje razvijati? Kako će se to financirati? Kako će sustav utjecati na ekonomsku održivost klijenta? Hoće li klijent imati visoke troškove održavanja? U većim projektima uz projektni plan može se napraviti i poseban plan održivosti, gdje su nabrojene konkretne mjere koje doprinose različitim dimenzijama održivosti (Amri i Bellamine Ben Saoud, 2014).

5.3 Provedba projekta

U fazi provedbe projekta aktivnosti se provode u skladu s projektnim planom. Tijekom provođenja aktivnosti društvena održivost očituje se u upravljanju projektnim timom, kao i u komunikaciji te suradnji sa svim zainteresiranim stranama. U svakom projektu može doći do nezadovoljstva i otpora od strane klijenta ili korisnika. Rješavanje problema učinkovito i na vrijeme također je dio društvene održivosti. Ekonomska održivost odnosi se na prilagođavanje budžeta po potrebi, kako bi projekt i dalje bio održiv ako se pojave neočekivani troškovi. Ekološki održive prakse provode se onako kako su planirane u prethodnoj fazi – primjerice, sve datoteke se dijele digitalno. Potrošni materijal koji se nabavlja i koristi za potrebe projekta trebao bi se ekološki zbrinuti nakon korištenja. Na isti način treba planirati zbrinjavanje opreme koja je nabavljena za potrebe projekta, a nakon završetka projekta neće se više koristiti. Planiranih karakteristika proizvoda, koje su vezane uz tehničku održivost, trebali bi biti svjesni svi koji rade na dizajnu i izradi proizvoda. Održivost se može uključiti u redovno projektno izvještavanje, gdje se stvarno uključivanje održivih praksi može usporediti s planiranim. Također, održive prakse mogu se uključiti i „u hodu“, tijekom provođenja projekta, ako voditelj ili članovi tima primijete da bi se neke aktivnosti mogle provoditi na održiviji način.

5.4 Zatvaranje projekta

U posljednjoj fazi, zatvaranju projekta, dobro je pogledati unatrag i razmotriti koliko je projekt bio održiv. Može se napraviti usporedba planiranih i

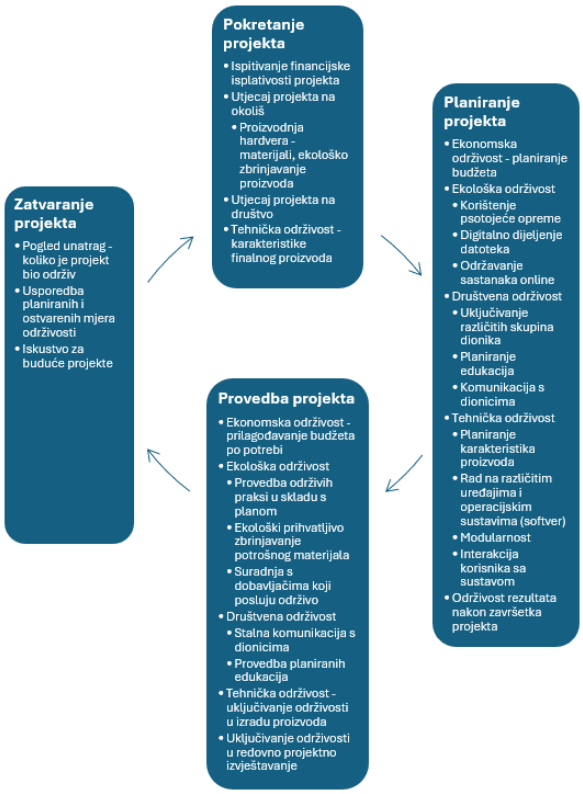
work are part of the social and individual dimensions of sustainability. When planning, the organization should definitely think about the characteristics of the product being developed, which falls into the sphere of technical sustainability. If software is being developed, the answers to the following questions should be rethought: On which devices will it work? Will it also work on older devices? Will it work on different information systems? Will the system be modular and allow for the replacement of one of its components without changing the entire system? How will the system be updated, and will it work on the same hardware after the update? What will be the interaction of end users with the system? Will they be able to make adjustments and changes themselves? How will they report and resolve problems?

Already in the planning phase, one can think about the sustainability of the project results after the project is completed. What is the expected lifespan of the product? How will the client and end users be supported? Will the product be further developed? How will it be financed? How will the system affect the client's economic sustainability? Will the client have high maintenance costs? In larger projects, a separate sustainability plan can be created in addition to the project plan, listing specific measures that contribute to different dimensions of sustainability (Amri and Bellamine Ben Saoud, 2014).

5.3 Project implementation

In the project implementation phase, activities are carried out in accordance with the project plan. During the implementation of activities, social sustainability is reflected in the project team's management, as well as in their communication and cooperation with all stakeholders. In every project, dissatisfaction and resistance from the client or user may occur. Solving problems effectively and promptly is also a key aspect of social sustainability. Economic sustainability refers to adjusting the budget as necessary, so that the project remains viable if unexpected costs arise. Environmentally sustainable practices are implemented as planned in the previous phase - for example, all files are shared digitally. Consumables purchased and used for the project should be environmentally disposed of after use. The disposal of equipment purchased for the project and no longer needed after it is completed should be planned in the same way. Everyone working on the design and construction of the product should be aware of planned product characteristics related to technical sustainability. Sustainability can be incorporated into regular project reporting, where the actual implementation of sustainable practices can be compared with the planned. Additionally, ostvarenih mjera održivosti, koliko su one zapravo koristile projektu te jesu li možda uzrokovale i neke probleme. Znanje i iskustvo stečeno na jednom projektu može se iskoristiti u budućim projektima (Soares et al., 2024).

Na Slici 2. prikazan je prijedlog održivih akcija prema fazama IT projekta.



Slika 2. Prijedlog održivih akcija po fazama IT projekta, izvor: autori

6 Primjer: održivost u projektu razvoja sustava za praćenje parametara u poljoprivredi

Nakon analize stručne literature vezane uz održivost u IT projektima, provedena je studija slučaja na konkretnom primjeru IT projekta. Radi se o projektu izrade uređaja za mjerenje parametara tla (temperature, vlage, pH vrijednosti) na polju te pripadnog softvera za upravljanje sustavom.

U ovom projektu održivost je uključena već od faze pokretanja projekta. Provedeno je istraživanje tržišta kako bi se ustanovilo postoji li interes za kupnju ovog sustava te ispitivanje financijske isplativosti projekta, što spada u ekonomsku održivost.

U fazi planiranja projekta uključena je ekološka održivost, primjerice prilikom nabave za potrebe projekta. Budući da tvrtka prije nije radila projekte u području poljoprivrede, bilo je potrebno nabaviti novu opremu. Provedena je analiza dobavljača te se nastojalo

sustainable practices can be incorporated "on the go" during the project implementation if the leader or team members notice that certain activities could be carried out in a more sustainable manner.

5.4 Project closure

In the final phase, project closure, it is good to look back and consider how sustainable the project was. A comparison can be made of the planned and implemented sustainability measures, how much they actually benefited the project and whether they may have caused any problems. The knowledge and experience gained in one project can be used in future projects (Soares et al., 2024).
Fig. 2 shows a proposal of sustainable actions throughout the phases of an IT project.

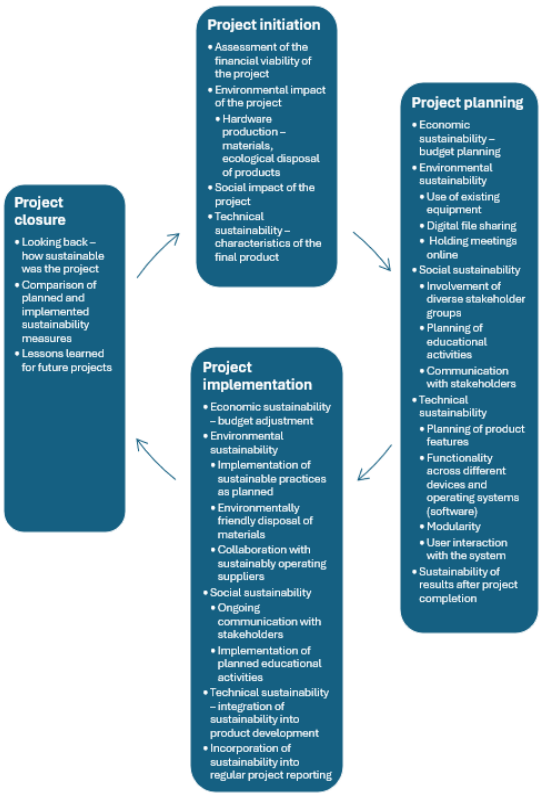


Figure 2. A proposal of sustainable actions throughout the phases of an IT project, source: authors

6 Case study: sustainability in an agricultural parameter monitoring system development project

After analyzing professional literature related to sustainability in IT projects, a case study was

poslovati s dobavljačima koji i sami posluju održivo. Također, kod projektiranja uređaja odabrani su materijali koji nisu štetni za tlo u kojem je uređaj postavljen. Planirano je i ekološki prihvatljivo zbrinjavanje opreme i materijala nakon završetka projekta.

U sve faze, od pokretanja do provedbe projekta uključeni su klijenti te je s njima održavana konstantna komunikacija, a u suradnji s drugim poduzećem – distributerom sustava organizirane su edukacije za korisnike sustava, što upućuje na društvenu održivost projekta. U fazi provedbe korištene su prakse koje se mogu smatrati ekološki održivima, kao što su digitalno dijeljenje datoteka i održavanje sastanaka online, iako primarni razlog tome nije bio težnja ekološkoj održivosti, već jednostavno praktičnost jer dio članova tima radi na daljinu. U fazi završetka projekta višak opreme i materijala ekološki je zbrinut, a sve što se može iskoristiti u budućem radu tvrtke je prenamijenjeno.

Ovo je primjer projekta koji nije samo održiv, već je i projekt „za održivost“. Sam sustav je tehnički održiv, a doprinosi trima dimenzijama održivosti – ekonomskoj, ekološkoj i društvenoj. Poljoprivrednicima omogućava efikasnije upravljanje proizvodnjom kroz praćenje parametara tla u realnom vremenu. Primjerice, praćenje količine minerala u tlu doprinosi racionalnijoj upotrebni gnojiva, što je dobro i za okoliš, i za smanjenje troškova poljoprivrednika. Također, pruža priliku manjim poljoprivrednim gospodarstvima da u svoje poslovanje uključe naprednu tehnologiju, koja nije preskupa i jednostavna je za korištenje.

7 Diskusija i zaključak

U ovom radu analizirana je održivost u IT projektima. Iako je najčešća asocijacija uz pojam održivost ona ekološka, u literaturi su evidentni i drugi oblici održivosti koji uključuju ekonomsku, društvenu, tehničku, individualnu i kulturnu dimenziju održivosti. Analiza znanstvene literature u bazi WoS ukazuje na činjenicu da održivost u upravljanju IT projektima nije u većoj mjeri zastupljena kroz dosadašnji rad istraživača. Ovaj rad dio je šireg istraživanja koje se provodi u okviru izrade diplomskog rada prve autorice, a koje uključuje analizu nekoliko studija slučaja na konkretnim primjerima IT projekata u organizacijama privatnog i javnog sektora u Republici Hrvatskoj, s ciljem istraživanja u kojoj mjeri su razne dimenzije održivosti zastupljene u praksi. Jedan od primjera provedene studije slučaja opisan je i ovom radu.

Obzirom na važnost održivosti na globalnoj razini te ubrzani razvoj tehnologije, potrebni su dodatni napori kako bi se tema održivosti u IT projektima i IT sektoru dodatno istražila i počela aktivnije primjenjivati u praksi. Ovaj rad pruža početne smjernice za aktivnosti koje potiču održivo ponašanje kroz faze životnog ciklusa IT projekta.

conducted on a concrete IT project. The project involved the development of a device for measuring soil parameters (temperature, moisture, pH level) in the field, along with accompanying software for system management.

In this project, sustainability was integrated from the initiation phase. Market research was carried out to determine whether there was interest in purchasing such a system, along with an assessment of the project's financial viability, which falls under economic sustainability.

In the planning phase, ecological sustainability was taken into account, for example, during procurement for the project. Since the company had not previously worked on agricultural projects, it was necessary to acquire a new equipment. A supplier analysis was conducted, and efforts were made to work with suppliers that operate sustainably. Additionally, during the device design phase, materials that are not harmful to the soil were selected. Environmentally friendly disposal of equipment and materials after the project's completion was also planned. Throughout all phases—from initiation to implementation—clients were actively involved and maintained in continuous communication.

In cooperation with another company, which served as the system distributor, training sessions were organized for users, indicating the project's social sustainability. During the implementation phase, practices considered environmentally sustainable were used, such as digital file sharing and holding online meetings. However, these were primarily adopted for practicality, as some team members worked remotely, rather than out of a specific commitment to ecological sustainability. At the project's conclusion, surplus equipment and materials were disposed of in an environmentally responsible way, and everything that could be reused in the company's future work was repurposed.

This is an example of a project that is not only sustainable but also a project for sustainability. The system itself is technically sustainable and contributes to all three dimensions of sustainability: economic, ecological, and social. It enables farmers to manage production more efficiently by monitoring soil parameters in real time. For instance, monitoring mineral levels in the soil supports more rational use of fertilizers, which benefits both the environment and reduces costs for farmers. It also provides smaller agricultural holdings with the opportunity to integrate advanced technology into their operations—technology that is affordable and easy to use.

7 Discussion and conclusion

This paper analyses sustainability in IT projects. Although the most common association with the

Reference

Amri, R. & Bellamine Ben Saoud, N. (2014). Towards a Generic Sustainable Software Model. *2014 Fourth International Conference on Advances in Computing and Communications (ICACC)*, 231-234. https://doi.org/10.1109/ICACC.2014.62

Clinning, G. & Marnewick, C. (2017). Incorporating sustainability into IT project management in South Africa. *South African Computer Journal* 29 (1), 1–26. https://doi.org/10.18489/sacj.v29i1.398

Doane, D. & MacGillivray, A. (2001). *Economic Sustainability - The business of staying in business.* New Economics Foundation.

ICT4S (2025). About the Conference. Preuzeto 18.6.2025. s https://conf.researchr.org/info/ict4s-2025/about

Morcov, S., Pintelon, L., & Kusters, R. (2021). Definitions, characteristics and measures of IT project complexity—A systematic literature review. *International Journal of Information Systems and Project Management*, 8(2), 5–21

Moreira, A., Araújo, J., Gralha, C., Goulão, M., Brito, I.S. & Albuquerque, D. (2023). A social and technical sustainability requirements catalogue. *Data & Knowledge Engineering*, 143, 102107. https://doi.org/10.1016/j.datak.2022.102107

Moura, R.L., Janes Carneiro, T.C. & Lemos Dias, T. (2023) VUCA environment on project success: The effect of project management methods. *Brasilian Business Review*. 20(3)

Pantović, V.; Vidojević, D.; Vujičić, S.; Sofijanić, S. & Jovanović-Milenković, M. (2024) Data-Driven Decision Making for Sustainable IT Project Management Excellence. *Sustainability* 2024, 16 (7), 3014. https://doi.org/10.3390/su16073014

Peters, L.J., Software Project Management (2024) *Methods and Techniques*. CRC Press. Taylor & Francis Group

Project Management Institute (2013) *A Guide to the Project Management body of Knowledge (PMBOK guide)*. Fifth Edition. Project Management Institute

Rugeviciute, A., Courboulay V. & Hilty, L.M. (2023). The research landscape of ICT for sustainability: harnessing digital technology for sustainable development. *2023 International Conference on ICT for Sustainability (ICT4S).* (97-107). Rennes, France.

Santarius, T. & Wagner, J. (2023). Digitalization and sustainability: A systematic literature analysis of ICT for Sustainability research. *GAIA - Ecological*

term sustainability is ecological, other forms of sustainability are evident in the literature, including economic, social, technical, individual, and cultural dimensions of sustainability. The analysis of the scientific literature in the WoS database suggests that sustainability in IT project management has not been extensively represented in the work of researchers to date. This work is part of a broader research project being conducted as part of the first author's master's thesis, which includes an analysis of several case studies conducted on specific examples of IT projects in private and public sector organizations in the Republic of Croatia to investigate the extent to which various dimensions of sustainability are represented in practice. One of the examples of the conducted case study is described in this paper.

Considering the importance of sustainability on a global level and the rapid development of technology, additional efforts are needed to investigate further the topic of sustainability in IT projects and the IT sector and to apply it more actively in practice. This paper provides initial guidelines for activities that promote sustainable behaviour throughout the phases of the IT project life cycle.

References

Amri, R. & Bellamine Ben Saoud, N. (2014). Towards a Generic Sustainable Software Model. *2014 Fourth International Conference on Advances in Computing and Communications (ICACC)*, 231-234. https://doi.org/10.1109/ICACC.2014.62

Clinning, G. & Marnewick, C. (2017). Incorporating sustainability into IT project management in South Africa. *South African Computer Journal* 29 (1), 1–26. https://doi.org/10.18489/sacj.v29i1.398

Doane, D. & MacGillivray, A. (2001). *Economic Sustainability - The business of staying in business.* New Economics Foundation.

ICT4S (2025). About the Conference. Retrieved from: https://conf.researchr.org/info/ict4s-2025/about

Morcov, S., Pintelon, L., & Kusters, R. (2021). Definitions, characteristics and measures of IT project complexity—A systematic literature review. *International Journal of Information Systems and Project Management*, 8(2), 5–21

Moreira, A., Araújo, J., Gralha, C., Goulão, M., Brito, I.S. & Albuquerque, D. (2023). A social and technical sustainability requirements catalogue. *Data & Knowledge Engineering*, 143, 102107. https://doi.org/10.1016/j.datak.2022.102107

Perspectives for Science and Society, 32(1), 21-32. https://doi.org/10.14512/gaia.32.S1.5

Soares, I.; Fernandes, G. & Santos, J.M.R.C.A. (2024). Sustainability in Project Management Practices. *Sustainability 2024*, 16 (10), 4275. https://doi.org/10.3390/su16104275

Soini, K. & Dessein, J. (2016). Culture-Sustainability Relation: Towards a Conceptual Framework. *Sustainability 2016*, 8 (2), 167. https://doi.org/10.3390/su8020167

Sutton, P. (2004). *A Perspective on environmental sustainability?* Victorian Commissioner for Environmental Sustainability.

Swacha, J. (2022). Models of Sustainable Software: A Scoping Review. *Sustainability 2022*, 14 (1), 551. https://doi.org/10.3390/su14010551

Schwalbe, K. (2019). Information Technology Project Management, 9th edition, Cengage Learning Inc.

Standish Group. (1995). *CHAOS report*. The Standish Group

Vidal, L., & Marle, F. (2008). Understanding project complexity: Implications on project management. *Kybernetes*, 37(8), 1094–1110. https://doi.org/10.1108/03684920810884928

VOSviewer (2025), Centre for Science and Technology Studies, Leiden University, The Netherlands, available at: https://www.vosviewer.com/

World Commission on Environment and Development (1987). *Our Common Future*, available at https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf

Woźniak, M. Sustainable Approach in IT Project Management—Methodology Choice vs. Client Satisfaction. *Sustainability 2021*, 13 (3), 1466. https://doi.org/10.3390/su13031466

Williams, R. (2025) *Reduce and Manage Technical Debt*, Gartner, available at: https://www.gartner.com/en/infrastructure-and-it-operations-leaders/topics/technical-debt

Zaleski, S. & Michalski, R. (2021) Success Factors in Sustainable Management of IT Service Projects: Exploratory Factor Analysis. *Sustainability* 2021, 13 (8), 4457. https://doi.org/10.3390/su13084457

Moura, R.L., Janes Carneiro, T.C. & Lemos Dias, T. (2023) VUCA environment on project success: The effect of project management methods. *Brasilian Business Review*. 20(3)

Pantović, V.; Vidojević, D.; Vujičić, S.; Sofijanić, S. & Jovanović-Milenković, M. (2024) Data-Driven Decision Making for Sustainable IT Project Management Excellence. *Sustainability* 2024, 16 (7), 3014. https://doi.org/10.3390/su16073014

Peters, L.J., Software Project Management (2024) *Methods and Techniques*. CRC Press. Taylor & Francis Group

Project Management Institute (2013) *A Guide to the Project Management body of Knowledge (PMBOK guide)*. Fifth Edition. Project Management Institute

Rugeviciute, A., Courboulay V. & Hilty, L.M. (2023). The research landscape of ICT for sustainability: harnessing digital technology for sustainable development. *2023 International Conference on ICT for Sustainability (ICT4S).* (97-107). Rennes, France.

Santarius, T. & Wagner, J. (2023). Digitalization and sustainability: A systematic literature analysis of ICT for Sustainability research. *GAIA - Ecological Perspectives for Science and Society*, 32(1), 21-32. https://doi.org/10.14512/gaia.32.S1.5

Soares, I.; Fernandes, G. & Santos, J.M.R.C.A. (2024). Sustainability in Project Management Practices. *Sustainability 2024*, 16 (10), 4275. https://doi.org/10.3390/su16104275

Soini, K. & Dessein, J. (2016). Culture-Sustainability Relation: Towards a Conceptual Framework. *Sustainability 2016*, 8 (2), 167. https://doi.org/10.3390/su8020167

Sutton, P. (2004). *A Perspective on environmental sustainability?* Victorian Commissioner for Environmental Sustainability.

Swacha, J. (2022). Models of Sustainable Software: A Scoping Review. *Sustainability 2022*, 14 (1), 551. https://doi.org/10.3390/su14010551

Schwalbe, K. (2019). Information Technology Project Management, 9th edition, Cengage Learning Inc.

Standish Group. (1995). *CHAOS report*. The Standish Group

Vidal, L., & Marle, F. (2008). Understanding project complexity: Implications on project management. *Kybernetes*, 37(8), 1094–1110. https://doi.org/10.1108/03684920810884928

VOSviewer (2025), Centre for Science and Technology Studies, Leiden University, The

Netherlands, available at: https://www.vosviewer.com/

World Commission on Environment and Development (1987). *Our Common Future*. Retrieved from: https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf

Woźniak, M. (2021) Sustainable Approach in IT Project Management—Methodology Choice vs. Client Satisfaction. *Sustainability 2021*, 13 (3), 1466. https://doi.org/10.3390/su13031466

Williams, R. (2025) *Reduce and Manage Technical Debt*, Gartner, available at: https://www.gartner.com/en/infrastructure-and-it-operations-leaders/topics/technical-debt

Zaleski, S. & Michalski, R. (2021) Success Factors in Sustainable Management of IT Service Projects: Exploratory Factor Analysis. *Sustainability* 2021, 13 (8), 4457. https://doi.org/10.3390/su13084457

The Influence of Blockchain-based Triple-entry Accounting on Accounting Information – a Theoretical and Conceptual Perspective

Tanja Sestanj-Peric
University of Zagreb Faculty of Organization and Informatics
Department of Economics
Pavlinska 2, Varaždin, Croatia
tanja.peric@foi.unizg.hr

Abstract. *This paper examines the conceptual and theoretical foundations of blockchain-based triple-entry accounting (TEA) as a potential evolution of the traditional double-entry system. By analyzing relevant economic theories and accounting concepts, the study explores how TEA can enhance the quality of financial information through increased transparency, security, and verifiability. The analysis highlights TEA's potential to reduce fraud, improve auditability, and strengthen trust among stakeholders in financial reporting. The findings offer a theoretical basis for understanding TEA's implications for accounting practices and support its potential role in the digital transformation of financial accounting.*

Keywords. triple-entry accounting, blockchain, distributed ledger, financial reporting, accounting information quality

1 Introduction

Digital transformation, automation, data science, and artificial intelligence already play a significant role in accounting and finance and are expected to become even more prominent as digital innovations continue to promote efficiency, accuracy, and transparency in financial processes.

Financial accounting, through financial statements, is an important source of business information, primarily serving external users. Despite a comprehensive regulatory framework, financial information remains vulnerable to manipulation and fraud, as seen in scandals such as Enron, WorldCom, and more recently, Toshiba.

Blockchain technology has significant potential to enhance transparency and accuracy in recording business transactions and financial reporting while reducing the risk of accounting fraud.

Although business transactions have become more complex, requiring continuous improvements in financial reporting standards, the fundamental principle by which business events and transactions are recorded, and later reflected in financial statements, has remained largely unchanged. This principle is the double-entry accounting system. Technological advancements are expected to allow for the evolution of this system into a triple-entry accounting system with the aim of preventing accounting fraud. The term triple-entry accounting was initially used in that context by Grigg in 2005 (Grigg, 2024). Distributed ledger technology, based on blockchain, could enable this transformation by preventing the manipulation of financial information and accounting fraud.

Most existing studies on the topic are theoretical and focus on the potential benefits of blockchain-based triple-entry accounting (hereinafter: TEA) without examining its relationship with underlying economic theories or fundamental accounting concepts.

This paper addresses that gap by exploring the potential impact of TEA on financial information from a theoretical and conceptual perspective.

The main objectives of this paper are to examine:

- Which issues, as explained by selected economic theories and mitigated by accounting information, would be further addressed by TEA-based financial reporting.
- Which accounting concepts would be influenced by TEA and how this would affect the quality of financial information.

The most important economic theories and accounting concepts underlying financial accounting will be analyzed to justify the transition from double-entry bookkeeping (accounting) to triple-entry accounting based on blockchain distributed ledger technology. This analysis provides a theoretical foundation for understanding the contribution of TEA to financial accounting, particularly its impact on security, transparency, and efficiency of financial reporting.

The paper is structured as follows: section two presents a review of selected papers; section 3 provides theoretical analysis; and section 4 concludes.

2 Review of Relevant Studies

Research about the application of blockchain in accounting started to grow beginning in 2019, according to searches of the Web of Science and Scopus databases.

Bonsón & Bednárová (2019), in their literature review, argue that blockchain technology has the potential to significantly transform accounting and auditing. It can enhance data reliability, transparency, and automation through features such as immutability, smart contracts, and decentralized consensus. The authors suggest that private blockchains may offer solutions for better auditability and reduced fraud. However, there are challenges such as scalability, cybersecurity, and regulatory acceptance.

Thies et al. (2023), in their systematic literature review on blockchain-based triple-entry accounting (hereinafter: TEA), note that the existing literature is almost entirely theoretical. As a result, various proposals for the design of TEA have emerged. Thies et al. adopt the view that TEA is a blockchain-supported extension of traditional double-entry accounting. Most of the authors whose articles they analyze have a positive outlook on this technology and its potential contribution to accounting and auditing, particularly in terms of fraud prevention, as well as faster availability and greater transparency of financial information. The challenges likely to arise in this context relate to legal and regulatory issues, cybersecurity, costs, and the fact that blockchain technology is not standardized.

Dai & Vasarhelyi (2017), in their theoretical work, argue for the significant potential contribution of blockchain to financial accounting, which still makes limited use of advances in information technology. They explore the idea of integrating Enterprise Resource Planning (ERP) systems with blockchain to create a triple-entry accounting system, enable real-time data processing and auditing, and transform the professional roles of accountants. The type of blockchain envisioned is a permissioned blockchain, in which only pre-approved entities can verify transactions. The triple-entry accounting system is seen as an enhancement of traditional double-entry accounting by adding a third entry—recorded on the blockchain—that creates immutable and transparent accounting data entries. This allows for real-time auditing and immediate availability of financial information to all relevant stakeholders. As a result, financial information becomes more transparent, and the risk of manipulation is reduced. These developments will influence accounting and auditing professionals, who are expected to take on more advisory roles while also requiring new knowledge in the field of information technology. To assess potential challenges, the authors apply the Technology-Organization-Environment (TOE) framework developed by Tornatzky and Fleischer (1990). From a technological standpoint, they highlight storage and computing resource costs, which necessitate adapting blockchain for accounting purposes. Organizationally, significant changes to business processes would be required, while from an environmental perspective, regulators would need to provide guidance and oversight regarding the application of blockchain in accounting.

Empirical research on the application of blockchain in accounting, as well as on the organizational factors influencing the decision to adopt blockchain in accounting, is still scarce.

Akter et al. (2024) investigated organizational factors that either promote or hinder the adoption of blockchain in accounting, as well as perceptions of the benefits of blockchain in this context. Semi-structured interviews were conducted with 19 accounting and blockchain experts, although none had actual experience with implementing blockchain-based accounting solutions. The researchers applied the Technology-Organization-Environment (TOE) framework. Among the key findings were challenges and a general lack of knowledge regarding the application of blockchain in accounting. The study identified nine organizational-level factors that either facilitate or hinder the intention to adopt blockchain in accounting. These include:

- Technological factors: perceived benefits, trialability, complexity, expected costs, and privacy risk
- Organizational factors: lack of knowledge, organizational innovativeness, and top management support
- Environmental factor: external pressure.

These insights highlight the multifaceted nature of blockchain adoption and the need to address both technological readiness and organizational capacity.

Cai (2021) uses case studies to explore triple-entry accounting. The selected cases focus on start-up projects aiming to establish blockchain-based triple-entry accounting systems. One case involves a public decentralized ledger project targeting clients with many customers and suppliers. The project is based on a cloud platform that could improve the efficiency and transparency of business processes related to customers and suppliers. However, potential issues include data privacy and the scalability of such a ledger. The main conclusions of the study are that, with the use of blockchain, data entry into the accounting system can become more efficient, while trust and transparency within the accounting system can be enhanced. The author emphasizes the need for further research into the conceptual and theoretical foundations of triple-entry accounting to strengthen the basis for its practical development.

Fang et al. (2023) conducted an empirical analysis on a sample of companies listed on Chinese stock exchanges to examine the impact of blockchain adoption on the quality of accounting information for the period from June 2016 to August 2020. The research shows that the adoption of blockchain significantly improves the quality of accounting information, as blockchain enhances corporate governance and creates synergy with large auditing firms, thereby improving the overall quality of financial reporting. A decentralized accounting information system strengthens internal controls and ensures the immutability of transactions, which helps prevent fraud. Furthermore, the study highlights that these positive effects are more pronounced in industries with less developed IT infrastructure and in cases where there have been no recent changes in auditing firms.

A study conducted by Oladejo et al. (2024) using a questionnaire at the end of 2020 and the beginning of 2021, involving 44 respondents, revealed that a majority of participants were skeptical about the potential of blockchain in the field of accounting. However, most of these respondents, although from the accounting or auditing domain, were not actively engaged with blockchain-related activities.

Busulwa et al. (2025) examined the knowledge of accountants in Australia and New Zealand regarding digital technologies during the period from mid-December 2021 to the end of March 2022. When it came to emerging technologies such as blockchain and other distributed ledgers and databases relevant to accounting, respondents rated their knowledge or understanding very low on a 1–5 scale, with both the median and mean values around 2. This underscores a key barrier to the adoption of emerging technologies in accounting.

Seshadrinathan & Chandra (2025) explored the impact of trust in blockchain technology on its adoption in accounting using the Technology-Organization-Environment (TOE) framework. In their model, TOE factors (firm-level influencers) mediate the relationship between trust in blockchain and the acceptance of blockchain for accounting purposes. The model was validated through qualitative interviews and quantitative surveys with accounting professionals.

3 Theoretical and Conceptual Foundations for Transitioning from Double-entry to Triple-entry Accounting

In line with Thies et al. (2023), TEA is in this paper considered as an enhancement of the traditional accounting system based on the concept of double-entry accounting (hereinafter: DEA), utilizing blockchain technology. In TEA, the traditional debit-credit recording in individual business ledgers is supplemented by a third entry: a cryptographically signed transaction receipt stored on a distributed ledger that is immutable and verifiable.

The key differences between DEA and TEA that are traced in literature are summarized in Table 1.

Table 1. Key differences between DEA and TEA

Feature	Double-Entry Accounting (DEA)	Triple-Entry Accounting (TEA)
Ledger Ownership	Internal	Distributed
Transparency	Limited (internal records)	High (shared, verifiable ledger)
Security	Prone to fraud and manipulation	Immutable and cryptographically secure
Auditability	Requires external audits	Real-time, automated external audits

Although accounting is heavily regulated, manipulation and fraud continue to undermine the value of accounting information.

This section explains how TEA could enhance financial accounting as a source of reliable financial information. It explores relevant economic theories and accounting concepts in this context.

3.1 Economic Theories Relevant to Accounting and TEA

To justify the transition from DEA to TEA, several economic theories can be applied. In this section the following theories are considered: efficient contracting theory, agency theory, stakeholder theory and fraud diamond theory.

3.1.1 Efficient Contracting Theory

Conceptual Framework for Financial Reporting, 2018 (hereinafter: CF) specifies that financial information must be decision useful since its primary users are investors, lenders and other creditors. Decision usefulness is one role of accounting information. Creditors and investors undertake financial statement analysis and use accounting information to predict cash flow and to make their decisions (Van Horne & Wachowicz, 2008, p. 128). They use accounting information to make decisions about investing in or lending to a firm. Subsequently, this information serves as credible evidence of past performance, allowing comparisons with forecasts and supporting the enforcement of contractual provisions, such as those stipulated in debt covenants within debt contracts.

Financial information also plays a stewardship role since it is used to assess if management fulfilled its goals or promises thus aligning the interests of managers and shareholders. If managers can

manipulate financial information trust is compromised (Scott, 2015, p. 16-17).

Efficient contracting theory explains how contracts are structured to maximize efficiency by aligning the interests of conflicting parties involved (Scott, 2015, p.313) - usually under conditions of asymmetric information and incomplete contracts. Applied to accounting i.e. financial reporting, this means debt or managerial compensation contracts that depend on accounting variables are the main reason why trustworthy accounting information is needed. Use of accounting information limits opportunistic behavior and enhances contracting efficiency by lowering contracting costs (Watts & Zimmerman, 1990).

TEA could promote contracting efficiency by reducing agency costs that are included in contracting costs, particularly in the following ways:

- Real-time availability of information
- Mutual confirmation of transactions (for transactions that involve other parties like suppliers or customers)
- Enhanced verifiability and auditability of information.

3.1.2 Agency Theory

Agency theory is characterized by agency relationship which consists of conflict of interests and asymmetry of information between principals and agents, like a firm and its creditors/investors (Wagenhofer, 2015).

Agency theory, when applied to financial accounting, helps explain why financial reporting standards, auditing practices, and corporate disclosures exist, all of that leading to reliable financial information. It focuses on how accounting information reduces the agency problems between agents (a firm or management) and principals (investor/creditor, shareholders or other stakeholders). Moral hazard represents hidden action taken by an agent that is not in the principal's interest; adverse selection is when an agent has more information than principal (Scott, 2015, pp 22-23).

Traditionally financial statements provide verifiable, standardized information that principals can use to evaluate agent's performance. However, basing compensation or debt contracts on accounting numbers gives managers an incentive to get involved in earnings management thus distorting accounting information (Wagenhofer, 2015). Auditing, conservatism integrated into accounting standards and disclosure requirements serve to deal with agency problems. TEA could further reduce both moral hazard and adverse selection problems by lowering agency costs via the following mechanisms:

- Immutable and secure transaction records that lead to more reliable financial information
- Decentralized and shared records that promote transparency and trust
- Enhanced auditability that prevents or diminishes incentives to manipulate.

3.1.3 Stakeholder theory

This theory emphasizes the importance of values in business, advocating for the creation of relationships of managers with not only shareholders but other stakeholders (for ex. employees, suppliers, customers, regulators etc.) to enhance value creation and rejecting the separation thesis between business and ethics. Shareholders and profits are important, but profits are the result of the value creation process. (Freeman et al., 2004)

In this view, accounting becomes a tool for accountability and transparency not only to investors but to all parties affected by a firm's actions. Stakeholder theory emphasizes ethical obligations in reporting, not just legal compliance. Traditionally different stakeholders (investors, auditors, regulators, business partners) depend on centralized financial records that may be altered or manipulated.

TEA supports ethical reporting by enabling transparent, verifiable access to information for multiple stakeholders. Access to a shared ledger would make transaction verifiable by all relevant parties thus reducing information asymmetry and enhancing transparency and trust.

3.1.4 Fraud Diamond Theory

This theory (Wolfe & Hermanson, 2004) attributes fraud occurrence to four factors: incentive, opportunity, rationalization and capability.

TEA addresses the "opportunity" component of the fraud diamond by ensuring transaction records are cryptographically verified, immutable, and transparent, thereby reducing the likelihood of fraudulent behavior.

3.2 Impact of TEA on Accounting Information Quality

3.2.1 Faithfull Representation

According to CF (2018), financial information must be relevant and faithfully represent economic transactions of an entity if it is to be decision useful primarily for investors and creditors, but also for other stakeholders. Faithfully represented information is the one that is maximally complete, neutral and free from error.

TEA would improve the last characteristic by reducing errors through automated, verifiable entries.

3.2.2 Qualitative Characteristics that Enhance the Usefulness of Information

Relevance and faithful representation, these two fundamental qualitative characteristics of financial information, are improved if financial information has additional characteristics of comparability, verifiability, timeliness and understandability. Comparability enables users to identify and understand similarities and differences among items. Information is verifiable if different knowledgeable and

independent observers can reach consensus that it represents what it purports to represent. Timeliness refers to information that is available in time to influence decisions. Information is understandable if presented clearly and concisely, allowing users to comprehend it. (Conceptual Framework for Financial Reporting, 2018)

TEA could positively affect some of these characteristics (see Table 2).

Table 2. TEA impact on financial information

Characteristic	TEA impact
Comparability	Improved through standardization and consistency
Verifiability	Strengthened through cryptographic proof and shared access
Timeliness	Enhanced via real-time recording and instant visibility

Firstly, standardized transactions structures and shared ledgers reduce inconsistencies between entities thus enabling better comparison of information of one entity with similar information about other entities.

Additionally, distributed ledger technology based on blockchain provides cryptographic proof of transactions and thus reduces dependence on subjective judgment or trust in internal systems so that other parties like auditors or regulators can verify transactions directly from the blockchain without needing company-specific access.

Finally, with TEA transactions would be recorded in real time and instantly visible to authorized stakeholders (suppliers, investors, auditors, …) improving timeliness.

4 Conclusion

This paper has examined the transformative potential of blockchain-based triple-entry accounting (TEA) through the lens of foundational economic theories and core accounting concepts. The analysis suggests that TEA, as an extension of traditional double-entry accounting, can enhance the transparency, accuracy, and reliability of financial information by leveraging blockchain's immutability, decentralization, and consensus mechanisms. Analyzing TEA through theories such as agency theory, stakeholder theory, efficient contracting, and the fraud diamond theory, the paper demonstrates how TEA could reduce information asymmetry, mitigate agency costs, and deter fraudulent behavior in financial reporting. Furthermore, it shows that TEA enhances key qualitative characteristics of financial information - particularly verifiability, timeliness, and faithful representation - offering a compelling theoretical basis for its potential integration into future financial accounting practices.

Further research and practical development will be essential in translating the theoretical advantages of TEA into real-world accounting practices.

References

Akter, M., Kummer, T.-F., & Yigitbasioglu, O. (2024). Looking beyond the hype: The challenges of blockchain adoption in accounting. International Journal of Accounting Information Systems, 53, 1–20. https://doi.org/10.1016/j.accinf.2024.100681

Bonsón, E., & Bednárová, M. (2019). Blockchain and its implications for accounting and auditing. Meditari Accountancy Research, 27(5), 725–740. https://doi.org/10.1108/MEDAR-11-2018-0406

Busulwa, R., Birt, J., Gepp, A., & Oates, G. (2025). Current State of Accountants' Knowledge of Digital Technologies: Evidence From Australia and New Zealand. Accounting & Finance. https://doi.org/10.1111/acfi.70010

Cai, C. W. (2021). Triple-entry accounting with blockchain: How far have we come? Accounting & Finance, 61(1), 71–93. https://doi.org/10.1111/acfi.12556

Conceptual Framework for Financial Reporting. (2018). IFRS Foundation. https://www.ifrs.org/issued-standards/list-of-standards/conceptual-framework/

Dai, J., & Vasarhelyi, M. A. (2017). Toward blockchain-based accounting and assurance. Journal of information systems, 31(3), 5–21. https://doi.org/10.2308/isys-51804

Fang, B., Liu, X., Ma, C., & Zhuo, Y. (2023). Blockchain technology adoption and accounting information quality. Accounting & Finance, 63(4), 4125–4156. https://doi.org/10.1111/acfi.13088

Freeman, R. E., Wicks, A. C., & Parmar, B. (2004). Stakeholder Theory and „The Corporate Objective Revisited". Organization Science, 15(3), 364–369.

Grigg, I. (2024). Triple Entry Accounting. Journal of Risk and Financial Management, 17(2), 1–12. https://doi.org/10.3390/jrfm17020076

Oladejo, M. T., Botes, V., Low, M., & Reeves, S. (2024). Blockchain technology disruptions: Exploring accounting and auditing academics and practitioners' perception. Accounting & Finance, acfi.13383. https://doi.org/10.1111/acfi.13383

Scott, W. R. (2015). Financial Accounting Theory (7th edition). Pearson.

Seshadrinathan, S., & Chandra, S. (2025). Trusting the trustless blockchain for its adoption in accounting: Theorizing the mediating role of technology-organization-environment framework. Financial

Innovation, 11(1), 1–39. https://doi.org/10.1186/s40854-024-00685-5

Thies, S., Kureljusic, M., Karger, E., & Krämer, T. (2023). Blockchain-based triple-entry accounting: A systematic literature review and future research agenda. Journal of Information Systems, 37(3), 101–118. https://doi.org/10.2308/ISYS-2022-029

Van Horne, J. C., & Wachowicz, J. M. (2008). Fundamentals of financial management. Pearson education.

Wagenhofer, A. (2015). Agency theory: Usefulness and implications for financial accounting. In: The Routledge Companion to Financial Accounting Theory. Routledge.

Watts, R. L., & Zimmerman, J. L. (1990). Positive accounting theory: A ten year perspective. The Accounting Review, 65(1), 131–156.

Wolfe, D. T., & Hermanson, D. R. (2004). The fraud diamond: Considering the four elements of fraud. Kennesaw State University. https://digitalcommons.kennesaw.edu/facpubs/1537/

Leveling Up Process Efficiency: AI as a Game-Changer for SMEs

Ivan Turčić, Ana-Marija Stjepić
Faculty of Economics and Business, University of Zagreb
Department of Entrepreneurship/ Department of Informatics
Trg J.F. Kennedy 6, HR-10000 Zagreb, Croatia
{iturcic, astjepic}@efzg.hr

Tomislav Jurić
Faculty of Economics and Business, University of Zagreb
Trg J.F. Kennedy 6, HR-10000 Zagreb, Croatia
tomislavjuric00Zg@gmail.com

Abstract. *This paper explores the artificial intelligence (AI) technology influence on business processes in small and medium-sized enterprises (SMEs) in the Republic of Croatia. Artificial intelligence tools that improve operational efficiency and competitiveness include big data analytics, pattern recognition, consumer behaviour prediction, and process automation. A survey was conducted among representatives of owners and directors of Croatian SMEs, and descriptive statistics were performed. SMEs in the Republic of Croatia face obstacles such as high upfront expenditures and security concerns, which hinder their implementation, despite the growing global adoption of AI. A Croatian study of 413 SMEs found that while AI's potential is widely recognized, its current application is still somewhat limited. According to the results, AI will gradually impact the operations of SMEs, improving their ability to compete in the market, be cost-effective, and be innovative.*

Keywords. artificial intelligence, business process, small and medium-sized enterprises

1 Introduction

Artificial intelligence (AI) is one of the main things that is driving the digital transformation of businesses today. Also, AI is becoming more widely available to small and medium-sized enterprises (SMEs) now that it has a range of tools for making operations more efficient, flexible, and responsive to the market. Before, AI was only available to large enterprises with a lot of technology infrastructure. Enterprises can use AI tools like big data analytics, pattern recognition, predicting consumer behavior, and automating processes to improve their operations, cut costs, and make decisions based on data. Since SMEs frequently function with limited financial, human, and technological resources, these capabilities are especially pertinent to them.

AI is not only a tool for improving processes, but it is also a strategic way for SMEs to stay competitive. AI can help enterprises create and deliver new and distinctive outcomes that meet the needs of specific customers in much less time. By adding AI to their operations, SMEs can react to the market and make their production systems more flexible. Therefore, in a business world that is getting more dynamic and unpredictable, digitalisation and automation help SMEs become stronger and more flexible.

According to empirical research, some fields, like transportation, logistics, automotive, and information technology, are adopting AI faster than others, like the chemical industry (Küpper et al., 2018). These disparities reflect differing expectations, sector-specific constraints, and varying levels of digital readiness. Furthermore, skepticism among some managers regarding the tangible value of AI contributes to cautious investment behavior, even when potential efficiency gains are well-documented.

It is also anticipated that AI will increasingly influence core business processes, from procurement and supply chain management to production, marketing, finance, and customer support. Notably, AI does not aim to replace existing business methodologies, but rather to augment them, primarily by automating routine jobs and helping employees to concentrate on higher-value, creative, and strategic roles (Küpper et al., 2018).

The adoption of AI will also entail structural shifts in the labor force. While some job functions may become obsolete due to automation, new roles will emerge in areas requiring complementary human skills and advanced digital literacy (Küpper et al., 2016). This transformation highlights the demand for constant employee education and training, particularly in SMEs seeking to build long-term competitive advantages.

This study examines the AI role in transforming business processes within Croatian SMEs, focusing on perceived benefits, adoption challenges, and the broader implications for organizational efficiency and competitiveness.

The following section outlines the theoretical framework that informs this analysis, including definitions of AI, classification of SMEs, and adaptation of AI technology in business processes. The third section presents the methodological framework used in this study to assess the adoption patterns and perceived impacts of AI among Croatian SMEs. The forth chapter presents an analysis and the fifth discussion of the results. The last chapter is a conclusion with stated limitations and recommendations.

2 Theoretical background

Artificial intelligence represents "a branch of computer science concerned with developing the ability of computers to perform tasks that require certain forms of intelligence. This may include navigating specific situations, learning to handle new circumstances, acquiring new concepts, analyzing data and drawing conclusions based on that analysis, understanding language, and many other activities" (Prister, 2019, p. 69).

Also, AI contains a wide range of computational techniques that allow machines to complete tasks typically demanding human intelligence, like learning, reasoning, problem-solving, and decision-making (Russell & Norvig, 2020). In a business context, AI technologies—including machine learning, natural language processing, computer vision, and robotics—facilitate automation, predictive analytics, and real-time data interpretation, thereby contributing to enhanced decision-making and operational efficiency.

Small and medium-sized enterprises are a vital component of national and regional economies, particularly within the European Union, where they represent around 99% of all enterprises and generate more than half of total value added (Di Bella et al., 2023). The number of employees (fewer than 250) and the annual turnover (not exceeding €50 million) are two common ways to define SMEs. However, definitions may differ by country (Official Journal of the European Union, 2003).

SMEs are crucial for economic growth, but they often face structural challenges, including insufficient funding, limited research and development capabilities, and slow growth. These things often make it difficult for them to invest in and use new technologies, like AI. Additionally, AI systems are highly complex, and SMEs often lack the in-house expertise to capitalize on digital opportunities.

Structured and related tasks that work together to make a product or service that meets specific customer needs are called business processes (Dumas et al., 2018). Adding AI to these processes can enable organisations to work more effectively by making them faster, more accurate, and more flexible (Davenport & Ronanki, 2018). Predictive maintenance in production systems, intelligent supply chain management, dynamic pricing models, fraud detection, and AI-enhanced customer service platforms are among the most common applications of AI in business process optimization (McKinsey & Company, 2021).

Artificial intelligence is a big step forward in how enterprises make money by automating tasks. AI-powered solutions are becoming increasingly important for "smart automation," which helps enterprises automate monotonous tasks, make workflows better, personalise customer experiences, and guess what will happen in the future (Davenport & Ronanki, 2018). This change affects more than just operations. It also affects strategic planning and the ability to come up with new ideas at every level of the value chain.

In manufacturing, AI has the potential to revolutionise production processes through task automation, predictive analytics, and the development of adaptive systems (Arinez et al., 2020). Artificial intelligence assists SMEs with intelligent manufacturing, predictive maintenance, customised production, and supply chain optimisation. These factors enhance the flexibility and efficiency of enterprises (Li et al., 2017; Buchmeister, 2019). Moreover, AI is helpful in some ways, but SMEs have problems implementing it since it demands many data, is hard to train models, is risky for cybersecurity, and uses more energy (Plathottam et al., 2023; Perrault et al., 2019). Nevertheless, SMEs can use AI to enhance productivity, product quality, and competitiveness in the market by carefully dealing with these problems and encouraging people and AI to work together.

Due to its ability to facilitate data-driven decision-making, real-time customer interaction, and process automation, AI in marketing is essential to increasing the competitiveness of SMEs. Machine learning, natural language processing, and predictive analytics are examples of AI technologies that help SMEs with performance optimisation, content personalisation, and market segmentation (Abrokwah-Larbi & Awuku-Larbi, 2023; Eriksson et al., 2020). Enterprises may engage with customers, cut costs on assistance, and provide service 24/7 with chatbots and other tools (Johannsen et al., 2018; Jin et al., 2019). Artificial intelligence also helps marketers by giving them important information about how customers act, which lets them set prices that change, automate content, and target customers more precisely (Enshassi et al., 2024). When SMEs start embracing AI, they need to modify the way they do things and the way their workers conduct their jobs. Though AI will make them more efficient, satisfied customers, and offer them an edge over their competitors in the long run (Verma et al., 2021; Enholm et al., 2021).

Artificial intelligence in sales helps SMEs by taking care of dull administrative work, which frees up sales staff to spend more time selling and getting things done faster (Chatham, 2023; HubSpot, n.d.). Also, AI looks at many data to help with lead scoring, dividing customers into groups, and coming up with more

innovative sales techniques. This makes people work harder and sell more (Chatham, 2023; Salesforce, 2024). Artificial intelligence's predictive analytics and automation tools give SMEs an edge over their competitors by allowing them to connect with customers in a more personal way and make strategic changes at the correct time (McKinsey, n.d.a).

AI is becoming more and more important to SMEs as a way to assist them in making better decisions by discovering patterns and predicting trends. This gives them an advantage over their rivals (Pozzo et al., 2024). Enterprises can make decisions faster and more accurately with the help of AI that also does things on its own, makes greater use of resources, and decreases the chance of human error. This helps enterprises generate new ideas and establish plans for the future (Filippi et al., 2024). A key benefit of AI is lowering prediction costs, much like how semiconductor technology reduced computational costs and transformed industries; this has expanded AI's application in inventory management, finance, and marketing, leading to better data-driven decisions (McKinsey, n.d.b). Despite AI's predictive superiority, human judgment remains essential for final decision-making (McKinsey, n.d.b).

3 Methodology

This research examines the opinions of SMEs owners and employees in Croatia about how AI affects their jobs. The study looked at how AI affected the business processes of a number of enterprises in a variety of fields, the challenges that came up when it was put into use, and the benefits that were expected to come from AI in the future.

The main goal of this study was to find out how AI affects the workflows of SMEs by asking about their processes, figuring out the best ways to use AI in these workflows, and figuring out how to put it into practice. The questionnaire was adapted from two prior studies conducted among German SMEs: one involving 98 enterprises (Szedlak et al., 2020) and another surveying 283 firms (Urlic & Frank, 2021).

Also, the survey was designed and conducted from August to October 2024 using a Google Forms questionnaire. The survey targeted SMEs across various industries and geographic locations within the Republic of Croatia. The anonymous questionnaire was distributed via email to enterprise addresses obtained from the FINA database. Out of 4,400 emails sent, 435 responses were collected, resulting in a response rate of 9.89%.

The survey consisted of 13 primarily closed-ended questions. Participants assessed the degree of AI's influence on different business processes on a 5-point Likert scale, ranging from 1 (no impact) to 5 (extremely high impact).

The first three questions aimed to gather essential enterprise information, including industry sector, enterprise size, and geographic location for classification purposes. Subsequent questions focused on the adoption and duration of AI technology use, as well as the primary applications of AI in daily business operations.

The survey then explored the extent of AI's impact on business activities, the importance attributed to AI integration in specific business processes, and perceived opportunities stemming from AI adoption. Additionally, respondents were asked about challenges or barriers encountered or anticipated during AI implementation. Finally, the survey assessed the perceived significance of investing in AI for the enterprise's future strategic development.

4 Analysis of results

A total of 435 responses were collected, of which 22 were from large enterprises; these were excluded from the analysis. The final sample consisted of responses from 413 enterprises.

The first survey question addressed the industries in which the participating enterprises operate. The largest share, 22.8% (99 enterprises), was in wholesale and retail trade, followed by 15.2% (66 enterprises) in construction and 14.3% (62 enterprises) in other service activities. Additionally, 9.2% (40 enterprises) operated in telecommunications, computer programming, and other IT services, while 8.3% (36 enterprises) were in manufacturing. Other industries accounted for 5% or less of the sample. These results indicate that the survey included enterprises from diverse industries, not solely SMEs focused on technology or closely related to AI.

The second question focused on enterprise size classification. Among respondents, 43.2% (188) identified as small enterprises, 30.3% (132) as medium, and 21.4% (93) as micro enterprises. Large enterprises accounted for 5.1% (22) and were excluded from further analysis due to the study's focus.

The third question addressed the geographical location of the enterprises' headquarters within Croatia. The majority, 57% (237 enterprises), are based in Central Croatia, followed by 18% (73) in Dalmatia, 13% (53) in Slavonia, and 12% (50) in Istria and Kvarner. The data indicates that most surveyed enterprises are situated in Central Croatia, with 165 of them headquartered in Zagreb.

The following question explored the duration of AI usage within the surveyed enterprises. A majority of 56.3% enterprises (240) reported not using any form of AI in their operations. Meanwhile, 16.3% (67) have adopted AI within the past year, 19.5% (81) between one and three years ago, 3% (9) for three to five years, and 4.8% (17) for over five years. This indicates that 43.7% of the respondents currently use AI technology, which is a significant proportion for SMEs in Croatia.

A related question addressed the frequency of AI use in daily business activities. Of the enterprises

surveyed, 44.8% (191) do not utilize AI regularly, 17% (70) use it rarely, 16.8% (69) occasionally, 12% (48) frequently, and 9.4% (37) very frequently.

The sixth question examined how enterprises incorporate AI technologies—whether as part of their product offerings, embedded in daily processes, or as services provided. Among 413 respondents, 6.1% enterprises (25) indicated that AI is used within their products, 33.9% (140) employ AI as part of their processes, and 22% (91) integrate AI within their service offerings, while 53.1% (231) do not use AI technologies.

The next question examined respondents' perceptions of AI's impact across various business functions. AI was seen to have any level of influence on finance (52%), human resources (51%), and accounting (55%), while 45 to 49 % of enterprises believe that AI does not affect the listed business functions. The highest percentage of any impact was reported in research and development and IT, both at 67% while 34 % of enterprises believe that there is no impact of AI in the aforementioned business areas. Sales (60%), logistics (55%), and production (47%) were also noted as influenced by any level of AI (low, moderate, high or extremely high). Overall, production, accounting, finance, and HR were rated as having low impact (means from 2 to 2,04), sales, logistics, and customer service as moderate (means from 2,14 to 2,34), while marketing, IT, and R&D had high but not extremely high influence (means from 2,71 to 2,82).

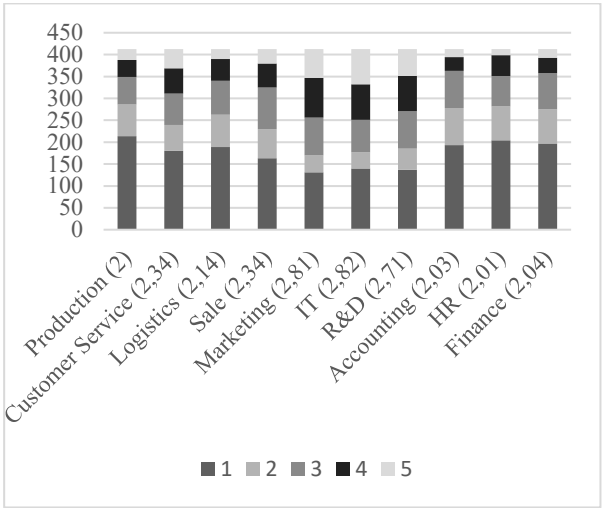


Figure 1. AI's influence on different business areas (mean in bracket)
Source: authors' work

In Figure 2, the results indicated that autonomous vehicles, resource management, and robotics were deemed the least important processes influenced by AI (means 2.03 and 2.27). Quality control, process management, predictive maintenance and logistics processes, were moderately important processes (means from 2.29 to 2.32), while chatbots, assistance (aid) systems, automation, and language assistants were considered the most critical processes being influenced by AI (means from 2.47 to 2.91).

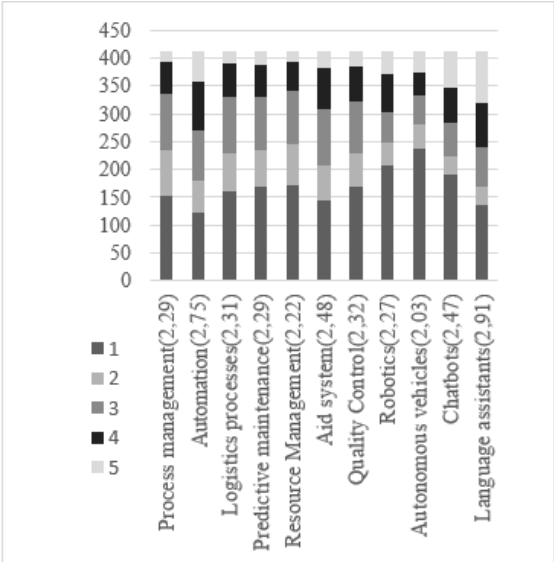


Figure 2. The importance of certain processes being influenced by AI (mean in bracket)
Source: authors' work

Enterprises identified key opportunities from AI implementation as process acceleration (61.4%), improved data utilization (56.6%), process automation (46.7%), and cost savings (46%). Enhanced decision-making (37.7%) and new business model development (32.6%) were also highlighted.

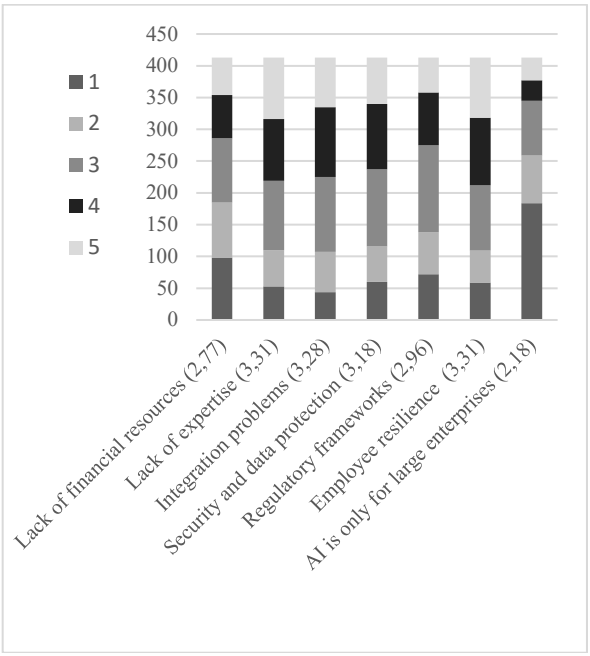


Figure 3. Challenges/obstacles to implementing AI (mean in bracket)
Source: authors' work

Regarding challenges, employee resistance and lack of expertise were identified as primary barriers (both mean 3.31). Integration issues (mean 3.28) and

data security (mean 3.18) were moderate concerns, while regulatory frameworks (mean 2.96), financial constraints (mean 2.77), and perceptions of AI being suitable only for large enterprises (mean 2.18) were less significant obstacles. Open-ended responses pointed to data inaccuracy and misinformation as additional implementation challenges.

When asked about AI's future impact, 91.3% of SMEs anticipated some level of influence, ranging from moderate (35.6%), minor (26.4%), high (19.1%) to extremely high (10.1%). Finally, 90.6% recognised the importance of investing in AI as part of their future development strategy, though opinions varied on the level of importance (23.7% low, moderate 34.9%, high 17.2% and 9.4% extremely high).

Table 1. The key findings of the paper

Main industries of the surveyed enterprises	Wholesale and retail trade (22.8%), Construction (15.2%), IT (9.2%)
Classifications of enterprises	43.2% Small, 30.2% Medium and 21.4% Micro
Geographical area of the surveyed enterprises	57% Central Croatia
Duration of AI usage by surveyed enterprises	56.3% do not use, 19.5% use AI in the last 1-3 years
How SMEs are using AI	As part of the process (32.2%)
The most influenced business areas by AI technologies in the surveyed enterprises	IT & Marketing
The most important processes to be influenced by AI	Language assistants
The main opportunities for the implementation of AI in the business of the surveyed enterprises	Speed up the process (61.4%)
The main challenges that the surveyed enterprises face for the implementation of AI	Lack of expertise and the resilience of employees to change
How many enterprises think AI will have an impact on their business in the future	35.6% - medium impact, 26.4% - low impact, 19.1% - high impact
The importance of the surveyed enterprises on investing in AI as a strategy for future development	34.9% - medium-important, 23.7% - minor, 17.2% - very important

Source: authors' work

The key findings are summarized in table 1.

5 Discussion

The survey collected responses from enterprises across various industries, encompassing a total of 413 participants. Among these, 22.8% were from the wholesale and retail trade sector, 15.2% from construction, 14.3% from other service activities, and 9.2% from the telecommunications and IT sector. This distribution indicates a broad representation of SMEs beyond those directly engaged in technology sectors.

Comparatively, a similar study conducted in Germany surveyed 98 SMEs, with 27% from service industries, 17% from IT, and 15% from manufacturing (Szedlak et al., 2020). A different German research of 283 SMEs indicated that 38% of them were in the services sector, 11% were in manufacturing, and 41% were in information (Urlich & Frank, 2021). These studies help us learn more about how SMEs in different nations use AI and what it does.

When it comes to using AI, 32.2% of Croatian SMEs stated they used it in their processes, 21.1% in their services, and 5.7% in their goods. The German research, on the other hand, indicated higher adoption rates, with 51% using AI in processes, 16% in services, and 6% in goods (Szedlak et al., 2020). The differences between the two countries could be because they are at different levels of AI knowledge and use.

The survey also looked at how respondents thought AI might affect different areas of business. Fifty-two percent of Croatians SMEs were asked said that AI had an effect on finance, 51% said it had an effect on human resources, 55% said it had an effect on accounting, 67% said it had an effect on R&D, 67% said it had an effect on IT, 60% said it had an effect on sales, 55% said it had an effect on logistics, and 47% said it had an effect on production. These results suggest that respondents are aware of the effects of AI on a range of moderate to high-level business tasks.

In contrast, the German studies revealed that AI had a higher effect on finance (88%), human resources (78%), accounting (80%), research and development (75%), IT (90%), sales (85%), and production (80%). These differences could suggest that the two countries are at different stages of being ready for AI and being able to employ it.

Seventy percent of SMEs in Croatia considered that automation processes must be under the influence of AI. This is the same as 80% of SMEs in Germany who said the same thing (Szedlak et al., 2020). Respondents in Croatia believe that the assistance (aid) system it is important to be influenced by AI for 65% of Croatian SMEs 65%, while it was important to 95% of German SMEs. These were the major discrepancies in the compared results.

When asked what the benefits of using AI were, 61.4% of SMEs in Croatia said it would speed up processes, 56.6% said it would make data use more efficient, 46.7% said it would automate processes, 46% said it would save money, 37.7% said it would help them make better decisions, 32.6% said it would help

them come up with new business models, and 22.1% said it would help them control risks better. The German study found that AI could sped up process 66%, data was used more efficiently 72%, processes were automated 77%, possible savings 55%, decisions were made better 53%, new business models were created 43%, and risks were controlled 31% (Urlich & Frank, 2021).

The study also found that 86% of Croatian SMEs were worried about their employees not wanting to use AI and not knowing enough about it. The German studies also showed that a lack of competencies and expertise is the main barrier to implementing AI for 65% enterprises (Urlich & Frank, 2021). These problems keep coming up, which shows how important it is to change the enterprise's culture and teach people new skills so they can use AI more easily.

Finally, Croatian SMEs know that AI has a lot of potential, but they don't seem to be using it or incorporating it as much as their German counterparts. This means that enterprises need to do more to teach their employees how to use AI, raise awareness of it, and build the infrastructure that will help them get the most out of its benefits.

6 Conclusion

AI technologies like machine learning and neural networks have made it possible for enterprises to look at large datasets, detect patterns, and make decisions based on data. With these technologies, enterprises can save money, work more efficiently, and improve their overall performance. Increasingly, departments in both large and small enterprises are utilizing AI. Production, logistics, marketing, human resources, and strategic management are some of these divisions. An enterprise that uses AI has an advantage over its competitors in today's fast-paced world because with AI, an enterprise can monitor market trends, predict what customers will want, and improve business strategies.

The results show that Croatian SMEs know how AI can help them, mostly by speeding up work, automating tasks, and keeping track of resources. A lot of people know about AI solutions, but they haven't started using them yet. Respondents think that the biggest problems are high start-up costs, not having enough skilled workers, and worries about keeping data safe and managing risk.

Yet enterprises that use AI early on are likely to get ahead of their competitors and do better in the market. The survey found that 91% of respondents think AI will have an effect on their business in the future, and 10% think it will have an extremely high effect. Furthermore, 90% consider investment in AI technologies important for their future development strategies, and 9.1% view it as critically important for long-term growth.

These findings highlight that Croatian SMEs are increasingly open to exploring AI, especially when its value is clearly linked to enterprise outcomes. Enterprises willing to adapt and strategically implement AI stand to not only improve performance but also contribute positively to the broader economy.

However, this study surveyed 413 SMEs in Croatia, representing only about 0.2% of the total SME population, which limits the generalizability of the findings. Data were collected solely through self-reported online questionnaires, introducing potential subjectivity and response bias. Additionally, the research was conducted within a specific timeframe, which may affect the applicability of results under changing economic conditions. Future studies should consider larger, more diverse samples and incorporate qualitative methods, such as in-depth interviews, to provide deeper insights into AI's impact on SMEs.

References

Abrokwah-Larbi, K., & Awuku-Larbi, Y. (2024). The impact of artificial intelligence in marketing on the performance of business organizations: Evidence from SMEs in an emerging economy. *Journal of Entrepreneurship in Emerging Economies, 16*(4), 1090–1117. https://doi.org/10.1108/JEEE-07-2022-0207

Arinez, J., Chang, Q., Gao, R. X., Xu, C., & Zhang, Y. (2020). Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook. *Journal of Manufacturing Science and Engineering*, 142(11), 110804. doi:10.1115/1.4047855

Buchmeister, B., Palcic, I. & Ojstersek, R. (2019). Artificial intelligence in manufacturing companies and broader: An overview. In B. Katalinić (Ed.), *DAAAM International Scientific Book*, (pp. 81–98). DAAAM International Vienna, Vienna doi: 10.2507/daaam.scibook.2019.07

Chatham, A., B. (2023). *7 ways small businesses are using AI for sales*. Retrieved from https://www.wix.com/blog/ai-for-sales

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review, 96*(1), 108–116.

Di Bella, L., Katsinis, A., Lagüera-González, J., Odenthal, L., Hell, M., & Lozar, B. (2023). *Annual Report on European SMEs 2022/2023* (JRC134336). Publications Office of the European Union. https://doi.org/10.2760/028705

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). *Fundamentals of Business Process Management* (2nd ed.). Springer. doi:10.1007/978-3-662-56509-4

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Artificial intelligence and business value: A literature review. *Information Systems Frontiers, 24*(6), 1709–1734. https://doi.org/10.1007/s10796-021-10186-w

Enshassi, M., Al-mulali, U., Nathan, R., J. i Soekmawati, S. (2024.), Potentials of artificial intelligence in digital marketing and financial technology for small and medium enterprises, IAES International Journal of Artificial Intelligence (IJ-AI) 13(1), 639-647, http://doi.org/10.11591/ijai.v13.i1.pp639-647

Eriksson, T., Bigi, A., & Bonera, M. (2020). Think with me, or think for me? On the future role of artificial intelligence in marketing strategy formulation. *The TQM Journal, 32*(4), 795–814. https://doi.org/10.1108/TQM-12-2019-0303

Filippi, E., Bannò, M., & Nencini, I. E. (2024). Artificial intelligence in business and decision making: Analysis of benefits and challenges for consultancy SMEs. *Piccola Impresa / Small Business*, (2), 88-107. https://doi.org/10.14596/pisb.4446

HubSpot. (n.d.). *Discover the Future of Sales with AI*. Retrieved from https://offers.hubspot.com/ai-sales

Jin, B. E., Cedrola, E., & Kim, N. (2019). Process innovation: Hidden secret to success and efficiency. In B. E. Jin & E. Cedrola (Eds.), *Process innovation in the global fashion industry* (pp. 1–15). Palgrave Pivot. https://doi.org/10.1057/978-1-137-52352-5_1

Johannsen, F., Leist, S., Konadl, D., & Basche, M. (2018). Comparison of Commercial Chatbot solutions for Supporting Customer Interaction. *Research Papers,* 158. https://aisel.aisnet.org/ecis2018_rp/158

Küpper, D., Kuhlmann, K., Köcher, S., Dauner, T. & Burggraaff, P. (2016). *The Factory of the Future*, BCG. Retrieved from https://www.bcg.com/publications/2016/leaning-manufacturing-operations-factory-of-future

Küpper, D., Lorenz, M., Kuhlmann, K., Bouffault, O., Wyck, J., Köcher, S. & Schlageter, J. (2018). *AI in the Factory of the Future*, BCG. Retrieved from https://www.bcg.com/publications/2018/artificial-intelligence-factory-future

Li, B., Hou, B., Yu, W., Lu, X., & Yang, C. (2017). Applications of Artificial Intelligence in Intelligent Manufacturing: A Review. *Frontiers of Information Technology & Electronic Engineering*, 18(1), 86–96. doi:10.1631/FITEE.1601885

McKinsey & Company. (2021). *The state of AI in 2021*. Retrieved from https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/global-survey-the-state-of-ai-in-2021

McKinsey. (n. d.a), AI-powered marketing and sales reach new heights with generative AI. Retrieved from https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/ai-powered-marketing-and-sales-reach-new-heights-with-generative-ai

McKinsey & Company (n. d.b), The economics of artificial intelligence, Retrieved from https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-economics-of-artificial-intelligence

Official Journal of the European Union. (2003). *Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises*. Retrieved from https://publications.europa.eu/resource/cellar/1bd0c013-0ba3-4549-b879-0ed797389fa1.0019.02/DOC_2

Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Mishra, S., & Niebles, J. C. (2019). *The AI Index 2019 Annual Report*. AI Index Steering Committee, Human-Centered AI Institute, Stanford University. Retrieved from https://wp.oecd.ai/app/uploads/2020/07/ai_index_2019_introduction.pdf

Plathottam, S. J., Rzonca, A., Lakhnori, R., & Iloeje, C. O. (2023). A review of artificial intelligence applications in manufacturing operations. *Journal of Advanced Manufacturing and Processing, 5*(3). https://doi.org/10.1002/amp2.10159

Pozzo, D. N., Gonzalez Beleño, C. A., Reales Correa, K., Garizabal Donado, M., Gomez Pedroza, F. J., & Moncada Diaz, J. E. (2024). Managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making: A study with Colombian SMEs. *Procedia Computer Science, 234*. https://doi.org/10.1016/j.procs.2024.06.119

Prister, V. (2019). Umjetna inteligencija. *Media, Culture and Public Relations*, *10*(1), 67–72. https://doi.org/10.32914/mcpr.10.1.7

Russell, S. J., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Hoboken, NJ: Pearson Education.

Salesforce. (2024). Top Generatice AI Statistics for 2024. Retrieved from https://www.salesforce.com/news/stories/generative-ai-statistics

Szedlak, C., Poetters, P., & Leyendecker, B. (2020). Application of artificial intelligence in small and medium-sized enterprises. In *Proceedings of the*

5th North American International Conference on Industrial Engineering and Operations Management (pp. 1574–1582). IEOM Society International.

Ulrich, P. S., & Frank, V. (2021). Relevance and adoption of AI technologies in German SMEs: Results from survey-based research. *Procedia Computer Science, 192*, 2152–2159.

Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. *International Journal of Information Management Data Insights, 1*(1), Article 100002. https://doi.org/10.1016/j.jjimei.2020.100002

Intelligent Systems

Tea Krčmar, Dina Šabanović, Miljenko Švarcmajer, Mirko Köhler, Ivica Lukić, Marko Carević
Beyond Traditional BI: Large Language Models as Game-Changers in Document Processing

Karlo Babić, Ana Meštrović
Pyramidal Recursive Composition of Multi-Word Units into Unified Representations

Petra Kuljiš, Lucia Načinović Prskalo, Marija Brkić Bakarić
Detecting Discrimination Using Selected Data Mining Methods

Saša Mitrović, Neven Vrček
Methodology for the comparative analysis of PCA and Autoencoders in dimensionality reduction: impact on classification accuracy and computational efficiency

Oliver Lohaj, Anastasiia Radishevska
From Black Box to User Insight: Explainability and Usability in Toxicity Detection

Ivana Kovačević, Ivan Aleksi, Tomislav Matić, Josip Zidar
Smart Cuvette: An Integrated Optical System for Ethanol Measurement in Liquid Samples

Bojan Radišić, Ivan Dunđer, Sanja Seljan
Impact of Imputation Methods on the Performance of Classification Algorithms

Dominik Čambál, Martin Pukanec, Monika Herchlová, Juraj Ďuďák, Matúš Nečas
Performance study of a bidirectional vanilla spiking neural network for signal noise suppression

Luka Lukač, Damjan Strnad, Štefan Horvat
Compressibility of WGANs in a Distributed Learning Environment

Marko Vještica, Elena Akik, Vladimir Dimitrieski, Lukas Hinterleitner, Jovan Erić, Frank-Christian Weidenfelder, Jonas Wieczorek
Towards User-Centric Chatbot for Q&A in Industry: Evaluation of Employees' Experience

Goran Bubaš
How ChatGPT and Gemini View the Elements of Communication Competence of Large Language Models: A Pilot Study

Dalibor Bužić, Jasminka Dobša
Analysis of Popular Croatian Song Lyrics

Beyond Traditional Business Intelligence: Large Language Models as Game-Changers in Document Processing

Tea Krčmar, Dina Šabanović, Miljenko Švarcmajer, Mirko Köhler, Ivica Lukić
J. J. Strossmayer University of Osijek, FERIT Osijek, Croatia
Cara Hadrijana 10b, Osijek
{tea.krcmar, dina.sabanovic, miljenko.svarcmajer, mirko.kohler, ivica.lukic}@ferit.hr

Marko Carević
Uvid d.o.o., Zagreb, Croatia
Hrelička 96, Zagreb
marko@uvid.hr

Abstract. *Unstructured data challenges traditional business intelligence (BI), particularly when processing layout-rich documents like scanned images or PDFs. This paper explores using large language models (LLMs) for extracting structured information, focusing on the LMDX (Language Model-based Document Information Extraction and Localization) approach. LMDX encodes document layout by inserting coordinate tokens directly into the prompt, eliminating the need for vision-based models or architectural changes. We also propose a blockchain-enabled, cloud-based pipeline that integrates LMDX for secure and scalable document intelligence. The paper presents key findings and outlines future research to enhance document processing and automated decision-making in modern BI systems.*

Keywords. Business Intelligence, Document Processing, Information Extraction, Layout Analysis, Machine Learning

1 Introduction

Today, most of the data generated in the digital world is classified as unstructured, characterized by the absence of a predefined model or organized structure, which prevents it from being stored in traditional relational databases. Unstructured data, often referred to as qualitative data, is recognized for its rich information, but is also criticized for its complexity in analysis due to its ambiguous and unstandardized nature (Oza et al., 2023). The exponential growth of unstructured data, mostly scanned images, handwritten notes, PDFs, emails, and text documents, requires intelligent document processing solutions (Mahadevkar et al., 2024; Singh & Hooda, 2023). Although structured data is well managed, difficulties are encountered in processing, retrieving, and understanding information from a variety of document formats by traditional business intelligence (BI) systems (Majid et al., 2024). Through the application of machine learning and natural language processing techniques, unstructured data can be analyzed to extract valuable information (Mahadevkar et al., 2024), which is then used to inform business decisions, enhance customer understanding, and support strategic planning for future needs (Sever, 2024).

Recent advances in deep learning, especially large-scale language models such as one proposed by (Perot et al., 2024) and pre-trained neural networks (Sage et al., 2019, 2021; X. Yang et al., 2017), have redefined how documents are analyzed and structured (Mahadevkar et al., 2024). Tasks such as text recognition, entity extraction, summarization, and multi-modal data processing are excelled in by these models (He et al., 2024; Yue et al., 2024), making them essential to automate processes that were previously required to be handled by considerable human effort. Unlike rule-based or template-based approaches, machine learning (ML) document processing can adapt to different document types, context can be inferred, and accuracy can be increased over time (Baviskar et al., 2021).

The potential of machine learning to transform document processing in business intelligence is explored in this paper. First, a comprehensive analysis of the impact of ML approaches and their use in BI is provided. Secondly, the problem and impact of unstructured data in business intelligence are examined. In the following chapter, LLMs for information extraction from layout-rich documents are explored with emphasis on LMDX by (Perot et al., 2024). Next, we propose a way to integrate LMDX into a cloud-based business intelligence system. Finally, key findings are summarized and future research directions are proposed to address the ongoing challenges of analyzing unstructured documents and their role in business intelligence.

2 Advancing Business Intelligence with Machine Learning

Machine learning has been recognized as a game-changer in business intelligence (Bharadiya, 2023; Sage et al., 2019), with its ability to enable organizations to move beyond traditional, static reporting to dynamic, data-driven decision-making (Hindle et al., 2019). Unlike rule-based systems, which are designed to rely on predetermined logic, ML models are trained on patterns in data, allowing hidden insights, anomalies, and accurate predictions to be uncovered (Bloch & Sacks, 2018; Taddy, 2018). Tasks involving high-dimensional data, where accuracy and efficiency are difficult to maintain with traditional methods, are particularly well-suited to these models (Taddy, 2018). Decision-making processes across all sectors are improved, operations are optimized, customer experience is enhanced, strategic growth is driven, and businesses are given a competitive advantage (Attaran & Deb, 2018; Prasanth et al., 2023). Additionally, new market opportunities and potential supply chain risks can be proactively identified before companies fall victim to predictive analytics, which is regarded as an essential part of machine learning in business intelligence. Rather than responding to problems as they arise, a forward-looking strategy is enabled, allowing proactive action to be taken (Pasupuleti et al., 2024).

An example of ML's capabilities is seen in demand forecasting, a crucial yet intricate aspect of supply chain management, where precise predictions of future demand must be made while accounting for numerous dynamic variables (Abolghasemi et al., 2019; Pasupuleti et al., 2024). Predefined statistical approaches are traditionally used in forecasting models, yet they struggle to capture nuanced market fluctuations, leading to inefficiencies such as overstocking or stockouts (J. Wang et al., 2018). In contrast, machine learning techniques, including ensemble methods and deep learning, are applied to vast and diverse datasets, ranging from historical sales records to external factors like economic indicators, seasonal trends, and consumer sentiment analysis (Pasupuleti et al., 2024). Predictive models are continuously refined with real-time data, significantly improving forecast accuracy and allowing inventory levels to be optimized, waste to be reduced, and resource allocation to be enhanced (Morariu et al., 2020).

Beyond forecasting, logistical operations such as vehicle routing, transportation planning, and inventory optimization are significantly refined through ML applications (Alam et al., 2020; G.-B. Huang et al., 2006). Real-time data sources, including traffic conditions, weather patterns, and customer order updates, are integrated into ML algorithms, allowing delivery schedules to be dynamically adjusted and routes to be optimized, reducing transportation costs and increasing efficiency (Jayaprakash et al., 2021). Through this adaptability, customer demands can be met more effectively, ensuring improved service reliability and satisfaction (Pasupuleti et al., 2024). Additionally, strategic decision-making is enhanced as market trends, shifts in consumer behavior, and supply chain vulnerabilities are uncovered. With these predictive insights, operational adjustments can be made proactively, allowing risks to be mitigated and emerging opportunities to be capitalized upon (M. Yang et al., 2019). Across industries, ML's transformative impact has been demonstrated through case studies, in which AI-driven analytics have been leveraged to streamline supply chain processes, minimize disruptions, and maintain a competitive edge (Pasupuleti et al., 2024).

3 The Issue of Unstructured Data in Business Intelligence

The main problem with unstructured data in BI lies in its inherent complexity (Attaran & Deb, 2018; Majid et al., 2024). Numerous information sources are unstructured or semi-structured, such as customer emails, web pages with competitor information, sales force reports, research paper repositories, and more (Baars & Kemper, 2008). These diverse types of data come in different formats—ranging from free-form text to images and audio(Baviskar et al., 2021)—which complicates the application of traditional BI tools, typically designed to handle structured, tabular data (Baars & Kemper, 2008). Additionally, extracting embedded information from visual data such as scanned documents, pictures and videos poses a challenge to current processing techniques (Adnan & Akbar, 2019). This lack of standardisation is exacerbated by the fact that BI systems rely on predefined models and structured data formats to produce meaningful insights (Sharma et al., 2021).

As the amount of unstructured data increases exponentially, traditional BI systems face considerable constraints in processing data and making strategic decisions. Traditional BI systems, which aggregate and process data in life cycles (Czvetko et al., 2022), fail to provide timely analysis for stratigic planning, and with it limit businesses' from fully capitalising on emerging trends in real-time. The sectors producing large volumes of non-structured content, such as legal contracts or medical records, are prime examples of data-driven decision-making problems (Kumar & Singh, 2019; S.-H. Park et al., 2021). This challenge also extends to businesses, which may miss out on valuable insights contained in customer feedback, market trends, social media content, feedback on news or operational reports. Lack of strategic planning is a significant factor contributing to the failure of businesses, especially those just starting (Chukwuelue, 2024).

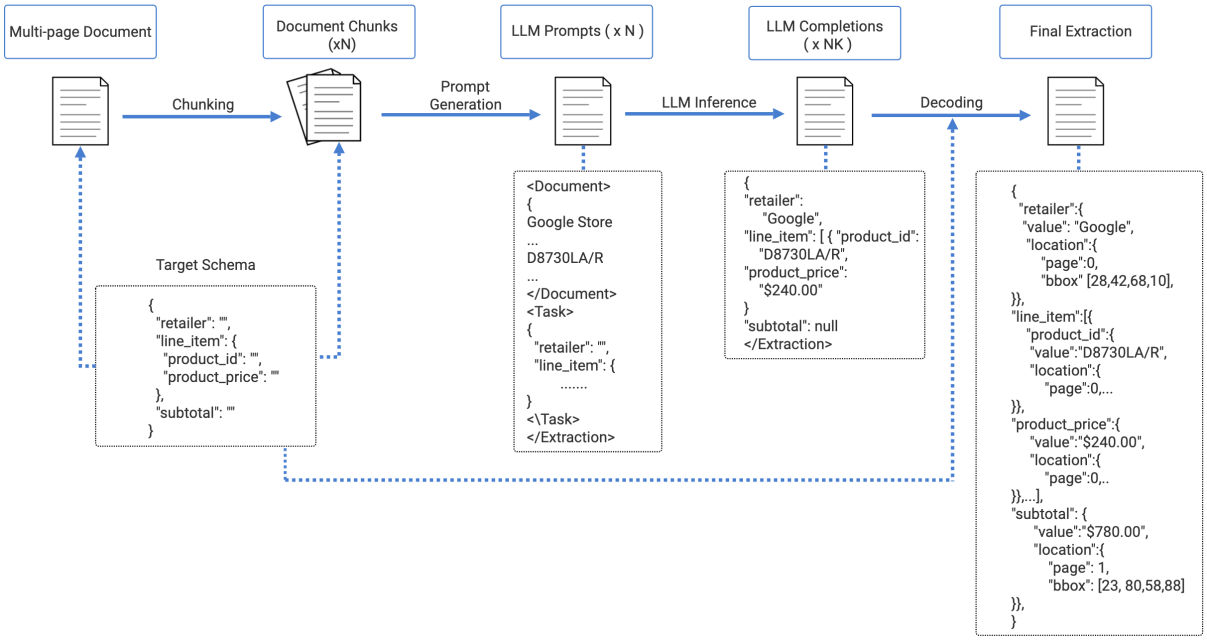


Figure 1. Methodology of LMDX

4 LLMs as Tools for Layout-Rich Document Data Extraction

Within organisations, the key information is often contained in layout-aware documents (LRDs), such as reports and presentations, which combine layout-aware visual elements (e.g., graphs, tables, charts) with text structure and content (Colakoglu et al., 2025; S. Park et al., 2019; Z. Wang et al., 2023; Zmigrod et al., 2024). LRDs are challenging traditional natural language processing (NLP) techniques designed for plain text (Cui et al., 2021; Tang et al., 2023) by incorporating visual and structural features to improve the extraction of information from LRDs (Colakoglu et al., 2025). However, these models require significant fine-tuning of the dataset to be accurate; users must manually annotate the training dataset with bounding boxes and extraction elements for each new document set (Colakoglu et al., 2025).

The emergence of large-scale language models, such as GPT-4 (Achiam et al., 2023), has significantly advanced the NLP domain due to their exceptional ability to understand and generate text (Liu et al., 2024; D. Xu et al., 2023). Retraining language teachers via automatic regression prediction allows them to capture natural patterns and semantic knowledge within the text corpus (Lyu et al., 2024; D. Xu et al., 2023). This enables LLMs to perform zero and few-shot learning, model various tasks consistently, and serve as tools for data augmentation. Additionally, LLMs can act as intelligent agents for complex planning and task execution, leveraging memory mining and various tools to increase efficiency and ensure successful task completion (D. Xu et al., 2023).

Since visually rich documents (VRDs) combine both textual and visual elements, with spatial positioning crucial for understanding, many approaches have explored custom architectures and pretraining strategies to model the relationship between text, layout, and image modalities. For instance, (Y. Xu et al., 2022) use a separate image encoder to add features to token encodings, while (Y. Huang et al., 2022) jointly model page image patches and tokens, using self-supervised pretraining tasks to learn modality connections (Perot et al., 2024). (Hong et al., 2022) propose encoding the relative 2D distances of text blocks in transformer attention and learning from unlabeled documents with an area-masking strategy. (Kim et al., 2022) and (Lee et al., 2023) abandon the text modality entirely, opting for a vision transformer encoder with an autoregressive decoder pretrained on pseudo-optical character recognition (OCR) and region masking tasks. LLMs for extraction have mostly been studied in the text domain (Keraghel et al., 2024), either generally (Laskar et al., 2023) or domain-specific (De Toni et al., 2022; Hu et al., 2023). (S. Wang et al., 2023) use an LLM to insert special tokens to mark the boundaries of target entities, building a layout-aware LLM with various document understanding capabilities (Perot et al., 2024).

Among various approaches and previous research, the LMDX (Language Model-based Document Information Extraction and Localization) framework by Perot et al. (2024) stands out. It reframes information extraction from visually rich documents as a sequence-to-JSON generation task, enabling large language models to perform layout-aware extraction using only text-based prompts. LMDX encodes document layout by appending normalized coordinate tokens to each line or word segment, allowing the LLM to infer

spatial relationships without needing vision encoders or architecture modifications. As shown in Fig.1, the pipeline begins by chunking documents: pages are processed individually and truncated line by line until each chunk satisfies the model's token limit. This approach minimizes entity fragmentation across pages. Each chunk is embedded in a structured prompt that includes the text, coordinate tokens, a task description, and a predefined extraction schema formatted in JSON. This schema supports singular, repeated, and hierarchical entities. During inference, multiple completions are sampled from the LLM for each chunk to introduce diversity. These completions are structured JSON outputs in which each extracted value is paired with its corresponding coordinate tokens. The decoding stage then verifies that the extracted text exists at the specified coordinates in the original document. If the referenced segment does not match exactly—either by content or location—the prediction is flagged as a hallucination and discarded. For valid outputs, bounding boxes are reconstructed by aggregating the positions of all verified segments. Finally, majority voting across completions is used to consolidate predictions, improving both consistency and resilience to spurious outputs. Zero-shot extraction allows processing of new document types without training, while fine-tuning improves accuracy with minimal data. Multiple responses are aggregated for reliability, making LMDX effective for automating document processing tasks. LMDX shows state-of-the-art results with space for improvement.

5 Integrating Data Extraction into Cloud

Integrating LMDX (Language Model-based Document Information Extraction and Localization) into a blockchain-enabled cloud-based business intelligence (BI) system, shown in Fig. 2, presents a novel approach to secure, decentralized, and intelligent document processing. LMDX leverages large language models (LLMs) to extract structured information from visually rich documents while ensuring entity localization. When combined with a blockchain infrastructure, this integration enhances data integrity, auditability, and automation in enterprise workflows.

The proposed architecture incorporates a cloud-hosted LMDX deployment that interacts with a blockchain-based storage and verification system. An automated system processes documents, with LMDX converting unstructured information into structured formats. To ensure tamper-resistant integrity and verifiable audit trails, the processed data is hashed and recorded on a blockchain ledger. The structured outputs are subsequently integrated into a BI framework for advanced analytics, visualization, and decision support. A decentralized storage layer maintains the immutability of raw document data while allowing for controlled access through smart contracts deployed on blockchain platforms such as Ethereum or Hyperledger Fabric.

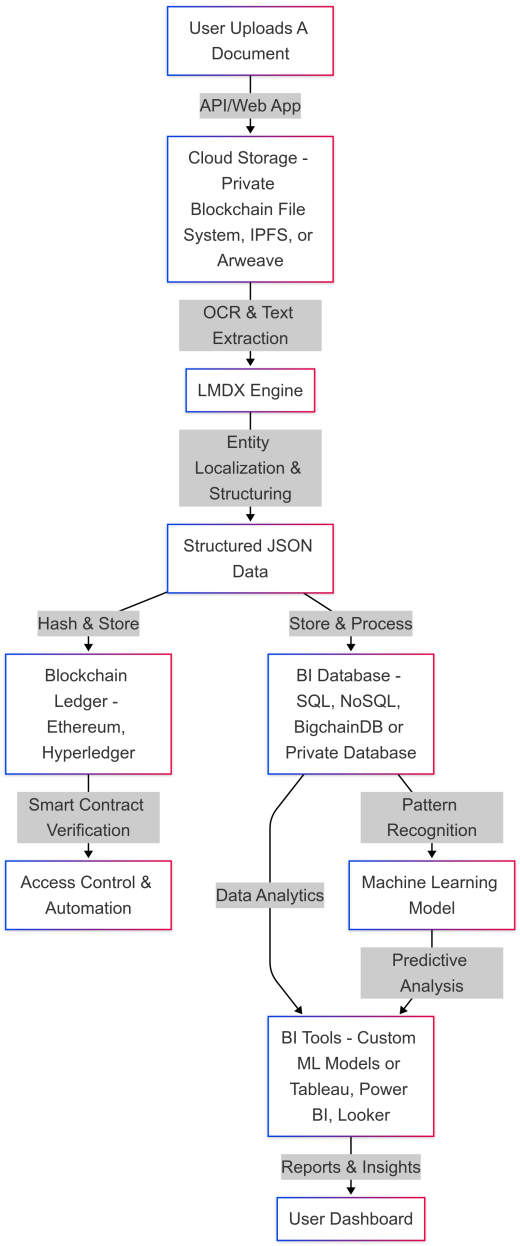


Figure 2. Implementation of LMDX into a business system in the cloud

This integration enables a secure, scalable, and automated workflow f or d ocument i ntelligence, particularly in applications such as financial a uditing, regulatory compliance, and enterprise contract analysis. By utilizing machine learning for data extraction and blockchain for authentication, companies can boost operational productivity, minimize fraud potential, and maintain compliance with legal standards such as GDPR and financial reporting s tandards. The B I system further enriches decision-making processes by providing real-time insights, trend analysis, and anomaly

detection based on structured document data. The synergy between LLM-based document intelligence and blockchain's decentralized trust framework establishes a robust, future-proof solution for enterprises seeking to streamline document-driven workflows in a secure manner.

6 Conclusion and Future Work

Machine learning, in particular through large-scale language models, has proved to be indispensable to overcome the limitations of traditional BI systems in the processing of unstructured data. By automating the extraction of information from structured documents, ML improves the accuracy of data analysis, reduces human effort and allows for real-time decision-making based on data. Although the best available techniques are showing promising results, further improvements in fine-tuning and model adaptability are needed to further optimise their use in a variety of document types. Future research will focus on using LMDX (Language Model-based Document Information Extraction and Localization) as a foundation for further advancing document extraction techniques. We plan to enhance its capabilities, refining the model for better accuracy and adaptability to various document types. Additionally, we aim to integrate this improved version of LMDX into cloud-based business intelligence systems, enabling real-time, scalable document processing. This integration will help streamline BI processes and empower organizations with more efficient, automated decision-making capabilities in the cloud.

Acknowledgments

This work was conducted as part of the project *Research of Advanced Algorithms and Solutions for Innovative Business Intelligence in the Cloud* - NPOO.C3.2.R3-I1.04.0128.

References

Abolghasemi, M., Gerlach, R., Tarr, G., & Beh, E. (2019). Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. https://doi.org/10.48550/arXiv.1909.13084

Achiam, O. J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H.-i., Bavarian, M., Belgum, J., Bello, I., … Zoph, B. (2023). Gpt-4 technical report. https://api.semanticscholar.org/CorpusID:257532815

Adnan, K., & Akbar, R. (2019). Limitations of information extraction methods and techniques for heterogeneous unstructured big data. *International Journal of Engineering Business Management*, *11*. https://doi.org/10.1177/1847979019890771

Alam, M., Alam, M., Roman, M., Tufail, M., Khan, U., & Khan, M. (2020). Real-time machine-learning based crop/weed detection and classification for variable-rate spraying in precision agriculture, 273–280. https://doi.org/10.1109/ICEEE49618.2020.9102505

Attaran, M., & Deb, P. (2018). Machine learning: The new 'big thing' for competitive advantage. *5*, 277–305. https://doi.org/10.1504/IJKEDM.2018.10015621

Baars, H., & Kemper, H.-G. (2008). Management support with structured and unstructured data - an integrated business intelligence framework. information systems management 25(2):132-148.doi: 10.1080/10580530801941058. *IS Management*, *25*, 132–148. https://doi.org/10.1080/10580530801941058

Baviskar, D., Ahirrao, S., Potdar, V., & Kotecha, K. (2021). Efficient automated processing of the unstructured documents using artificial intelligence: A systematic literature review and future directions. *IEEE Access*, *PP*, 1–1. https://doi.org/10.1109/ACCESS.2021.3072900

Bharadiya, J. P. (2023). The role of machine learning in transforming business intelligence. *Int. J. Comput. Artif. Intell.*, *4*(1), 16–24. https://doi.org/10.33545/27076571.2023.v4.i1a.60

Bloch, T., & Sacks, R. (2018). Comparing machine learning and rule-based inferencing for semantic enrichment of bim models. *Automation in Construction*, *91*, 256–272. https://doi.org/10.1016/j.autcon.2018.03.018

Chukwuelue, A. (2024). Surpassing $1to3$ million revenue threshold: Analyzing why small businesses miss the mark. *Journal of Technology and Systems*, *6*, 14–38. https://doi.org/10.47941/jts.2037

Colakoglu, G., Solmaz, G., & Fürst, J. (2025). Problem solved? information extraction design space for layout-rich documents using llms. https://arxiv.org/abs/2502.18179

Czvetko, T., Kummer, A., Ruppert, T., & Abonyi, J. (2022). Data-driven business process management-based development of industry 4.0 solutions. *CIRP Journal of Manufacturing Science and Technology*, *36*, 117–132. https://doi.org/10.1016/j.cirpj.2021.12.002

De Toni, F., Akiki, C., De La Rosa, J., Fourrier, C., Manjavacas, E., Schweter, S., & Van Strien, D. (2022, May). Entities, dates, and languages: Zero-shot on historical texts with t0. In A. Fan, S. Ilic, T. Wolf, & M. Gallé (Eds.), *Proceedings of bigscience episode #5 – workshop on challenges & perspectives in creating large language models*

(pp. 75–83). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.bigscience-1.7

He, X., Zhou, M., Xu, X., Ma, X., Ding, R., Du, L., Gao, Y., Jia, R., Xu, C., Han, S., Yuan, Z., & Zhang, D. (2024). Text2analysis: A benchmark of table question answering with advanced data analysis and unclear queries. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*, 18206–18215. https://doi.org/10.1609/aaai.v38i16.29779

Hindle, G., Kunc, M., Mortenson, M., Oztekin, A., & Vidgen, R. (2019). Business analytics: Defining the field and identifying a research agenda. *European Journal of Operational Research*, *281*. https://doi.org/10.1016/j.ejor.2019.10.001

Hong, T., Kim, D., Ji, M., Hwang, W., Nam, D., & Park, S. (2022). Bros: A pre-trained language model focusing on text and layout for better key information extraction from documents. https://arxiv.org/abs/2108.04539

Hu, Y., Ameer, I., Zuo, X., Peng, X., Zhou, Y., Li, Z., Li, Y., Li, J., Jiang, X., & Xu, H. (2023). Improving large language models for clinical named entity recognition via prompt engineering. *Journal of the American Medical Informatics Association : JAMIA*, *31*, 1812–1820. https://api.semanticscholar.org/CorpusID:257805032

Huang, G.-B., Zhu, Q.-Y., & Siew, C. (2006). Real-time learning capability of neural networks. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, *17*, 863–78. https://doi.org/10.1109/TNN.2006.875974

Huang, Y., Lv, T., Cui, L., Lu, Y., & Wei, F. (2022). Layoutlmv3: Pre-training for document ai with unified text and image masking. *Proceedings of the 30th ACM International Conference on Multimedia*, 4083–4091. https://doi.org/10.1145/3503161.3548112

Jayaprakash, S., Nagarajan, M. D., Prado, R. P. d., Subramanian, S., & Divakarachari, P. B. (2021). A systematic review of energy management strategies for resource allocation in the cloud: Clustering, optimization and machine learning. *Energies*, *14*(17), 5322.

Keraghel, I., Morbieu, S., & Nadif, M. (2024). Recent advances in named entity recognition: A comprehensive survey and comparative study. https://api.semanticscholar.org/CorpusID:267060999

Kim, G., Hong, T., Yim, M., Nam, J., Park, J., Yim, J., Hwang, W., Yun, S., Han, D., & Park, S. (2022, October). Ocr-free document understanding transformer. https://doi.org/10.1007/978-3-031-19815-1_29

Kumar, S., & Singh, M. (2019). Big data analytics for healthcare industry: Impact, applications, and tools. *Big Data Mining and Analytics*, *2*(1), 48–57. https://doi.org/10.26599/BDMA.2018.9020031

Laskar, M. T. R., Bari, M. S., Rahman, M., Bhuiyan, M. A. H., Joty, S. R., & Huang, J. (2023). A systematic study and comprehensive evaluation of chatgpt on benchmark datasets. *ArXiv*, *abs/2305.18486*. https://api.semanticscholar.org/CorpusID:258967462

Lee, K., Joshi, M., Turc, I., Hu, H., Liu, F., Eisenschlos, J., Khandelwal, U., Shaw, P., Chang, M.-W., & Toutanova, K. (2023). Pix2struct: Screenshot parsing as pretraining for visual language understanding. https://arxiv.org/abs/2210.03347

Liu, Q., He, Y., Xu, T., Lian, D., Liu, C., Zheng, Z., & Chen, E. (2024). Unimel: A unified framework for multimodal entity linking with large language models, 1909–1919. https://doi.org/10.1145/3627673.3679793

Lyu, Y., Niu, Z., Xie, Z., Zhang, C., Xu, T., Wang, Y., & Chen, E. (2024). Retrieve-plan-generation: An iterative planning and answering framework for knowledge-intensive llm generation, 4683–4702. https://doi.org/10.18653/v1/2024.emnlp-main.270

Mahadevkar, S., Patil, S., Kotecha, K., Soong, L., & Choudhury, T. (2024). Exploring ai-driven approaches for unstructured document analysis and future horizons. *Journal of Big Data*, *11*. https://doi.org/10.1186/s40537-024-00948-z

Majid, M., Marinova, D., Hossain, A., & Rummani, F. (2024). Use of conventional business intelligence (bi) systems as the future of big data analysis. *American Journal of Information Systems*, *9*, 1–10. https://doi.org/10.12691/ajis-9-1-1

Morariu, C., Morariu, O., Raileanu, S., & Borangiu, T. (2020). Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems. *Computers in Industry*, *120*, 103244. https://doi.org/10.1016/j.compind.2020.103244

Oza, R., Punjani, D., & Domadiya, D. (2023). Analysis of unstructured data using artificial intelligence. *International Journal of Creative Research Thoughts*, *11*(5), 2320–2882.

Park, S., Shin, S., Lee, B., Lee, J., Surh, J., Seo, M., & Lee, H. (2019). Cord: A consolidated receipt dataset for post-ocr parsing. https://api.semanticscholar.org/CorpusID:207900784

Park, S.-H., Lee, D.-G., Park, J.-S., & Kim, J. W. (2021). A survey of research on data analytics-based legal tech. *Sustainability*, *13*, 8085. https://doi.org/10.3390/su13148085

Pasupuleti, V., Thuraka, B., Kodete, C. S., & Malisetty, S. (2024). Enhancing supply chain agility and sustainability through machine learning: Optimization techniques for logistics and inventory management. *Logistics*, *8*, 73. https://doi.org/10.3390/logistics8030073

Perot, V., Kang, K., Luisier, F., Su, G., Sun, X., Boppana, R. S., Wang, Z., Wang, Z., Mu, J., Zhang, H., Lee, C.-Y., & Hua, N. (2024). LMDX: Language model-based document information extraction and

localization. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Findings of the association for computational linguistics: Acl 2024* (pp. 15140–15168). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-acl.899

Prasanth, A., Vadakkan, D. J., Surendran, P., & Thomas, B. (2023). Role of artificial intelligence and business decision making. *International Journal of Advanced Computer Science and Applications*, *14*(6). https://doi.org/10.14569/IJACSA.2023.01406103

Sage, C., Aussem, A., Elghazel, H., Eglin, V., & Espinas, J. (2019). Recurrent neural network approach for table field extraction in business documents. *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 1308–1313. https://doi.org/10.1109/ICDAR.2019.00211

Sage, C., Douzon, T., Aussem, A., Eglin, V., Elghazel, H., Duffner, S., Garcia, C., & Espinas, J. (2021). Data-efficient information extraction from documents with pre-trained language models. *ICDAR Workshops*. https://api.semanticscholar.org/CorpusID:237345634

Sever, M. M. (2024). Digital transformation and business intelligence (bi) in the industry 4.0 (i 4.0) age. In P. Keikhosrokiani (Ed.), *Data-driven business intelligence systems for socio-technical organizations* (pp. 28–54). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-1210-0.ch002

Sharma, K., Shetty, A., Jain, A., & Dhanare, R. (2021). A comparative analysis on various business intelligence (bi), data science and data analytics tools, 1–11. https://doi.org/10.1109/ICCCI50826.2021.9402226

Singh, S., & Hooda, S. (2023). A study of challenges and limitations to applying machine learning to highly unstructured data. *7th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 1–6. https://doi.org/10.1109/ICCUBEA58933.2023.10392115

Taddy, M. (2018, February). *The Technological Elements of Artificial Intelligence* (NBER Working Papers No. 24301). National Bureau of Economic Research, Inc. https://ideas.repec.org/p/nbr/nberwo/24301.html

Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*. https://api.semanticscholar.org/CorpusID:117248575

Wang, S., Sun, X., Li, X., Ouyang, R., Wu, F., Zhang, T., Li, J., & Wang, G. (2023). Gpt-ner: Named entity recognition via large language models. *ArXiv*, *abs/2304.10428*. https://api.semanticscholar.org/CorpusID:258236561

Wang, Z., Zhou, Y., Wei, W., Lee, C.-Y., & Tata, S. (2023). Vrdu: A benchmark for visually-rich document understanding. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5184–5193. https://doi.org/10.1145/3580305.3599929

Xu, D., Chen, W., Peng, W., Zhang, C., Xu, T., Zhao, X., Wu, X., Zheng, Y., & Chen, E. (2023). Large language models for generative information extraction: A survey. *Frontiers Comput. Sci.*, *18*, 186357. https://api.semanticscholar.org/CorpusID:266690657

Xu, Y., Xu, Y., Lv, T., Cui, L., Wei, F., Wang, G., Lu, Y., Florencio, D., Zhang, C., Che, W., Zhang, M., & Zhou, L. (2022). Layoutlmv2: Multi-modal pre-training for visually-rich document understanding. https://arxiv.org/abs/2012.14740

Yang, M., Lim, M. K., Qu, Y., Ni, D., & Xiao, Z. (2019). Supply chain risk management with machine learning technology: A literature review and future research directions. *Computers & Industrial Engineering*, *175*, 108859.

Yang, X., Yumer, E., Asente, P., Kraley, M., Kifer, D., & Lee Giles, C. (2017). Learning to extract semantic structure from documents using multimodal fully convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yue, C., Xu, X., Ma, X., Du, L., Ding, Z., Han, S., Zhang, D., & Zhang, Q. (2024). Extract information from hybrid long documents leveraging llms: A framework and dataset. https://doi.org/10.48550/arXiv.2412.20072

Zmigrod, R., Wang, D., Sibue, M., Pei, Y., Babkin, P., Brugere, I., Liu, X., Navarro, N., Papadimitriou, A., Watson, W., Ma, Z., Nourbakhsh, A., & Shah, S. (2024). Buddie: A business document dataset for multi-task information extraction. https://arxiv.org/abs/2404.04003

Pyramidal Recursive Composition of Multi-Word Units into Unified Representations

Piramidalna rekurzivna kompozicija višerječnih izraza u objedinjene reprezentacije

Karlo Babić, Ana Meštrović
University of Rijeka
Faculty of Informatics and Digital Technologies
Radmile Matejčić 2, Rijeka, Croatia
{karlo.babic, amestrovic}@uniri.hr

Abstract. *In this paper, we explore the composition of word embeddings to create richer, more meaningful representations of multi-word units. Existing methods, such as averaging word embeddings, provide simple and efficient approaches. However, they often fail to capture the complexity of multi-word interactions. To address this, we employ the Pyramidal Recursive learning (PyRv) method, which recursively combines word embeddings into unified representations. Originally developed for constructing representations hierarchically from subwords to phrases, PyRv is well-suited for progressively merging individual word vectors into phrase vectors. We evaluate the effectiveness of PyRv for embedding composition using fastText embeddings on the dependency relation labeling task. Using a single fastText word embedding yields an accuracy of 71%. Averaging five fastText word embeddings (the middle word and its four neighboring words) results in a significant drop in accuracy to 34%. In contrast, by composing five word embeddings with PyRv, we achieve an accuracy of 77%, demonstrating the superior ability of PyRv to integrate multiple word embeddings into more expressive representations. These findings highlight the potential of PyRv as a lightweight yet powerful technique for word embedding composition.*

Keywords. Composition, embeddings, pyramidal, recursive neural network, text representation learning

Sažetak. *U ovom radu istražujemo kompoziciju vektorskih reprezentacija riječi (eng. "word embeddings") s ciljem stvaranja semantički bogatijih i smislenijih reprezentacija višerječnih izraza. Postojeće metode, poput prosječne vrijednosti vektorskih reprezentacija riječi, nude jednostavan i učinkovit pristup. Međutim, takvi pristupi često ne uspijevaju obuhvatiti složenost odnosa između riječi. Kako bismo to prevladali, koristimo metodu Pyramidal Recursive learning (PyRv), koja rekurzivno kombinira vektorske reprezentacije riječi u objedinjene reprezentacije. Izvorno razvijena za hijerarhijsku izgradnju reprezentacija od podriječi do fraza, PyRv je prikladna za postupno spajanje pojedinačnih vektora riječi u vektore fraza. Evaluaciju metode PyRv za kompoziciju vektorskih reprezentacija provodimo korištenjem fastText vektora riječi u zadatku označavanja zavisnosnih odnosa. Korištenjem fastText vektorske reprezentacije jedne riječi postiže se točnost od 71%. Prosječna vrijednost pet fastText vektora riječi (središnje riječi i njezine četiri susjedne) rezultira značajnim padom točnosti na 34%. Suprotno tome, kompozicijom pet vektora riječi metodom PyRv postiže se točnost od 77%, što pokazuje nadmoćnu sposobnost metode PyRv u integraciji više vektora riječi u izražajnije reprezentacije. Ovi rezultati ukazuju na potencijal metode PyRv kao lagane, ali snažne tehnike za kompoziciju vektorskih reprezentacija riječi.*

Ključne riječi. Kompozicija, vektorske reprezentacije, piramidalno, rekurzivna neuronska mreža, učenje reprezentacija teksta

1 Introduction

Word embeddings are foundational to many natural language processing (NLP) tasks, providing a way to map words into continuous vector spaces that capture semantic relationships between them. By converting words

1 Uvod

Vektorske reprezentacije riječi temelj su mnogih zadataka obrade prirodnog jezika (NLP), jer omogućuju preslikavanje riječi u kontinuirane vektorske prostore koji zadržavaju semantičke odnose među njima. Pre-

into numerical representations, word embeddings allow machines to process and understand text in a way that retains important linguistic properties. Popular word embedding methods, such as Word2Vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014), and FastText (Bojanowski et al., 2017), have demonstrated the utility of these representations by positioning semantically similar words closer in the vector space. The effectiveness of NLP models often hinges on the quality of these embeddings, as richer and more informative representations can lead to improved performance across various downstream tasks.

In many real-world applications, however, understanding text at the word level alone is insufficient. The need to represent larger linguistic units, such as phrases or sentences, necessitates techniques for combining word embeddings into more complex structures. Word composition, which involves aggregating individual word embeddings to represent multi-word expressions or entire sentences, serves this purpose. By integrating information from multiple word vectors, compositional methods aim to capture both the meanings of individual words and the syntactic and semantic interactions between them, particularly in morphologically complex languages, such as Croatian, which was used for evaluation in this paper.

There are several established methods for combining word embeddings. Simple techniques include element-wise operations such as addition, averaging, or multiplication, which produce a composite vector by leveraging individual word vectors (Arora et al., 2017; Joulin et al., 2016). These methods, while computationally efficient, may fail to fully capture the complexity of phrase or sentence meaning. More sophisticated approaches employ weighted combinations, context-aware methods, or syntactic structures to improve the expressiveness of the resultant embeddings (Babić et al., 2020; Bahdanau, 2014; Socher et al., 2013).

While transformer-based models, such as BERT (Devlin et al., 2018) and GPT (Brown et al., 2020), offer robust pre-trained sentence embeddings by learning deep contextual representations, they are often computationally expensive and may not always align with the specific needs of certain tasks. Word composition methods provide a more lightweight and flexible alternative, especially in cases where transparency and control over the aggregation process are crucial. Additionally, word-level composition techniques can better retain the granularity of individual word meanings, which is sometimes diluted in sentence-level embeddings produced by transformer models.

Word composition provides a valuable approach for constructing meaningful representations of multi-word units, balancing computational efficiency and interpretability. These methods remain relevant, particularly in domains where sentence embedding techniques may obscure important details or where domain-

tvaranjem riječi u numeričke reprezentacije, vektori riječi omogućuju računalima obradu i razumijevanje teksta na način koji očuva važne jezične značajke. Popularne metode za generiranje vektorskih reprezentacija riječi, poput Word2Vec (Mikolov i dr., 2013), GloVe (Pennington i dr., 2014) i FastText (Bojanowski i dr., 2017), pokazale su korisnost ovih reprezentacija tako što pozicioniraju semantički slične riječi bliže jedna drugoj u vektorskom prostoru. Učinkovitost NLP modela često uvelike ovisi o kvaliteti ovih reprezentacija, budući da bogatije i informativnije reprezentacije mogu dovesti do boljih rezultata u različitim zadacima.

Međutim, u mnogim stvarnim primjenama, razumijevanje teksta na razini pojedinačnih riječi nije dovoljno. Potreba za reprezentacijom većih jezičnih jedinica, poput fraza ili rečenica, zahtijeva tehnike za kombiniranje vektora riječi u složenije strukture. Kompozicija riječi, koja podrazumijeva agregaciju pojedinačnih vektorskih reprezentacija kako bi se prikazale višerječne izraze ili čitave rečenice, ima upravo tu svrhu. Integriranjem informacija iz više vektora riječi, kompozicijske metode nastoje obuhvatiti značenja pojedinih riječi, kao i sintaktičke i semantičke odnose među njima – osobito u morfološki složenim jezicima, poput hrvatskog, koji je korišten u evaluaciji ovog rada.

Postoji više poznatih metoda za kombiniranje vektora riječi. Jednostavne tehnike uključuju elementne operacije, poput zbrajanja, prosječavanja ili množenja, koje stvaraju kompozitni vektor korištenjem pojedinačnih vektora riječi (Arora i dr., 2017; Joulin i dr., 2016). Iako su te metode računski učinkovite, često ne uspijevaju u potpunosti obuhvatiti značenje izraza ili rečenice. Složeniji pristupi koriste ponderirane kombinacije, metode osjetljive na kontekst ili sintaktičke strukture kako bi poboljšali izražajnost dobivenih reprezentacija (Babić i dr., 2020; Bahdanau, 2014; Socher i dr., 2013).

Iako modeli temeljeni na transformerima, poput BERT-a (Devlin i dr., 2018) i GPT-a (Brown i dr., 2020), nude snažne unaprijed naučene vektorske reprezentacije rečenica učenjem dubokih kontekstualnih značenja, često su računski zahtjevni i ne odgovaraju uvijek specifičnim potrebama određenih zadataka. Metode kompozicije riječi predstavljaju lakšu i fleksibilniju alternativu, osobito u slučajevima gdje su transparentnost i kontrola nad procesom agregacije od presudne važnosti. Osim toga, metode temeljene na kompoziciji na razini riječi bolje zadržavaju granularnost značenja pojedinih riječi, što se ponekad gubi u rečeničnim reprezentacijama koje generiraju modeli transformera.

Kompozicija riječi pruža vrijedan pristup za izgradnju smislenih reprezentacija višerječnih izraza, uravnotežujući računalnu učinkovitost i interpretabilnost. Ove metode ostaju relevantne osobito u područjima gdje tehnike vektorskog reprezentiranja rečenica mogu prikriti važne pojedinosti ili gdje je potrebna domensko-specifična prilagodba kompozicije vektora.

specific customization of embedding composition is required.

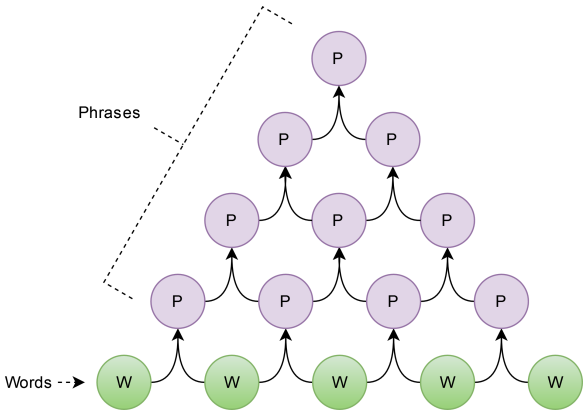
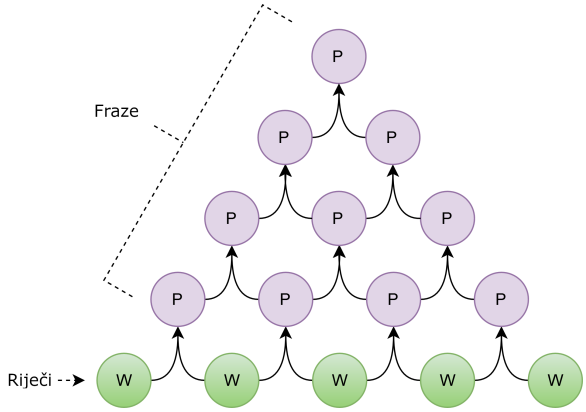


Figure 2: A visual representation of pyramidal recursion in the PyRv+FT method. The lowest-level nodes correspond to fastText-embedded words. As we move upward, the nodes represent combined word embeddings, capturing phrase-level meanings.



Slika 2: Vizualni prikaz piramidalne rekurzije u metodi PyRv+FT. Najniži čvorovi predstavljaju riječi u fastText vektorskoj reprezentaciji. Viši čvorovi prikazuju kombinirane vektore riječi koji obuhvaćaju značenja na razini fraza.

In this work, we leverage the Pyramidal Recursive learning (PyRv) method, introduced in our previous paper (Babić and Meštrović, 2024), to compose multiple word embeddings into unified representations. PyRv facilitates structured composition through its hierarchical learning approach, recursively combining representations at each level. Initially developed for constructing representations from tokens (subwords) up to phrases (as illustrated in Figure 1), PyRv is well-suited for combining word embeddings by progressively merging individual word vectors into phrase vectors (as is shown in Figure 2).

One of the main properties of the PyRv method is representation compositionality, which enables the composition of multiple embeddings into a coherent, semantically rich representation. This property aligns with the objective of this paper, where the focus is on effectively combining word embeddings to capture more complex linguistic structures. By recursively merging embeddings, PyRv maintains the semantic integrity of each word while constructing increasingly abstract representations at higher levels of the hierarchy.

The primary contribution of this paper is the introduction of a method for composing multi-word units into unified representations using Pyramidal Recursive learning (PyRv). To assess the effectiveness of this approach, we train the PyRv embedding model on Croatian texts and compare its performance to the widely-used baseline method of averaging word embeddings. In addition, we explore the structure of the representation space generated by PyRv's composition method. Our evaluation results validate PyRv's compositionality property.

Following this introduction, the subsequent sections of this paper are structured as follows: In Section 2,

U ovom radu koristimo metodu Pyramidal Recursive learning (PyRv), predstavljenu u našem prethodnom radu (Babić i Meštrović, 2024), kako bismo kompozicijom više vektora riječi dobili objedinjene reprezentacije. PyRv omogućuje strukturirano kombiniranje kroz hijerarhijski pristup učenju, rekurzivno kombinirajući reprezentacije na svakoj razini. Izvorno razvijena za izgradnju reprezentacija od tokena (podriječi) do fraza (kao što je prikazano na slici 1), metoda PyRv posebno je pogodna za kombiniranje vektora riječi postupnim spajanjem pojedinačnih vektora u vektorske reprezentacije fraza (kao što je prikazano na slici 2).

Jedno od glavnih svojstava metode PyRv jest kompozicionalnost reprezentacija, koja omogućuje stvaranje koherentnih, semantički bogatih reprezentacija iz više vektora. Ovo svojstvo izravno se poklapa s ciljem ovog rada, koji se fokusira na učinkovito kombiniranje vektorskih reprezentacija riječi kako bi se obuhvatile složenije jezične strukture. Rekurzivnim spajanjem vektora, PyRv održava semantički integritet svake riječi pri izgradnji reprezentacija na višim razinama hijerarhije.

Glavni doprinos ovog rada je predstavljanje metode za kompoziciju višerječnih izraza u objedinjene reprezentacije pomoću piramidalnog rekurzivnog učenja (PyRv). Kako bismo evaluirali učinkovitost ovog pristupa, treniramo PyRv model reprezentacija na hrvatskim tekstovima i uspoređujemo njegovu učinkovitost s često korištenom referentnom metodom prosječavanja vektora riječi. Također istražujemo strukturu vektorskog prostora koji proizlazi iz PyRv metode kompozicije. Rezultati evaluacije potvrđuju svojstvo kompozicionalnosti metode PyRv.

Nakon ovog uvoda, sljedeća poglavlja rada organizirani su kako slijedi: U poglavlju 2, "Povezani ra-

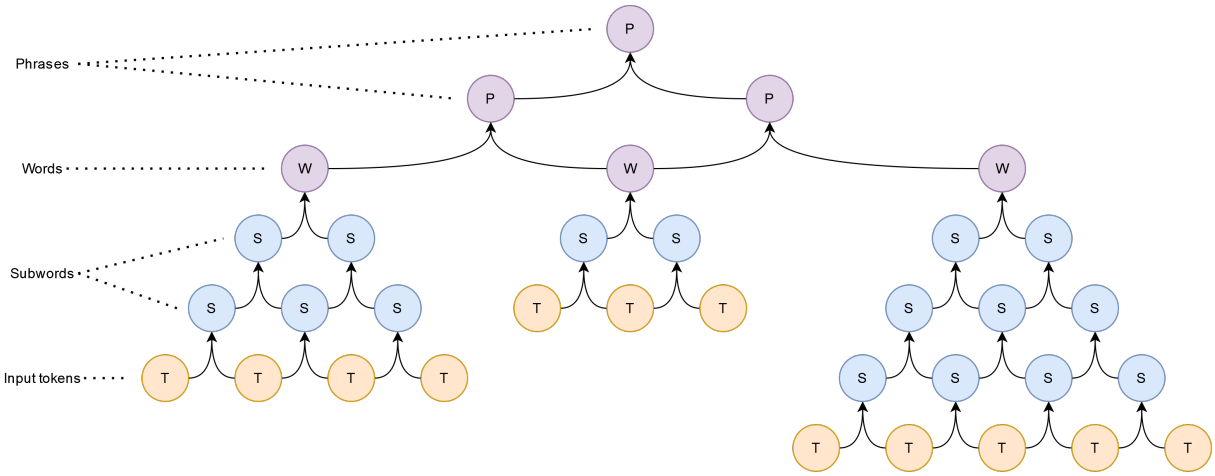


Figure 1. A visualized example of a pyramidal recursion in the PyRv method. The lowest-level nodes correspond to input tokens. Moving upward, the nodes within the three pyramids represent combined subword embeddings. At the pyramid peaks, nodes represent word embeddings, and higher nodes signify combined word embeddings, representing phrases.

Slika 1. Vizualizirani primjer piramidalne rekurzije u metodi PyRv.Čvorovi na najnižoj razini odgovaraju ulaznim tokenima. Na višim razinama unutar triju piramida nalaze se čvorovi koji predstavljaju kombinirane podriječne vektore. Na vrhovima piramida nalaze se vektori riječi, a još viši čvorovi predstavljaju kombinirane vektore riječi, tj. fraze.

"Related Work," we explore prior research, highlighting methods that compose word embeddings. Section 3, "Methodology," presents our method for composition of word embeddings. Section 4, "Evaluation," details the datasets, experiment setup, and results. Finally, in Section 5, "Conclusion," we conclude with a summary of our contributions to the field and discuss future directions.

2 Related work

Composing word embeddings to generate meaningful representations of larger text units remains a critical area of study in natural language processing. Approaches to this problem span from simple aggregation techniques that compose word embeddings to more complex neural architectures that embed entire sentences, each offering unique advantages in different contexts.

A common approach is to average word embeddings to generate a single vector representing a phrase or sentence. Averaged word embeddings were used with fastText for efficient text classification (Joulin et al., 2016). This technique, inspired by the continuous bag of words (CBOW) model (Mikolov et al., 2013), offers a computationally lightweight solution that performs competitively with deeper models in various NLP tasks.

Building upon this, an enhanced version was proposed where word embeddings are combined using weighted averages, followed by post-processing through PCA/SVD (Arora et al., 2017). The weighting scheme they propose significantly improves performance on tex-

dovi," prikazujemo dosadašnja istraživanja i metode kompozicije vektora riječi. Poglavlje 3, "Metodologija," predstavlja naš pristup kompoziciji vektorskih reprezentacija. Poglavlje 4, "Evaluacija," donosi podatkovne skupove, eksperimentalnu postavu i rezultate. Na kraju, u poglavlju 5, "Zaključak," sažimamo doprinos rada i iznosimo smjernice za budući rad.

2 Povezani radovi

Kompozicija vektorskih reprezentacija riječi s ciljem generiranja smislenih prikaza većih jezičnih jedinica ostaje ključno područje istraživanja u obradi prirodnog jezika. Pristupi ovom problemu kreću se od jednostavnih tehnika agregacije, koje kombiniraju vektore riječi, do složenijih neuronskih arhitektura koje vektoriziraju cijele rečenice, pri čemu svaki pristup ima specifične prednosti u različitim kontekstima.

Uobičajen pristup je prosječno kombiniranje vektora riječi radi generiranja jednog vektora koji predstavlja frazu ili rečenicu. Prosječni vektori riječi korišteni su s modelom fastText za učinkovitu klasifikaciju teksta (Joulin i dr., 2016). Ova tehnika, inspirirana modelom continuous bag of words (CBOW) (Mikolov i dr., 2013), nudi računalno učinkovit pristup koji u brojnim NLP zadacima konkurira dubljim modelima.

Na tome se nadograđuje poboljšana verzija u kojoj se vektori riječi kombiniraju pomoću ponderiranih prosjeka, a zatim se dodatno obrađuju metodama poput PCA/SVD (Arora i dr., 2017). Predloženi sustav ponderiranja značajno poboljšava rezultate u zadacima se-

tual similarity tasks. This method demonstrates that simple compositional techniques can rival more complex architectures, especially in unsupervised settings.

A comparative study highlighted the trade-offs between simple word averaging and more complex models like LSTMs for sentence embedding (Wieting et al., 2015). Their findings showed that while LSTMs perform well on in-domain data, simple word averaging techniques tend to outperform LSTMs in out-of-domain tasks. This suggests that straightforward compositional methods, despite their simplicity, are robust and generalizable across diverse datasets.

A recursive neural network based on syntactic parse trees was proposed to generate phrase and sentence representations, useful in tasks like sentiment analysis (Socher et al., 2013). A Self-Adaptive Hierarchical Sentence Model was introduced, using recursive structures without relying on syntax and showing that supervised learning can effectively capture compositional semantics in a non-syntactic hierarchy (Zhao et al., 2015).

Transformer (Vaswani, 2017) architectures use attention mechanism to compose embeddings. The attention mechanism was introduced in neural machine translation, allowing models to dynamically focus on different parts of the input sequence during decoding (Bahdanau, 2014). The introduction of attention helped relieve the encoder from compressing all information into a single fixed-length vector, thus enabling more flexible and effective composition of representations over sequential data.

In this work, we build on these approaches by applying the Pyramidal Recursive learning (PyRv) method (Babić and Meštrović, 2024) to recursively combine word embeddings into more abstract representations. Unlike the averaging techniques (Arora et al., 2017; Joulin et al., 2016), PyRv enables hierarchical composition, capturing both word-level and higher-level semantic structures in a more structured way. Unlike most other recursive models, which often rely on syntactic trees (Socher et al., 2013), PyRv operates without requiring explicit parse structures. Additionally, unlike hierarchical models that require supervision (Zhao et al., 2015), PyRv is fully unsupervised. Additionally, by recursively merging embeddings, our approach offers a simpler alternative to attention mechanisms for constructing rich representations of text.

3 Methodology

In this section, we introduce word embeddings and common composition techniques, followed by a detailed explanation of how the PyRv method is applied to compose word embeddings.

mantičke sličnosti teksta. Ova metoda pokazuje da jednostavne kompozicijske tehnike mogu biti usporedive sa složenijim arhitekturama, osobito u ne-nadgledanim okruženjima.

Jedno komparativno istraživanje istaknulo je kompromise između prosječnog kombiniranja riječi i složenijih modela poput LSTM-ova za rečenične reprezentacije (Wieting i dr., 2015). Rezultati pokazuju da LSTM-ovi dobro funkcioniraju na podacima iz iste domene, dok jednostavne tehnike prosjeka često nadmašuju LSTM-ove u zadacima izvan domena. To sugerira da su jednostavne kompozicijske metode, unatoč svojoj jednostavnosti, robusne i prenosive na različite skupove podataka.

Predložena je rekurzivna neuronska mreža temeljena na sintaktičkim stablima za generiranje reprezentacija fraza i rečenica, korisna u zadacima poput analize sentimenta (Socher i dr., 2013). Predstavljen je i Self-Adaptive Hierarchical Sentence Model koji koristi rekurzivne strukture bez oslanjanja na sintaksu te pokazuje da nadgledano učenje može učinkovito uhvatiti semantiku kompozicije u nesintaktičkoj hijerarhiji (Zhao i dr., 2015).

Arhitekture transformera (Vaswani, 2017) koriste mehanizam pažnje (attention) za kompoziciju reprezentacija. Taj mehanizam prvi je put uveden u kontekstu neuronskog strojnog prevođenja, gdje je modelima omogućeno da se tijekom dekodiranja dinamički fokusiraju na različite dijelove ulazne sekvence (Bahdanau, 2014). Uvođenje mehanike pažnje oslobodilo je enkoder potrebe da sve informacije sažme u jedan vektor fiksne duljine, čime se omogućila fleksibilnija i učinkovitija kompozicija reprezentacija nad sekvencijalnim podacima.

U ovom radu gradimo na spomenute pristupe koristeći metodu piramidalnog rekurzivnog učenja (PyRv) (Babić i Meštrović, 2024) kako bismo rekurzivno kombinirali vektorske reprezentacije riječi. Za razliku od metoda prosjeka (Arora i dr., 2017; Joulin i dr., 2016), PyRv omogućuje hijerarhijsku kompoziciju, uzimajući u obzir semantičke strukture na razini riječi i višim razinama na strukturiraniji način. Za razliku od većine drugih rekurzivnih modela, koji se često oslanjaju na sintaktička stabla (Socher i dr., 2013), PyRv ne zahtijeva eksplicitne strukture za parsiranje. Također, za razliku od hijerarhijskih modela koji zahtijevaju nadgledano učenje (Zhao i dr., 2015), PyRv je potpuno nenadgledan. Nadalje, rekurzivnim spajanjem reprezentacija, naš pristup nudi jednostavniju alternativu mehanizmima pažnje u konstrukciji semantički bogatih reprezentacija teksta.

3 Metodologija

U ovom poglavlju predstavljamo vektorske reprezentacije riječi (eng. "word embeddings") i uobičajene tehnike kompozicije, nakon čega slijedi detaljno objašnjenje kako se metoda piramidalnog rekurzivnog učenja

3.1 Word embedding

Word embeddings, such as those produced by fastText (Bojanowski et al., 2017), are commonly used to represent individual words. To represent multiple words, basic techniques like averaging or concatenation can be applied. However, both approaches come with notable limitations.

When averaging embeddings (e.g., taking the mean of multiple word vectors), important information, particularly word order, is lost. Concatenation preserves all information but presents two significant challenges.

First, concatenated embeddings vary in dimensionality depending on the number of words, which complicates their use as input for models that require a fixed input size. Second, the dimensionality of concatenated representations can grow excessively large. For instance, a 5-word phrase embedded with 300-dimensional fastText results in a 1500-dimensional vector (5x300).

In our evaluation, we use averaging to compose Croatian fastText embeddings (Grave et al., 2018) to maintain a consistent dimensionality across all methods, ensuring that the results are comparable.

3.2 PyRv with fastText word embeddings

Pyramidal Recursive learning (PyRv) is a method designed to construct hierarchical representations of text, moving progressively from low-level units such as characters or subwords to higher-level representations such as words, phrases, sentences, and even paragraphs. PyRv combines representations recursively, forming increasingly abstract and semantically rich embeddings at each level of the hierarchy.

To address the limitations of averaging and concatenation, we explore the use of PyRv for composing multiple word embeddings into a single, unified representation. Unlike our previous work on PyRv (Babić and Meštrović, 2024), where the recursion starts from subwords or tokens, in this study we begin with word embeddings produced by fastText. This hybrid approach is referred to as PyRv+FT.

For this paper, we use pre-trained Croatian fastText word vectors (Grave et al., 2018) to embed words, which are then recursively combined into phrase embeddings via the PyRvNN model. The PyRv+FT embeddings are trained on Croatian Wikipedia texts (Wikimedia public dump, January 11, 2020) for 10 epochs.

We evaluate PyRv+FT on two downstream tasks, described in detail in the next section. Through this process, we investigate how PyRv improves the com-

(eng. "Pyramidal Recursive learning", PyRv) primjenjuje za kompoziciju vektorskih reprezentacija riječi.

3.1 Vektorske reprezentacije riječi

Vektorske reprezentacije riječi (eng. "word embeddings"), poput onih koje generira fastText (Bojanowski i dr., 2017), često se koriste za reprezentaciju pojedinačnih riječi. Za reprezentaciju više riječi mogu se primijeniti osnovne tehnike poput prosjeka ili konkatenacije. Međutim, obje metode imaju značajna ograničenja.

Kod prosjeka (npr. računanja srednje vrijednosti više vektora riječi), gubi se važna informacija, posebice redoslijed riječi. Konkatenacija zadržava sve informacije, ali donosi dva velika izazova.

Prvo, vektori dobiveni konkatenacijom variraju u dimenzionalnosti ovisno o broju riječi, što otežava njihovu primjenu kao ulaz za modele koji zahtijevaju fiksnu veličinu ulaza. Drugo, dimenzionalnost konkateniranih reprezentacija može postati pretjerano velika. Na primjer, fraza od 5 riječi vektorizirana pomoću fastText-a s 300 dimenzija rezultira vektorom od 1500 dimenzija (5x300).

U našoj evaluaciji koristimo metodu prosjeka za kompoziciju hrvatskih fastText vektora riječi (Grave i dr., 2018) kako bismo održali konzistentnu dimenzionalnost među svim metodama i time omogućili usporedivost rezultata.

3.2 PyRv s fastText vektorskim reprezentacijama riječi

Piramidalno rekurzivno učenje (eng. "Pyramidal Recursive learning", PyRv) je metoda osmišljena za izgradnju hijerarhijskih reprezentacija teksta, koja progresivno prelazi s nižih razina poput znakova ili podriječi na više razine poput riječi, fraza, rečenica pa čak i odlomaka. PyRv rekurzivno kombinira reprezentacije, formirajući sve apstraktnije i semantički bogate vektore na svakoj razini hijerarhije.

Kako bismo zaobišli ograničenja prosjeka i konkatenacije, istražujemo primjenu metode PyRv za kompoziciju više vektorskih reprezentacija riječi u jednu objedinjenu reprezentaciju. Za razliku od našeg prethodnog rada o metodi PyRv (Babić i Meštrović, 2024), gdje rekurzija započinje s podriječima ili tokenima, u ovom istraživanju polazimo od vektorskih reprezentacija riječi generiranih pomoću fastText-a. Ovaj hibridni pristup nazivamo PyRv+FT.

U ovom radu koristimo unaprijed istrenirane hrvatske fastText vektore riječi (Grave i dr., 2018) za vektorizaciju riječi, koje se zatim rekurzivno kombiniraju u frazne reprezentacije pomoću modela PyRvNN.
PyRv+FT vektori trenirani su na tekstovima s hrvatske Wikipedije (Wikimedia javni podaci, 11. siječnja 2020.) kroz 10 epoha.

Evaluiramo PyRv+FT na dva zadatka krajnje pri-

positionality of word embeddings, while maintaining manageable dimensionality and enhancing representational quality.

3.3 Mathematical Formulation

To formalize the distinction between averaging and our proposed PyRv+FT approach, we define their respective composition functions. Let a phrase P be a sequence of n words, $P = (w_1, w_2, \ldots, w_n)$. Each word w_i is mapped to a d-dimensional vector embedding $\boldsymbol{v}_i \in \mathbb{R}^d$ using fastText. The goal is to compose the sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ into a single phrase embedding $\boldsymbol{e}_P \in \mathbb{R}^d$.

Averaging. The averaging method computes the phrase embedding $\boldsymbol{e}_P^{\text{avg}}$ by taking the element-wise mean of the constituent word vectors:

$$\boldsymbol{e}_P^{\text{avg}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{v}_i \tag{1}$$

The primary limitation of this model is apparent from its formulation. Since vector addition is commutative ($\boldsymbol{v}_i + \boldsymbol{v}_j = \boldsymbol{v}_j + \boldsymbol{v}_i$), the resulting embedding $\boldsymbol{e}_P^{\text{avg}}$ is invariant to the order of words in the phrase, losing crucial syntactic and semantic information.

Pyramidal Recursive Composition (PyRv+FT). The PyRv+FT method constructs the phrase embedding $\boldsymbol{e}_P^{\text{PyRv}}$ through a hierarchical and recursive process. The composition starts at the lowest level (level 0), which consists of the initial word embeddings $(\boldsymbol{v}_1^{(0)}, \boldsymbol{v}_2^{(0)}, \ldots, \boldsymbol{v}_n^{(0)})$.

At each subsequent level k, adjacent pairs of vectors from level $k-1$ are combined using a parameterized composition function f_θ (the PyRvNN model) to form a new, higher-level representation. This process is repeated until a single vector remains. A new vector $\boldsymbol{v}_i^{(k)}$ at level k is generated from level $k-1$ as follows:

$$\boldsymbol{v}_i^{(k)} = f_\theta(\boldsymbol{v}_i^{(k-1)}, \boldsymbol{v}_{i+1}^{(k-1)}) \tag{2}$$

The final phrase embedding $\boldsymbol{e}_P^{\text{PyRv}}$ is the single vector remaining at the top of the pyramid.

Unlike averaging, the composition function f_θ is non-commutative because concatenation is order-dependent (i.e., $[\boldsymbol{a}; \boldsymbol{b}] \neq [\boldsymbol{b}; \boldsymbol{a}]$). This allows the model to capture word order. Furthermore, the learnable parameters θ enable the model to learn complex compositional semantics from data, offering a significant advantage over the static, order-agnostic nature of averaging.

mjene, detaljno opisana u sljedećem poglavlju. Kroz ovaj proces istražujemo kako PyRv poboljšava kompozicionalnost vektorskih reprezentacija riječi, dok istovremeno održava prihvatljivu dimenzionalnost i poboljšava kvalitetu reprezentacije.

3.3 Matematička formulacija

Kako bismo formalizirali razliku između metode prosjeka i našeg predloženog PyRv+FT pristupa, definiramo njihove odgovarajuće kompozicijske funkcije. Neka je fraza P niz od n riječi, $P = (w_1, w_2, \ldots, w_n)$. Svaka riječ w_i preslikava se u d-dimenzionalnu vektorsku reprezentaciju $\boldsymbol{v}_i \in \mathbb{R}^d$ pomoću fastTexta. Cilj je sastaviti niz vektora $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ u jedinstvenu vektorsku reprezentaciju fraze $\boldsymbol{e}_P \in \mathbb{R}^d$.

Prosjek. Metoda prosjeka računa vektorsku reprezentaciju fraze $\boldsymbol{e}_P^{\text{avg}}$ uzimanjem srednje vrijednosti po elementima pripadajućih vektora riječi:

$$\boldsymbol{e}_P^{\text{avg}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{v}_i \tag{1}$$

Glavno ograničenje ovog modela vidljivo je iz njegove formulacije. Budući da je zbrajanje vektora komutativno ($\boldsymbol{v}_i + \boldsymbol{v}_j = \boldsymbol{v}_j + \boldsymbol{v}_i$), rezultirajuća reprezentacija $\boldsymbol{e}_P^{\text{avg}}$ je neosjetljiva na redoslijed riječi u frazi, čime se gube bitne sintaktičke i semantičke informacije.

Piramidalna rekurzivna kompozicija (PyRv+FT). PyRv+FT metoda konstruira vektorsku reprezentaciju fraze $\boldsymbol{e}_P^{\text{PyRv}}$ kroz hijerarhijski i rekurzivni proces. Kompozicija započinje na najnižoj razini (razina 0), koja se sastoji od početnih vektorskih reprezentacija riječi $(\boldsymbol{v}_1^{(0)}, \boldsymbol{v}_2^{(0)}, \ldots, \boldsymbol{v}_n^{(0)})$.

Na svakoj sljedećoj razini k, susjedni parovi vektora s razine $k-1$ kombiniraju se pomoću parametrizirane kompozicijske funkcije f_θ (model PyRvNN) kako bi se oblikovala nova reprezentacija više razine. Ovaj proces se ponavlja dok ne preostane samo jedan vektor. Novi vektor $\boldsymbol{v}_i^{(k)}$ na razini k generira se iz razine $k-1$ na sljedeći način:

$$\boldsymbol{v}_i^{(k)} = f_\theta(\boldsymbol{v}_i^{(k-1)}, \boldsymbol{v}_{i+1}^{(k-1)}) \tag{2}$$

Konačna vektorska reprezentacija fraze $\boldsymbol{e}_P^{\text{PyRv}}$ je jedinstveni vektor koji preostaje na vrhu piramide.

Za razliku od metode prosjeka, kompozicijska funkcija f_θ nije komutativna jer konkatenacija ovisi o redoslijedu (tj. $[\boldsymbol{a}; \boldsymbol{b}] \neq [\boldsymbol{b}; \boldsymbol{a}]$). To omogućuje modelu da zabilježi redoslijed riječi. Nadalje, učljivi parametri θ omogućuju modelu učenje složene kompozicijske semantike iz podataka, što nudi značajnu prednost u odnosu na statičnu prirodu prosjeka koja je neosjetljiva na redoslijed.

4 Evaluation

In this section, we describe the evaluation of fastText and PyRv+FT embeddings on two important NLP tasks: Universal Part-of-Speech tagging (UPOS) and Universal Dependency Relation labeling (DEPREL). UPOS tags represent core grammatical categories such as nouns, verbs, and adjectives, while DEPREL captures the syntactic relationships between words in a sentence, indicating dependencies like subjects, objects, and modifiers.

For this evaluation, we use the hr500k 2.0 dataset (Ljubešić et al., 2016), a Croatian corpus with labeled data for both UPOS and DEPREL tasks (amongst others). This dataset contains 901 texts, 24,763 sentences, and a total of 499,635 tokens.

Additionally, we perform a qualitative analysis of PyRv+FT embeddings by visualizing the representation space to gain deeper insights into its structure and characteristics.

4.1 Experiment setup

To assess the performance of different embedding methods on the downstream tasks of UPOS and DEPREL, we use a multi-layer perceptron (MLP) model with one hidden layer. The hidden layer consists of 1,000 neurons and uses the ReLU activation function. The input to the model is a 300-dimensional vector (the size of both fastText and PyRv+FT embeddings). The output layer, with softmax activation, adjusts to the number of classes in each task: 17 classes for UPOS and 37 classes for DEPREL. Each evaluation is conducted by training the MLP for one epoch.

The embedding procedure remains consistent across both UPOS and DEPREL tasks, differing only in the MLP output labels.

The method of embedding a word from a sentence depends on the embedding strategy employed:

- **fastText 1 word**: Embeds only the target word, ignoring its surrounding context.
- **mean fastText N words**: Represents the target word by embedding all N words (the target word and its N-1 neighboring context words) separately and averaging the embeddings to obtain a final representation.
- **PyRv+FT N words**: Embeds each word using fastText, but instead of averaging the N embeddings, it recursively combines them using the PyRvNN model to generate a single, unified embedding.

4.2 Quantitative results

UPOS. Part-of-speech tagging is a relatively simple task where the surrounding word context does not provide significant benefits for classification. We include UPOS evaluation primarily to demonstrate how aver-

4 Evaluacija

U ovom poglavlju opisujemo evaluaciju fastText i PyRv+FT vektorskih reprezentacija na dva bitna zadatka obrade prirodnog jezika: označavanje univerzalnih vrsta riječi (UPOS) i označavanje univerzalnih sintaktičkih odnosa (DEPREL). Oznake UPOS predstavljaju osnovne gramatičke kategorije poput imenica, glagola i pridjeva, dok DEPREL bilježi sintaktičke odnose među riječima u rečenici, poput subjekata, objekata i atributa.

Za ovu evaluaciju koristimo skup podataka hr500k 2.0 (Ljubešić i dr., 2016), hrvatski korpus s anotiranim podacima za UPOS i DEPREL zadatke (među ostalim). Skup podataka sadrži 901 tekst, 24.763 rečenice i ukupno 499.635 tokena.

Dodatno, provodimo kvalitativnu analizu PyRv+FT reprezentacija vizualizacijom njihovog reprezentacijskog prostora, kako bismo dobili dublji uvid u njegovu strukturu i obilježja.

4.1 Postavke eksperimenta

Za procjenu performansi različitih metoda vektorskog reprezentiranja na zadacima UPOS i DEPREL koristimo višeslojni perceptron (MLP) model s jednom skrivenom slojem. Skriveni sloj sadrži 1.000 neurona i koristi ReLU aktivacijsku funkciju. Ulaz u model je 300-dimenzionalni vektor (što odgovara dimenziji fastText i PyRv+FT vektora). Izlazni sloj koristi softmax aktivaciju i prilagođava se broju klasa svakog zadatka: 17 za UPOS i 37 za DEPREL. Svaka evaluacija provodi se treniranjem modela MLP kroz jednu epohu.

Postupak vektorizacije ostaje jednak za UPOS i DEPREL, a razlikuju se samo oznake izlaznih klasa.

Način vektorizacije riječi iz rečenice ovisi o korištenoj strategiji:

- **fastText 1 riječ**: Vektorizira samo ciljnu riječ, zanemarujući njezin kontekst.
- **prosjek fastText N riječi**: Reprezentira ciljnu riječ tako da se zasebno vektoriziraju sve N riječi (ciljna i N-1 susjedne) i zatim izračuna prosjek.
- **PyRv+FT N riječi**: Svaka riječ se vektorizira fastTextom, ali se umjesto prosjeka tih N vektora koristi model PyRvNN koji ih rekurzivno kombinira u jednu objedinjenu reprezentaciju.

4.2 Kvantitativni rezultati

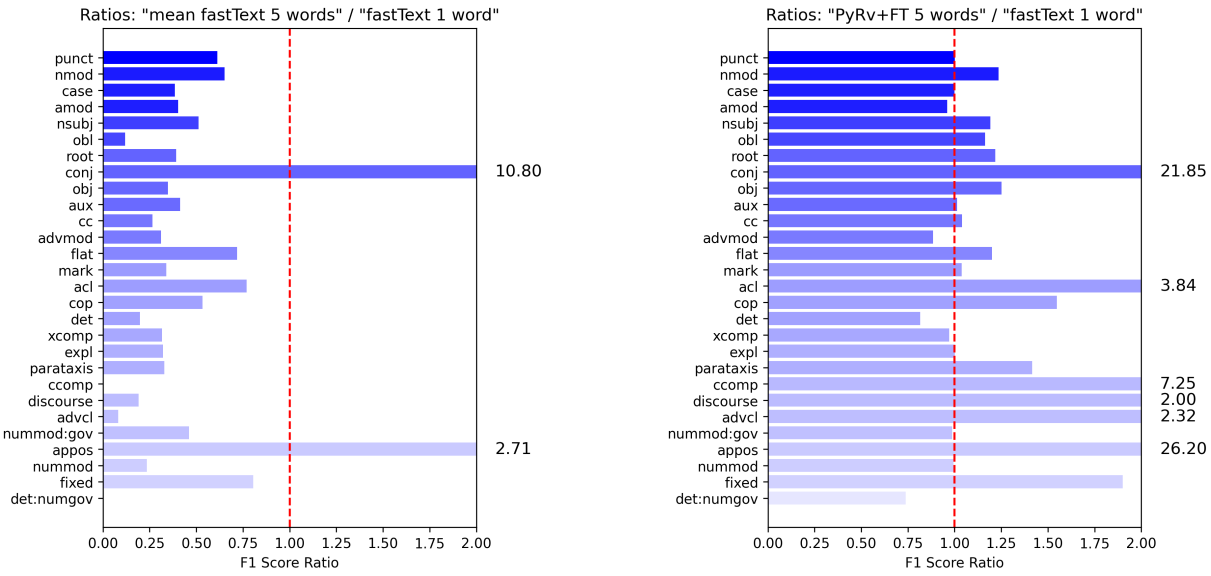
UPOS. Označavanje vrsta riječi relativno je jednostavan zadatak u kojem kontekst susjednih riječi ne donosi velike prednosti. UPOS evaluaciju uključujemo kako bismo pokazali da prosječna fastText reprezenta-

Table 1. UPOS results, Macro and Weighted averages. **Tablica 1**. Rezultati UPOS-a, makro i ponderirani prosjeci.

	Accuracy	**Precision**		**Recall**		**F1 score**	
		M. avg	W. avg	M. avg	W. avg	M. avg	W. avg
fastText 1 word	**0.95**	**0.91**	**0.95**	**0.89**	**0.95**	0.89	**0.95**
mean fastText 3 words	0.61	0.57	0.61	0.59	0.61	0.57	0.61
PyRv+FT 3 words	0.93	0.9	0.93	**0.89**	0.93	**0.9**	0.93

Table 2. DEPREL results, Macro and Weighted averages. **Tablica 2**. Rezultati DEPREL-a, makro i ponderirani prosjeci.

	Accuracy	**Precision**		**Recall**		**F1 score**	
		M. avg	W. avg	M. avg	W. avg	M. avg	W. avg
fastText 1 word	0.71	0.52	0.68	0.48	0.71	0.47	0.68
mean fastText 5 words	0.34	0.25	0.34	0.19	0.34	0.19	0.31
PyRv+FT 5 words	**0.77**	**0.58**	**0.77**	**0.55**	**0.77**	**0.56**	**0.76**



(a) "mean fastText 5 words" F1 scores divided by "fastText 1 word" F1 scores.
(a) Omjer F1 rezultata za "prosjek fastText 5 riječi" u odnosu na "fastText 1 riječ".

(b) "PyRv+FT 5 words" F1 scores divided by "fastText 1 word" F1 scores.
(b) Omjer F1 rezultata za "PyRv+FT 5 riječi" u odnosu na "fastText 1 riječ".

Figure 3. DEPREL relative F1 score ratios (by class) comparing different composition methods. The left plot (a) shows the ratio of F1 scores for "mean fastText 5 words" versus "fastText 1 word", while the right plot (b) compares "PyRv+FT 5 words" versus "fastText 1 word". Each bar represents a class, and the length of the bar indicates the relative performance of the model. Classes are ordered by support value in the test set (larger on the top).
Slika 3. Relativni F1 omjeri (po klasi) za DEPREL, uspoređujući metode kompozicije. Lijevi graf (a) prikazuje omjer F1 rezultata za "prosjek fastText 5 riječi" prema "fastText 1 riječ", dok desni graf (b) prikazuje odnos za "PyRv+FT 5 riječi". Svaka traka predstavlja klasu, a njezina duljina pokazuje relativne performanse. Klase su poredane po zastupljenosti u testnom skupu (najzastupljenije na vrhu).

aging fastText word embeddings can degrade downstream performance, while combining fastText word embeddings using PyRvNN preserves much of the embedding quality.

When using fastText to embed a single word without context, 95% accuracy can be achieved. Mean pooling can degrade performance, while PyRvNN composition preserves most of the quality.

cija može narušiti rezultate, dok kombiniranje fastText vektora putem PyRvNN-a zadržava većinu kvalitete reprezentacije.

Kada se koristi fastText za vektorizaciju samo jedne riječi bez konteksta, postiže se točnost od 95%. Me-

out considering its context, we achieve an accuracy of 95%. However, averaging fastText embeddings over three words leads to a substantial drop in performance, with accuracy falling to 61%. In contrast, when combining fastText embeddings for three words using PyRvNN, the performance degradation is minimized, yielding an accuracy of 93%. These results highlight how PyRvNN can effectively mitigate the loss of information that occurs when averaging word embeddings. Detailed results are shown in Table 1.

DEPREL. The dependency relation task is more complex than UPOS, as it requires understanding the syntactic relationships between words. In this case, enriching word embeddings with surrounding context can significantly improve classification performance.

When embedding a single word using fastText (without its context), the model achieves an accuracy of 71%. However, averaging five fastText word embeddings, including the target word and its four neighbors, results in a sharp decline in performance, with accuracy dropping to 34%. This reduction in accuracy reflects how averaging word embeddings leads to the loss of important information, including word order and syntactic structure. By contrast, composing five fastText word embeddings using PyRvNN boosts accuracy to 77%, demonstrating the method's ability to capture more nuanced relationships between words.

Evaluation results are presented in Table 2 with more detailed results (by class) available in Tables 5, 6, and 7 in the Appendix.

Figure 3 presents bar plots comparing F1 scores between two composition methods, mean averaging and PyRv, relative to single word fastText embeddings' performance. The left plot 3a illustrates the F1 score ratio between "mean fastText 5 words" and "fastText 1 word", while the right plot 3b contrasts "PyRv+FT 5 words" with "fastText 1 word".

In the following analysis, we focus on three notable classes: punctuation, conjuncts, and adnominal clauses. These were selected because punctuation classification does not rely on context, classifying conjuncts without context is nearly impossible, and for adnominal clauses, context is helpful but averaging tends to degrade performance.

Punctuation (punct) refers to punctuation marks such as ".", "?", "!", and ",". Since punctuation is straightforward to classify without context, a single fastText embedding for the target word alone achieves a perfect F1 score of 1. Averaging five fastText embeddings (the target word and its four neighboring words) significantly degrades performance, reducing the F1 score to 0.61. However, using PyRv to compose these embeddings preserves the high performance, maintaining an F1 score of 1.

Conjunct (conj) denotes a relation between elements connected by coordinating conjunctions like "and," "or," or ",". In coordinate structures, the first element is conventionally treated as the head, with subsequent el-

đutim, prosjek triju fastText vektora dovodi do značajnog pada performansi – točnost pada na 61%. Suprotno tome, kada se isti vektori kombiniraju koristeći PyRvNN, pad performansi je znatno manji – točnost iznosi 93%. Ovi rezultati pokazuju kako PyRvNN učinkovito ublažava gubitak informacija do kojeg dolazi kada se računa prosjek. Detaljni rezultati prikazani su u Tablici 1.

DEPREL. Zadatak označavanja sintaktičkih odnosa među riječima u rečenici znatno je složeniji jer zahtijeva razumijevanje odnosa među riječima. U ovom slučaju, obogaćivanje vektora kontekstom može značajno unaprijediti klasifikaciju.

Korištenjem fastText-a za jednu riječ bez konteksta postiže se točnost od 71%. No, kada se koristi prosjek pet fastText vektora (ciljna riječ i četiri susjedne), točnost drastično pada na 34%. Taj pad ukazuje na to da se prosjekom gubi važne informacije – poput redoslijeda riječi i sintaktičke strukture. S druge strane, PyRv kompozicija pet vektora rezultira točnošću od 77%, što potvrđuje da metoda bolje zahvaća dublje odnose među riječima.

Rezultati evaluacije prikazani su u Tablici 2, dok su detaljni rezultati po klasama dostupni u Tablicama 5, 6 i 7 u Dodatku.

Na Slici 3 prikazani su grafikoni s usporedbama F1 rezultata između metoda prosjeka i PyRv kompozicije u odnosu na fastText za jednu riječ. Lijevi prikaz 3a prikazuje omjer za "prosjek fastText 5 riječi", a desni 3b za "PyRv+FT 5 riječi".

U sljedećoj analizi fokusiramo se na tri istaknute klase: interpunkciju, koordinirane veznike i odnosne surečenice. Odabrane su jer klasifikacija interpunkcije ne ovisi o kontekstu, klasificiranje koordiniranih veznika bez konteksta gotovo je nemoguće, dok je za odnosne surečenice kontekst koristan, ali prosječno kombiniranje vektora narušava performanse.

Interpunkcija (punct) odnosi se na interpunkcijske znakove poput ".", "?", "!" i ",". Budući da je interpunkciju jednostavno klasificirati bez konteksta, korištenje fastText reprezentacije jedne riječi postiže savršen F1 rezultat od 1. Međutim, prosječno kombiniranje pet fastText reprezentacija (ciljna riječ i četiri susjedne) znatno narušava performanse, smanjujući F1 rezultat na 0,61. S druge strane, kompozicija ovih reprezentacija pomoću metode PyRv zadržava visoku točnost, održavajući F1 rezultat na 1.

Konjunktor (conj) označava relaciju između elemenata povezanih koordinacijskim veznicima poput "i", "ili" ili ",". U koordinacijskim strukturama, prvi element konvencionalno se tretira kao glava, a ostali se

ements connected through the conj relation. For example, in the sentence "Bill is *big* and *honest*," the word "honest" is labeled as conj (connected to "big"). Similarly, in "He *came* home, *took* a shower and immediately *went* to bed," the words "took" and "went" are both labeled as conj (connected to "came"). Classifying conjuncts accurately requires contextual information. A single fastText word embedding, without any context, yields a poor F1 score of 0.03. Averaging the embeddings of the target word and its four neighbors improves performance significantly, achieving an F1 score of 0.32 (a 10.8x improvement). Composing these embeddings using PyRv further boosts performance, reaching an F1 score of 0.64 (a 21.85x improvement over the single word embedding).

Adnominal clause (acl) refers to finite or non-finite clauses that modify a nominal. For instance, in "the *issues* as he *sees* them," the word "sees" is labeled as acl (connected to "issues"). In "There are many online *sites offering* booking facilities," the word "offering" is labeled as acl (connected to "sites"). Using a single fastText word embedding results in an F1 score of 0.14. Averaging the target word's embedding with its four neighbors actually degrades performance slightly, reducing the F1 score to 0.11. In contrast, composing these word embeddings with PyRv substantially improves performance, raising the F1 score to 0.53 (a 3.84x improvement over the single word embedding).

4.3 Qualitative analysis

To gain qualitative insights into the structure of PyRv+FT embeddings, we visualize the representation space. A portion of this space is shown in Figures 4 and 5 (in the Appendix). In these visualizations, each node represents a phrase consisting of two or three words. A two-word phrase is connected to a three-word phrase if the shorter phrase is part of the longer one.

We highlight specific clusters within the visualized space, with detailed examples of phrases from these clusters (translated to English) provided in Tables 3 and 4. The areas circled in the figures contain phrases built around the prepositions "u" (Croatian for "in") and "na" (Croatian for "on"). For example, Area C contains two-word phrases like "primjena **na**" (eng. "application on"), while Area A includes phrases such as "**na** svijet" (eng. "on the world"). Similarly, Area B features three-word phrases like "primjena **na** svijet" (eng. "application on the world"), and Area D includes phrases like "**na** svijet oko" (eng. "on the world around"). Same holds for the preposition "u".

The organization of phrases in the representation space is not random: phrases with similar syntactic structures (e.g., where the preposition appears at the beginning, middle, or end of the phrase) tend to cluster together. Furthermore, within these broader areas, smaller sub-clusters form based on the specific prepo-

povezuju conj relacijom. Na primjer, u rečenici "Bill je *velik* i *pošten*", riječ "pošten" označena je kao conj (povezana s "velik"). Slično, u "On je *došao* kući, *istuširao* se i odmah *otišao* na spavanje," riječi "istuširao" i "otišao" označene su kao conj (povezane s "došao"). Točna klasifikacija conj relacije zahtijeva kontekstualne informacije. FastText reprezentacija jedne riječi bez konteksta daje vrlo loš F1 rezultat od 0,03. Prosječno kombiniranje reprezentacija ciljne riječi i četiri susjeda znatno poboljšava performanse, postižući F1 rezultat od 0,32 (10,8 puta bolje). Kompozicija istih reprezentacija pomoću metode PyRv dodatno povećava točnost na 0,64 (21,85 puta bolje u odnosu na pojedinačnu riječ).

Atributna rečenica (acl) odnosi se na finitne ili nefinitne zavisne rečenice koje pobliže određuju ili opisuju imensku riječ. Na primjer, u rečenici "*problemi* kako ih on *vidi*", riječ "vidi" označena je kao acl (povezana s "problemi"). U rečenici "Postoji mnogo internetskih *stranica* koje *nude* mogućnost rezervacije," riječ "nude" označena je kao acl (povezana sa "stranica"). Korištenje fastText reprezentacije jedne riječi rezultira F1 rezultatom od 0,14. Prosječno kombiniranje s četiri susjeda zapravo dodatno narušava performanse, smanjujući F1 rezultat na 0,11. Nasuprot tome, kompozicija tih reprezentacija pomoću metode PyRv značajno poboljšava performanse, povećavajući F1 rezultat na 0,53 (3,84 puta bolje u odnosu na pojedinačnu riječ).

4.3 Kvalitativna analiza

Kako bismo dobili kvalitativni uvid u strukturu PyRv+FT reprezentacija, vizualiziramo njihov reprezentacijski prostor. Dio tog prostora prikazan je na Slikama 4 i 5 (u Dodatku). U tim vizualizacijama svaki čvor predstavlja frazu od dvije ili tri riječi. Fraza od dvije riječi povezana je s frazom od tri riječi ako je kraća fraza podskup duže.

Ističemo određene klastere unutar prikazanog prostora, s detaljnim primjerima fraza iz tih klastera (prevedenima na engleski) prikazanima u Tablicama 3 i 4. Područja zaokružena na slikama sadrže fraze izgrađene oko prijedloga "u" (hrvatski za "in") i "na" (hrvatski za "on"). Na primjer, Područje C sadrži dvočlane fraze poput "primjena **na**" (engl. "application on"), dok Područje A uključuje izraze poput "**na** svijet" (engl. "on the world"). Slično tome, Područje B sadrži tročlane fraze poput "primjena **na** svijet" (engl. "application on the world"), a Područje D uključuje fraze poput "**na** svijet oko" (engl. "on the world around"). Isti obrazac vrijedi i za fraze s prijedlogom "u".

Organizacija fraza u reprezentacijskom prostoru nije nasumična: fraze sa sličnom sintaktičkom strukturom (npr. prijedlog na početku, u sredini ili na kraju fraze) sklone su grupiranju. Nadalje, unutar šireg područja pojavljuju se manji podklasteri koji se dodatno organiziraju prema tome sadrži li fraza prijedlog "u" ili "na".

sition ("u" or "na") present in the phrase.

Table 3. Phrases by areas (A, B, and C) in the PyRv+FT representation space, translated to English (some phrases are longer when translated).

Preposition "on" (cro. "na")		
Area C	**Area B**	**Area A**
application on	application on the world	on the world
are on	are on local	on local
assistant on	assistant on the subject	on the subject
media on	media on protest	on the protest
vat on	vat on tickets	on tickets
based on	based on data	on data
finance on	finance on revenues	on revenues
os on	os on which	on which
dollars on	dollars on google	on google
found on	found on the third	on the third
relation on	relation on the past	on the past
Preposition "in" (cro. "u")		
Area C	**Area B**	**Area A**
enthusiast in	enthusiast in the river	in the river
currently in	currently in testing	in testing
released in	released in circulation	in circulation
circulation in	circulation in June	in June
by activity in	by activity in teaching	in teaching
work in	work in the wood	in the wood
tickets in	tickets in europe	in europe
musician in	musician in croatia	in croatia
only in	only in the past	in the past
drop in	drop in the sea	in the sea
is in	is in the past	in the past
year in	year in croatia	in croatia
percent in	percent in relation	in relation
and in	and in the average	in the average
. in	. in this	in this

Tablica 3. Izrazi po područjima (A, B i C) u PyRv+FT prostoru reprezentacije (izvorni hrvatski izrazi).

Prijedlog "na" (engl. "on")		
Područje C	**Područje B**	**Područje A**
primjena na	primjena na svijet	na svijet
su na	su na lokalnim	na lokalnim
asistentent na	asistentent na predmetu	na predmetu
medije na	medije na prosvjed	na prosvjed
pdv-a na	pdv-a na ulaznice	na ulaznice
baziranim na	baziranim na podacima	na podacima
financija na	financija na prihodima	na prihodima
os na	os na koji	na koji
dolara na	dolara na google	na google
nalazi na	nalazi na trećoj	na trećoj
odnosu na	odnosu na prošlu	na prošlu
Prijedlog "u" (engl. "in")		
Područje C	**Područje B**	**Područje A**
entuzijasta u	entuzijasta u rijeci	u rijeci
trenutno u	trenutno u testiranju	u testiranju
puštena u	puštena u opticaj	u opticaj
opticaj u	opticaj u lipnju	u lipnju
aktivnošću u	aktivnošću u nastavi	u nastavi
rada u	rada u drvnoj	u drvnoj
ulaznice u	ulaznice u europi	u europi
glazbenika u	glazbenika u hrvatskoj	u hrvatskoj
samo u	samo u protekle	u protekle
kap u	kap u moru	u moru
se u	se u protekloj	u protekloj
godini u	godini u hrvatskoj	u hrvatskoj
posto u	posto u odnosu	u odnosu
i u	i u prosjeku	u prosjeku
. u	. u ovoj	u ovoj

Table 4. Phrases in area D in the PyRv+FT representation space, translated to English (some phrases are longer when translated).

Area D	
Preposition "on" (cro. "na")	**Preposition "in" (cro. "u")**
on the world around	in the river which
on local elections	in testing and
on the subject of organization	in circulation in
on the protest of musicians	in june 2009
on tickets for	in teaching 1982
on the data collected	in the wood industry
on budget revenues	in europe .
on which this	in croatia only
on google play	in the past two
on the third position	in the sea of state
on the past year	in the past year
	in croatia is not
	in relation to
	in the average spends
	in this praise

Tablica 4. Izrazi u području D u PyRv+FT prostoru reprezentacije (izvorni hrvatski izrazi).

Područje D	
Prijedlog "na" (engl. "on")	**Prijedlog "u" (engl. "in")**
na svijet oko	u rijeci koji
na lokalnim izborima	u testiranju i
na predmetu organizacije	u opticaj u
na prosvjed glazbanika	u lipnju 2009.
na ulaznice u	u nastavi 1982.
na podacima prikupljenim	u drvnoj industriji
na prihodima proračuna	u europi .
na koji ova	u hrvatskoj samo
na google play	u protekle dvije
na trećoj poziciji	u moru državnog
na prošlu godinu	u protekloj godini
	u hrvatskoj nije
	u odnosu na
	u prosjeku troši
	u ovoj hvale

5 Conclusion

In this paper, we explored the use of Pyramidal Recursive learning (PyRv) for the composition of word embeddings and evaluated its ability to generate meaningful representations of multi-word units. Our findings show that PyRv outperforms simple averaging methods in embedding composition.

In the part-of-speech tagging task, where word context is less crucial, single word embeddings achieve an accuracy of 95%. Averaging 3-word context embeddings reduces this to 61% due to the loss of word order information, while PyRv retains a high accuracy of 93% by effectively preserving word order. In the more complex task of dependency relation labeling, where single word embeddings reach 71% accuracy, averaging embeddings for 5-word contexts results in a sharp decline to 34%. In contrast, composing context words with PyRv attains a significantly higher accuracy of 77%, demonstrating its superior capability in integrating multiple word embeddings into cohesive and expressive representations. This validates the effectiveness of the compositionality property of PyRv in real-world tasks.

The primary contribution of this work is the introduction of a method for composing multi-word units into unified representations using PyRv. We provided an evaluation of its effectiveness compared to averaging word embeddings and validated its compositionality property. Additionally, we explored the structure of the representation space generated by PyRv's compositional approach. By training the PyRv model on Croatian texts, we demonstrated its flexibility and potential for application across diverse languages.

Future work could further probe PyRv's effectiveness as a compositional model. A natural next step would be its evaluation on standard English benchmarks for tasks highly sensitive to semantic composition. This would also enable a direct comparison against other composition strategies, particularly those employed by state-of-the-art transformer-based models. Investigating architectural variations, such as enhancing the recursive function with attention mechanisms, also presents a promising research avenue.

Acknowledgments

This work has been fully supported by the University of Rijeka project uniri-iskusni-drustv-23-95.

5 Zaključak

U ovom radu istražili smo primjenu piramidalnog rekurzivnog učenja (PyRv) za kompoziciju vektorskih prikaza riječi te smo evaluirali njegovu sposobnost generiranja smislenih reprezentacija višerječnih izraza. Naši rezultati pokazuju da PyRv nadmašuje jednostavne metode prosjeka u zadacima kompozicije vektora.

U zadatku označavanja vrsta riječi, gdje kontekst ima manju ulogu, pojedinačni vektori riječi postižu točnost od 95%. Vektor dobiven metodom prosjeka konteksta sa 3 riječi značajno smanjuje točnost na 61% zbog gubitka informacije o redoslijedu riječi, dok PyRv zadržava visoku točnost od 93% zahvaljujući učinkovitom očuvanju strukture rečenice. U složenijem zadatku označavanja sintaktičkih odnosa, gdje pojedinačni vektori postižu točnost od 71%, prosjek 5 riječi dovodi do naglog pada na 34%. S druge strane, PyRv kompozicija podiže točnost na 77%, što potvrđuje njegovu sposobnost integracije konteksta u koherentne i informativne reprezentacije. Time se dokazuje učinkovitost kompozicionalnog pristupa koji PyRv omogućuje u stvarnim NLP zadacima.

Glavni doprinos ovog rada je predstavljanje metode za kompoziciju višerječnih izraza u objedinjene vektorske prikaze pomoću PyRv-a. Prikazali smo evaluaciju njegove učinkovitosti u usporedbi s metodom prosjeka vektora riječi i potvrdili njegovo kompozicionalno svojstvo. Također smo analizirali strukturu reprezentacijskog prostora koju PyRv generira. Trenirajući model na hrvatskim tekstovima, pokazali smo njegovu prilagodljivost i potencijalnu primjenu na raznim jezicima.

U budućem radu mogla bi se dalje istraživati učinkovitost metode PyRv kao kompozicijskog modela. Prirodan sljedeći korak bila bi njezina evaluacija na standardnim engleskim benchmarkovima za zadatke koji su izrazito osjetljivi na semantičku kompoziciju. To bi također omogućilo izravnu usporedbu s drugim kompozicijskim strategijama, posebice onima koje rabe najsuvremeniji modeli temeljeni na transformerima. Istraživanje arhitektonskih varijacija, poput obogaćivanja rekurzivne funkcije mehanizmima pažnje, također predstavlja obećavajući pravac istraživanja.

Zahvala

Ovaj rad je u potpunosti financiran iz projekta Sveučilišta u Rijeci uniri-iskusni-drustv-23-95.

References / Literatura

Arora, S., Liang, Y., & Ma, T. (2017). A simple but tough-to-beat baseline for sentence embeddings. *International conference on learning representations*.

Babić, K., Guerra, F., Martinčić-Ipšić, S., & Meštrović, A. (2020). A comparison of approaches for measuring the semantic similarity of short texts based on word embeddings. *Journal of information and organizational sciences*, *44*(2), 231–246.

Babić, K., & Meštrović, A. (2024). Recursively autoregressive autoencoder for pyramidal text representation. *IEEE access*, *12*, 71361–71370.

Bahdanau, D. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, *5*, 135–146.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805*.

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

Ljubešić, N., Klubička, F., Agić, Ž., & Jazbec, I.-P. (2016). New inflectional lexicons and training corpora for improved morphosyntactic annotation of croatian and serbian. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 4264–4270.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv:1301.3781*.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the 2013 conference on empirical methods in natural language processing*, 1631–1642.

Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

Wieting, J., Bansal, M., Gimpel, K., & Livescu, K. (2015). Towards universal paraphrastic sentence embeddings. *arXiv preprint arXiv:1511.08198*.

Zhao, H., Lu, Z., & Poupart, P. (2015). Self-adaptive hierarchical sentence model. *arXiv preprint arXiv:1504.05070*.

A Appendix / Dodatak

Table 5. DEPREL evaluation results using the fastText embedding method (single-word embeddings), where P stands for Precision, R for Recall, F1 for F1 score, and S for Support.
Tablica 5. Rezultati DEPREL evaluacije korištenjem fastText metode reprezentiranja (jednorječne vektorske reprezentacije), pri čemu P označava preciznost, R odziv, F1 F1-mjeru, a S potporu.

Class	P	R	F1	S
punct	1	1	1	3037
nmod	0.6	0.58	0.59	2437
case	0.96	0.98	0.97	2364
amod	0.79	0.93	0.85	2355
nsubj	0.53	0.66	0.59	1725
obl	0.48	0.57	0.52	1607
root	0.45	0.67	0.53	1136
conj	0.19	0.02	0.03	1134
obj	0.49	0.44	0.46	1072
aux	0.75	0.96	0.84	1037
cc	0.83	0.97	0.89	887
advmod	0.8	0.88	0.84	825
flat	0.54	0.69	0.61	689
mark	0.85	0.91	0.88	471
acl	0.44	0.08	0.14	452
cop	0.58	0.22	0.32	415
det	0.87	0.87	0.87	400
xcomp	0.61	0.7	0.65	350
expl	0.85	1	0.92	302
parataxis	0.63	0.38	0.47	300
ccomp	0.18	0.02	0.03	230
discourse	0.48	0.21	0.29	208
advcl	0.29	0.08	0.12	198
nummod:gov	0.72	0.9	0.8	187
appos	1	0.01	0.02	130
nummod	0.82	0.63	0.71	117
fixed	0.34	0.33	0.34	100
csubj	0	0	0	40
det:numgov	0.73	0.66	0.69	29
orphan	0	0	0	13
advmod:emph	0	0	0	5
flat:foreign	0	0	0	4
vocative	0	0	0	3
compound	0	0	0	1
macro avg	0.52	0.48	0.47	
weighted avg	0.68	0.71	0.68	

Table 6. DEPREL evaluation results using the fastText embedding method (averaging embeddings of five words), where P stands for Precision, R for Recall, F1 for F1 score, and S for Support.
Tablica 6. Rezultati DEPREL evaluacije korištenjem fastText metode reprezentiranja (prosjek vektorskih reprezentacija pet riječi), pri čemu P označava preciznost, R odziv, F1 F1-mjeru, a S potporu.

Class	P	R	F1	S
punct	0.54	0.71	0.61	3037
nmod	0.42	0.35	0.39	2437
case	0.28	0.56	0.37	2364
amod	0.3	0.4	0.34	2355
nsubj	0.3	0.3	0.3	1725
obl	0.29	0.03	0.06	1607
root	0.26	0.17	0.21	1136
conj	0.27	0.38	0.32	1134
obj	0.26	0.12	0.16	1072
aux	0.35	0.34	0.35	1037
cc	0.22	0.25	0.24	887
advmod	0.29	0.24	0.26	825
flat	0.44	0.43	0.44	689
mark	0.33	0.27	0.3	471
acl	0.23	0.07	0.11	452
cop	0.22	0.14	0.17	415
det	0.25	0.13	0.17	400
xcomp	0.35	0.15	0.21	350
expl	0.26	0.35	0.3	302
parataxis	0.76	0.09	0.16	300
ccomp	0	0	0	230
discourse	0.67	0.03	0.06	208
advcl	0.14	0.01	0.01	198
nummod:gov	0.27	0.6	0.37	187
appos	0.2	0.02	0.04	130
nummod	0.24	0.13	0.17	117
fixed	0.31	0.24	0.27	100
csubj	0	0	0	40
det:numgov	0	0	0	29
orphan	0	0	0	13
advmod:emph	0	0	0	5
flat:foreign	0	0	0	4
vocative	0	0	0	3
compound	0	0	0	1
macro avg	0.25	0.19	0.19	
weighted avg	0.34	0.34	0.31	

Table 7. DEPREL evaluation results using the PyRv+FT embedding method (composing embeddings of five words), where P stands for Precision, R for Recall, F1 for F1 score, and S for Support.
Tablica 7. Rezultati DEPREL evaluacije korištenjem PyRv+FT metode reprezentiranja (kompozicija vektorskih reprezentacija pet riječi), pri čemu P označava preciznost, R odziv, F1 F1-mjeru, a S potporu.

Class	**P**	**R**	**F1**	**S**
punct	1	1	1	3037
nmod	0.74	0.71	0.73	2437
case	0.98	0.97	0.97	2364
amod	0.81	0.82	0.82	2355
nsubj	0.7	0.7	0.7	1725
obl	0.59	0.63	0.61	1607
root	0.61	0.69	0.65	1136
conj	0.66	0.62	0.64	1134
obj	0.52	0.65	0.58	1072
aux	0.78	0.94	0.85	1037
cc	0.91	0.95	0.93	887
advmod	0.68	0.82	0.74	825
flat	0.77	0.69	0.73	689
mark	0.91	0.91	0.91	471
acl	0.6	0.47	0.53	452
cop	0.69	0.39	0.5	415
det	0.72	0.69	0.71	400
xcomp	0.67	0.59	0.63	350
expl	0.86	0.99	0.92	302
parataxis	0.83	0.56	0.67	300
ccomp	0.41	0.16	0.23	230
discourse	0.65	0.52	0.58	208
advcl	0.38	0.22	0.28	198
nummod:gov	0.82	0.76	0.79	187
appos	0.42	0.38	0.4	130
nummod	0.67	0.76	0.71	117
fixed	0.67	0.61	0.64	100
csubj	0	0	0	40
det:numgov	0.5	0.52	0.51	29
orphan	0	0	0	13
advmod:emph	0	0	0	5
flat:foreign	0	0	0	4
vocative	0	0	0	3
compound	0	0	0	1
macro avg	0.58	0.55	0.56	
weighted avg	0.77	0.77	0.76	

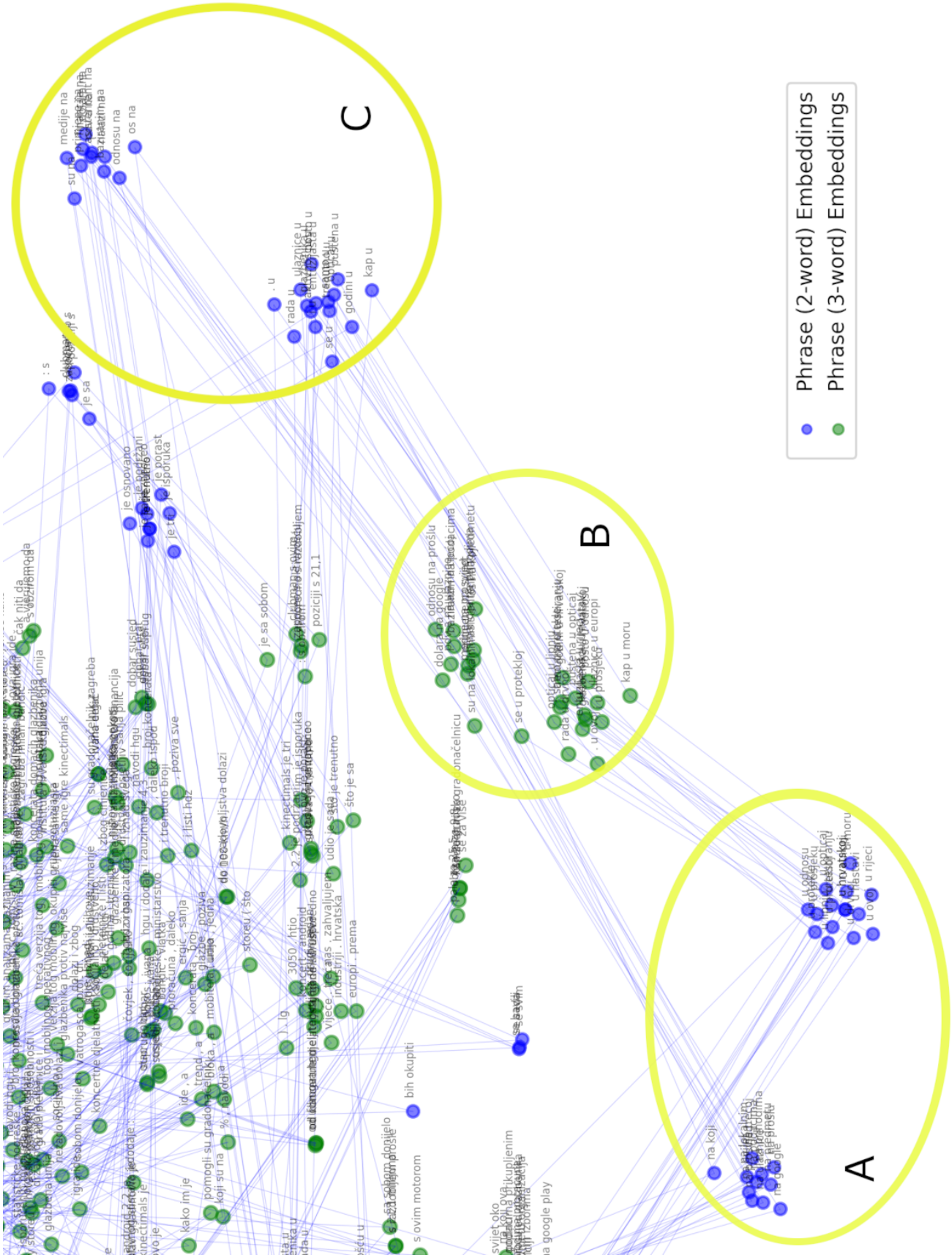


Figure 4: Visualization of the PyRv+FT representation space (reduced from 300 dimensions using t-SNE). Highlighted areas A, B, and C contain phrases with the prepositions 'na' (eng. 'on') and 'u' (eng. 'in'). Tables 3 and 4 (translated: 3 and 4) provide detailed examples of these phrases and their connections within the space.
Slika 4: Vizualizacija PyRv+FT reprezentacijskog prostora (reduciranog s 300 dimenzija korištenjem t-SNE-a). Istaknuta područja A, B i C sadrže fraze s prijedlozima 'na' (engl. 'on') i 'u' (engl. 'in'). Tablice 3 i 4 (prevedeno: 3 i 4) donose detaljne primjere ovih fraza i njihovih veza unutar prostora.

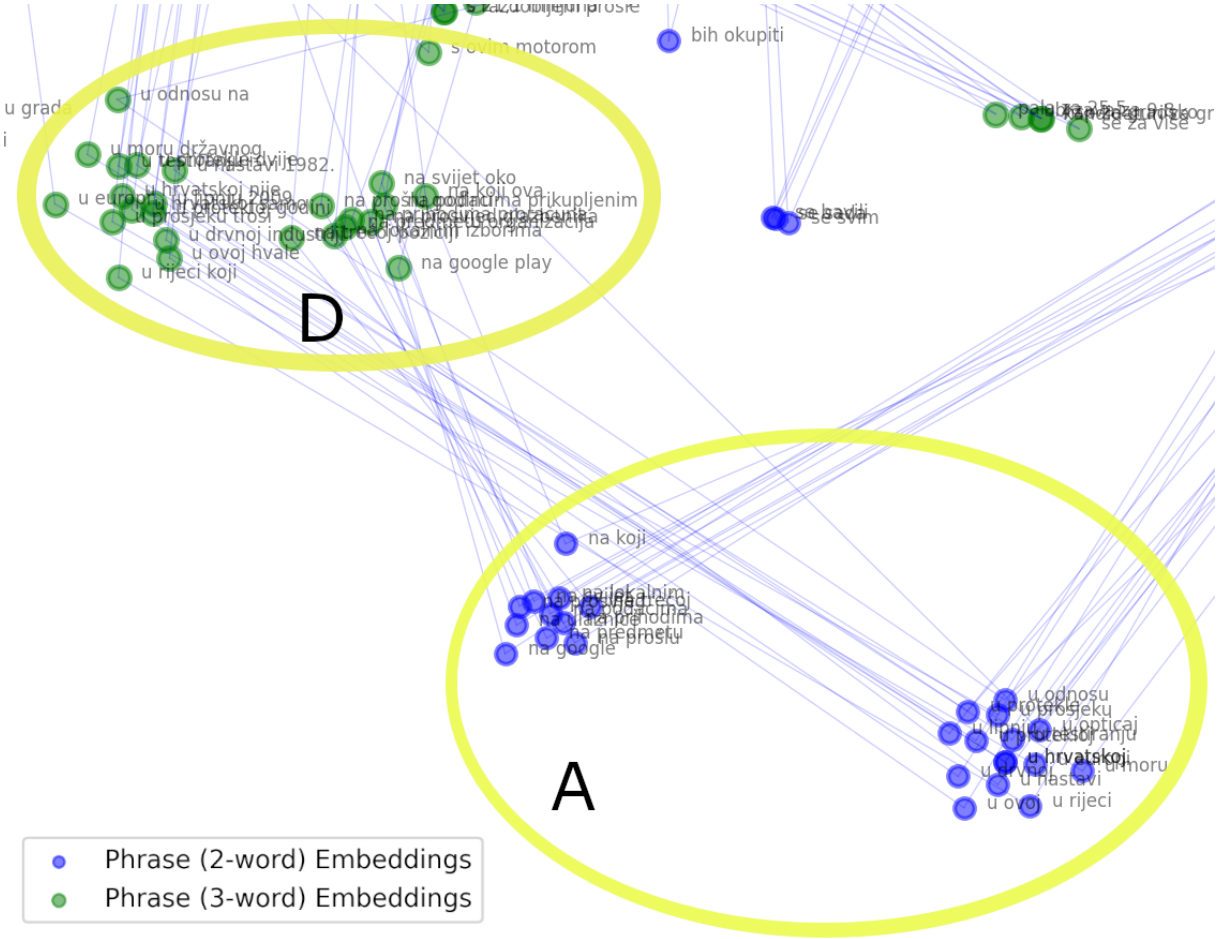


Figure 5. Visualization of the PyRv+FT representation space (reduced from 300 dimensions using t-SNE). Highlighted areas A and D contain phrases structured around the prepositions 'na' (eng. 'on') and 'u' (eng. 'in'). Tables 3 and 4 (translated: 3 and 4) present examples of these phrases and their relationships in the embedding space.
Slika 5. Vizualizacija PyRv+FT reprezentacijskog prostora (reduciranog s 300 dimenzija korištenjem t-SNE-a). Istaknuta područja A i D sadrže fraze strukturirane oko prijedloga 'na' (engl. 'on') i 'u' (engl. 'in'). Tablice 3 i 4 (prevedeno: 3 i 4) donose primjere ovih fraza i njihovih odnosa u vektorskom prostoru.

Detecting Discrimination Using Selected Data Mining Methods

Otkrivanje diskriminacije pomoću odabranih postupaka dubinske analize podataka

Petra Kuljiš, Lucia Načinović Prskalo, Marija Brkić Bakarić
University of Rijeka
Faculty of Informatics and Digital Technologies
Radmile Matejčić 2, 51000 Rijeka, Croatia
petrakuljis@gmail.com, {lnacinovic, mbrkic}@inf.uniri.com

Abstract. *The study uses advanced data analysis techniques to better understand patterns and manifestations of gender discrimination and to provide a basis for the development of effective policies aimed at reducing it. A publicly available dataset is analysed using selected data mining methods. Descriptive analysis provided deep insights into the basic characteristics of the data, and clustering was used to identify countries with similar patterns of gender inequality. Regression analysis was used to identify the factors contributing to the increase or decrease in gender inequality. The Isolation Forest model identified countries that represent extreme cases, while the comparative analysis revealed Croatia's advantages compared to the global average as well as areas for improvement.*

Keywords. data mining, gender discrimination, gender inequality, descriptive analysis, regression analysis, clustering, anomaly detection

Sažetak. *U radu su primijenjene napredne tehnike analize podataka s ciljem boljeg razumijevanja uzoraka i manifestacija rodne diskriminacije te stvaranja temelja za oblikovanje učinkovitih politika usmjerenih k smanjenju iste. Primjenom odabranih tehnika dubinske analize podataka analizira se javno dostupan podatkovni skup. Deskriptivnom analizom dobiveni su duboki uvidi u osnovne karakteristike podataka, a grupiranjem su identificirane zemlje sa sličnim obrascima rodne nejednakosti. Regresijska analiza korištena je za utvrđivanje čimbenika koji doprinose povećanju ili smanjenju rodne nejednakosti. Modelom izolacijskih šuma prepoznate su zemlje koje predstavljaju ekstreme, dok su komparativnom analizom utvrđene prednosti Hrvatske u odnosu na svjetski prosjek, ali i identificirana područja za poboljšanje.*

Ključne riječi. dubinska analiza podataka, rodna diskriminacija, rodna nejednakost, deskriptivna analiza, regresijska analiza, grupiranje, otkrivanje anomalija

1 Introduction

The primary objective of this study is to show how advanced data analysis techniques can be used to detect patterns of discrimination and to conduct a comprehensive analysis of gender inequality at the global level and compare these findings with the Croatian context. The analysis focuses on specific domains in which gender inequality is most pronounced, as well as on the key factors influencing it. By identifying countries that exhibit similar patterns of gender inequality, alongside outliers that deviate substantially from prevailing trends, the study positions Croatia within a broader global

1 Uvod

Primarni cilj ovog istraživanja je prikazati kako se napredne tehnike analize podataka mogu koristiti za otkrivanje obrazaca diskriminacije te provesti sveobuhvatnu analizu rodne nejednakosti na globalnoj razini kako bi se mogla izvršiti usporedba s Hrvatskom. Analiza se usredotočuje na specifična područja u kojima je rodna nejednakost najizraženija, kao i na ključne čimbenike koji na nju utječu. Identificiranjem zemalja koje pokazuju

framework. This approach enables researchers and policy makers to gain a better understanding of the causes and manifestations of discrimination and to develop informed strategies to overcome it. However, as Kabeer (2016) notes, even when policy recommendations are present, there is often shortage of concrete outcomes.

Gender equality "refers to the equal rights, responsibilities and opportunities of women and men and girls and boys" (UN Women, 2025). This includes equal opportunities for education, employment, participation in the political process and access to healthcare (World Economic Forum, 2025).

Gender discrimination is a particular form of discrimination based on sex or sexual orientation. The term "discrimination" comes from the Latin term *dis-criminare*, which means "to distinguish between". In a social context, discrimination refers to the behaviour based on prejudice in which people are treated unfairly because they belong to a particular group, regardless of their individual characteristics. Such practices seriously hinder the realization of equality, social development and peace. Gender discrimination manifests itself in the systematic preferential treatment of one gender at the expense of another. Despite progress, it is still a serious social problem with global consequences (United Nations, 2025, World Economic Forum, 2025, Grant Thornton International, 2025).

Gender discrimination can manifest itself in different domains. In the area of education, gender discrimination can lead to unequal opportunities for girls and boys, often resulting in lower literacy rates and educational attainment for women. In the area of healthcare, women often have limited access to certain medical services and poorer health outcomes. In developing countries, women often face a higher risk of death due to complications during pregnancy and childbirth (United Nations, 2025). Even though women's workforce participation globally has risen, gender-based industry segregation persists, with women still concentrated in lower-paying, people-centric industries (United Nations, 2025). Unequal pay for equal work among male and female workers is one of the possible causes of the unadjusted gender pay gap (Leythienne & Pérez-Julián, 2022, Grissom et al., 2021, Poli & Maier, 2022). This inequality is particularly pronounced in certain sectors such as the financial industry and managerial roles (Grant Thornton International, 2025). Political discrimination is another major challenge as women remain significantly underrepresented in the political sphere (United Nations, 2025). This inequality is exacerbated by social norms and legal systems that often do not offer equal protection and opportunities to all genders.

The knowledge discovery in databases (KDD) process encompasses steps ranging from the data collection and preparation, application of analytical

slične obrasce rodne nejednakosti, uz izdvajanje ekstremnih slučajeva koji značajno odstupaju od prevladavajućih trendova, Hrvatska se smješta u širi globalni okvir. Ovakav pristup omogućuje istraživačima i donositeljima odluka bolje razumijevanje uzroka i manifestacija diskriminacije te razvoj informiranih strategija za njeno prevladavanje. Međutim, čak i kada su donesene odgovarajuće politike, često izostaju konkretni rezultati (Kabeer, 2016).

Rodna jednakost odnosi se na ista prava, odgovornosti i mogućnosti žena i muškaraca te djevojčica i dječaka (UN Women, 2025). To uključuje jednake mogućnosti za obrazovanje, zapošljavanje, sudjelovanje u političkom procesu i zdravstvenu skrb (World Economic Forum, 2025).

Rodna diskriminacija predstavlja specifičan oblik diskriminacije koji se temelji na rodu ili rodnoj orijentaciji. Pojam „diskriminacija" potječe od latinskog izraza *dis-criminare*, što znači „razlikovati između". U društvenom kontekstu, diskriminacija podrazumijeva postupanje utemeljeno na predrasudama, pri čemu se osobe nepravedno tretiraju zbog pripadnosti određenoj skupini, neovisno o njihovim individualnim kvalitetama. Takva praksa ozbiljno ugrožava postizanje jednakosti, društvenog razvoja i mira. Rodna diskriminacija se očituje u sustavnom favoriziranju jednog roda nauštrb drugoga. Unatoč napretku, rodna diskriminacija i dalje predstavlja ozbiljan društveni problem s globalnim posljedicama (United Nations, 2025, World Economic Forum, 2025, Grant Thornton International, 2025).

Rodna diskriminacija se može manifestirati u raznim sferama života. U području obrazovanja, rodna diskriminacija može rezultirati nejednakim mogućnostima za djevojčice i dječake, često uzrokujući niže stope pismenosti i obrazovnih uspjeha među ženama. U području zdravstva, žene se često suočavaju s ograničenim pristupom određenim zdravstvenim uslugama te lošijim zdravstvenim ishodima. U zemljama u razvoju, žene su nerijetko izložene većem riziku od smrtnosti zbog komplikacija tijekom trudnoće i poroda (United Nations, 2025). Iako je sudjelovanje žena u radnoj snazi globalno poraslo, rodna segregacija po industrijama i dalje postoji, pri čemu su žene još uvijek koncentrirane u slabije plaćenim, na ljude usmjerenim industrijama (United Nations, 2025). Nejednaka plaća za jednak rad među muškim i ženskim radnicima jedan je od mogućih uzroka neprilagođenog rodnog jaza u plaćama (Leythienne & Pérez-Julián, 2022, Grissom et al., 2021, Poli & Maier, 2022). To je posebno izraženo u određenim sektorima, primjerice u financijskoj industriji te rukovodećim pozicijama (Grant Thornton International, 2025). Politička diskriminacija predstavlja još jedan značajan izazov budući da su žene još uvijek podzastupljene na političkoj sceni (United Nations, 2025). Ova nejednakost dodatno se

techniques, to the interpretation and presentation of results (Tan et al., 2019). The process that was applied in this study was structured using CRISP-DM phases (Cross-Industry Standard Process for Data Mining) (Shearer, 2000). CRISP-DM is designed as a universal framework, independent of specific tools, industries or the problems to be analysed. It defines the key steps in the knowledge discovery process, such as understanding the problem, understanding the data, data preparation, modelling, model evaluation and deployment.

In order to achieve a comprehensive understanding of gender inequality at the global level and compare these findings with the Croatian context, selected data mining techniques are applied. Analysis is conducted using the Kaggle platform's World Data of Gender Inequality Index" dataset.

Descriptive analysis examines existing data to uncover patterns, relationships, and specific domains in which gender inequality is most pronounced. Linear regression estimated by OLS (ordinary least squares) is used to understand and quantify the relationship between variables in order to discover key discrimination factors. While clustering enables the identification of countries that exhibit similar patterns of gender inequality, anomaly detection enables the identification of extreme cases. Finally, comparative analysis positioned Croatia within a broader global framework.

The next section briefly presents related work, followed by an overview of the methodology and the dataset in Section 3. Section 4 reports the results of the selected data mining techniques, while the final section provides a discussion, the main conclusions, and directions for future research.

2 Related Work

Even though gender discrimination can manifest itself in various domains, the literature review reveals that existing studies mostly focus on analysing labour market discrimination. While some researchers focus on uncovering discriminatory behaviour of employers (Hangartner et al., 2021), others such as Edelsztein & Waisbrot (2023), Bonaccolto-Töpfer et al. (2023) and (Alatrista-Salas et al. (2017) focus on wage gaps, on geographical distribution of land ownership and wealth (Fisher & Naidoo, 2016), and on health benefit indicators (Heinz et al., 2020).

Even though gender bias and favoritism can also be detected using natural language processing techniques and sentiment analysis, as demonstrated by Torres-Berrú et al. (2023), the dataset used in this study contains only structured attributes. A detailed overview of text analysis methods used to identify and detect gender bias, inequality, and violence in language can be found in (Nacinovic Prskalo, 2024).

produbljuje uslijed društvenih normi i pravnih sustava koji često ne osiguravaju ravnopravnu zaštitu i mogućnosti.

Proces otkrivanja znanja u bazama podataka (KDD) uključuje korake od prikupljanja i pripreme podataka, preko primjene analitičkih tehnika, pa sve do interpretacije i prezentacije rezultata (Tan et al., 2019). Proces koji je primijenjen u ovom istraživanju je strukturiran po uzoru na CRISP-DM (engl. *Cross-Industry Standard Process for Data Mining*) (Shearer, 2000). CRISP-DM je osmišljen kao univerzalni okvir, neovisan o konkretnim alatima, industrijskim sektorima ili specifičnim problemima koji se analiziraju. On definira ključne korake u procesu otkrivanja znanja poput razumijevanja problema, razumijevanja podataka, pripreme podataka, modeliranja, evaluacije modela i primjene. Analiza je provedena na skupu podataka „World Data of Gender Inequality Indeks" preuzetom s platforme Kaggle.

Deskriptivna analiza ispituje postojeće podatke kako bi se otkrili obrasci, odnosi i specifična područja u kojima je rodna nejednakost najizraženija. Linearna regresija metodom najmanjih kvadrata (OLS) koristi se za razumijevanje i kvantificiranje odnosa između varijabli s ciljem otkrivanja ključnih čimbenika diskriminacije. Grupiranje omogućuje identifikaciju zemalja koje pokazuju slične obrasce rodne nejednakosti, dok detekcija anomalija omogućuje prepoznavanje ekstremnih slučajeva. Na kraju, komparativna analiza smješta Hrvatsku u širi globalni okvir.

U sljedećem poglavlju dan je kratki prikaz povezanih istraživanja, nakon čega u trećem poglavlju slijedi prikaz metodologije i korištenog skupa podataka. Četvrto poglavlje sadrži prikaz rezultata odabranih tehnika dubinske analize podataka. Posljednje poglavlje sadrži interpretaciju rezultata, glavne zaključke i smjernice za buduća istraživanja.

2 Srodna istraživanja

Iako se rodna diskriminacija može očitovati u raznim područjima, pregled literature pokazuje da se većina istraživanja fokusira na analizu diskriminacije na tržištu rada. Dok se neki istraživači bave otkrivanjem diskriminatornog ponašanja poslodavaca (Hangartner et al., 2021), drugi poput Edelsztein & Waisbrot (2023), Bonaccolto-Töpfer et al. (2023) i Alatrista-Salas et al. (2017) proučavaju razlike u plaćama, geografsku raspodjelu vlasništva nad zemljištem i bogatstvom (Fisher & Naidoo, 2016), te zdravstvene indikatore (Heinz et al., 2020).

Iako se rodna pristranost i favoriziranje mogu otkriti korištenjem tehnika obrade prirodnog jezika i

Since this research aims at conducting a comprehensive analysis of gender inequality at the global level in order to compare the findings with the Croatian context, effort was put into finding similar studies. In the research covering the countries of the European Union (Vinska & Tokar, 2021), an agglomerative hierarchical clustering with full association based on the Euclidean distance obtained five groups based on gross domestic product and the Gender Equality Index (GEI) as indicators of economic development and gender equality. The research has shown that there are no distinct "laggards" among EU member states in terms of gender equality; on the contrary, more developed countries almost all belong to the group of "leaders" in gender equality (GEI>90% of the EU average), while less developed countries are equally divided into those that are leaders and those who are gradually adopting the principles of gender equality (GEI between 75-90% of the EU average). It can be concluded that countries with a higher level of development have better GEI scores. Alatrista-Salas et al. (2017) used clustering to investigate gender wage gaps. The cluster analysis conducted by Heinz et al. (2020) challenged the popular belief that greater gender equality automatically means a smaller difference in social well-being among girls and boys and suggested that higher levels of gender equality are associated with increased pressure, dissatisfaction, and related challenges among girls. A global study covering 47 countries suggested that different dimensions of gender inequality may have different causes and occur independently of each other, but also that measures such as income potentially underestimate the extent of gender inequality (Fisher & Naidoo, 2016).

Regression methods help to determine not only whether discrimination exists (e.g., a statistically significant gender effect on wages or another outcome variable), but also how much of the difference remains unexplained after accounting for relevant factors – and it is this unexplained part that is often interpreted as an indicator of discrimination. Research conducted by Bonaccolto-Töpfer et al. (2023) found that the wage gap narrowed significantly among lower-paid workers but widened among the highest-paid, suggesting a persistent "glass ceiling" for women at the top of the income ladder. The authors applied an adapted Oaxaca-Blinder decomposition to estimate and decompose changes between two time points. The approach is based on the method of least squares to estimate the parameters of the regression model.

In this study Croatia is positioned within a broader global framework by discovering key discrimination factors, and by identifying countries that exhibit similar patterns of gender inequality, alongside outliers that deviate substantially from the prevailing trends.

analize sentimenta, npr. Torres-Berrú i sur. (2023), u ovom istraživanju analizira se skup podataka koji sadrži samo strukturirane atribute. Međutim, detaljan pregled metoda analize teksta koje se koriste za prepoznavanje i otkrivanje rodne pristranosti, nejednakosti i nasilja u jeziku može se pronaći u radu (Nacinovic Prskalo, 2024).

Budući da je cilj ovog istraživanja provesti sveobuhvatnu analizu rodne nejednakosti na globalnoj razini kako bi se rezultati usporedili s hrvatskim kontekstom, nastojala su se pronaći slična istraživanja. U istraživanju koje obuhvaća zemlje Europske Unije (Vinska & Tokar, 2021) na temelju Euklidske udaljenosti je aglomerativnim hijerarhijskim grupiranjem s potpunim povezivanjem dobiveno pet grupa na temelju bruto domaćeg proizvoda i indeksa rodne jednakosti (GEI) kao pokazatelja ekonomskog razvoja i rodne jednakosti. Istraživanje je pokazalo da među članicama EU nema izrazito „zaostalih" po pitanju rodne jednakosti. Štoviše, razvijenije države gotovo sve spadaju u skupinu "lidera" u rodnoj jednakosti (GEI>90% EU prosjeka), dok su manje razvijene države podjednako podijeljene na one koje su lideri i one koje su "pristupnici" koji postupno usvajaju načela rodne jednakosti (GEI između 75-90% EU prosjeka). Može se zaključiti da zemlje s višim stupnjem razvoja imaju bolje GEI rezultate. Alatrista-Salas i sur. (2017) koriste grupiranje kako bi istražili rodne razlike u plaćama. Rezultati grupiranja koje su proveli Heinz i sur. (2020) osporili su uvriježeno mišljenje da veća rodna ravnopravnost automatski znači i manju razliku u zdravstvenoj dobrobiti između djevojaka i mladića i ukazuju da je kod većeg stupnja rodne jednakosti izgledniji veći pritisak i nezadovoljstvo kod djevojaka. Globalno istraživanje provedeno u 47 zemalja sugerira da različite dimenzije rodne nejednakosti mogu imati različite uzroke i javljati se neovisno jedna o drugoj, ali i da mjere poput dohotka mogu potencijalno podcijeniti razmjere rodne nejednakosti (Fisher & Naidoo, 2016).

Metode regresije mogu ne samo utvrditi postoji li diskriminacija (npr. statistički značajan učinak na plaće ili neku drugu zavisnu varijablu), nego i koliki dio razlike ostaje neobjašnjen nakon što se uzmu u obzir relevantni čimbenici, a upravo se taj neobjašnjeni dio često tumači kao pokazatelj diskriminacije. Istraživanje koje su proveli Bonaccolto-Töpfer i sur. (2023) pokazalo je da se jaz u plaćama značajno smanjio među niže plaćenim radnicima, ali je porastao među najviše plaćenima, što upućuje na postojan "stakleni strop" za žene na vrhu ljestvice dohodaka. Autori su primijenili prilagođenu Oaxaca-Blinder dekompoziciju za procjenu i dekompoziciju promjena između dvije vremenske točke. Pristup se temelji na metodi najmanjih kvadrata za procjenu parametara regresijskog modela.

3 Methodology and Data

Advances in technology and the proliferation of big data have facilitated the application of in-depth analytical methods capable of uncovering latent patterns within complex datasets. These insights contribute to the development of more effective business strategies, the enhancement of healthcare systems, and a deeper understanding of social phenomena. By applying advanced analytical techniques and machine learning algorithms, an in-depth analysis enables the discovery of hidden relationships that can serve as the basis for developing effective strategies and interventions to reduce gender discrimination and achieve gender balance.

3.1 Methodology

This research follows the CRISP-DM methodology. Data processing and modelling were performed using the R programming language. The workflow is shown in Fig. 1.

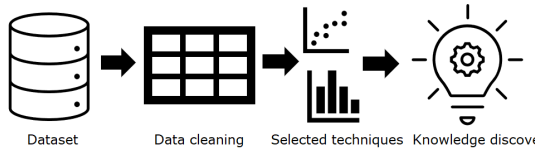


Figure 1. Phases of the research process

First, descriptive analysis was applied to the existing data to uncover patterns, relationships, and specific domains in which gender inequality is most pronounced. The descriptive analysis provided a deep insight into the basic characteristics of the data and revealed important patterns and trends that indicate the presence of discrimination.

Regression analysis is a classic statistical method for quantifying discrimination, and it is the second technique selected for this in-depth analysis. To understand and quantify the relationship between the predictors and the dependent variable, linear regression model estimated via OLS is used.

Clustering algorithms allow visualization of discrimination patterns that follow geographical, cultural, or economic lines, which is valuable for targeted policymaking. We used it to identify countries that exhibit similar patterns of gender inequality. The K-Means algorithm was selected due to its efficiency and simplicity. To determine the optimal number of clusters, the elbow method was used. This method visually illustrates the changes in total variance (sum of squared distances within groups) for different numbers of clusters. The optimal number corresponds to the point at which the decrease in variance begins to slow significantly – a

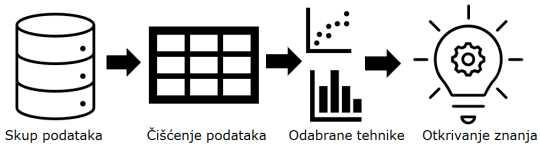
U ovom se istraživanju Hrvatska smješta u širi globalni okvir otkrivanjem ključnih čimbenika diskriminacije i prepoznavanjem zemalja koje pokazuju slične obrasce rodne nejednakosti, uz izdvajanje onih koje znatno odstupaju od prevladavajućih trendova.

3 Metodologija i skup podataka

Zahvaljujući razvoju tehnologije i dostupnosti velike količine podataka, dubinska analiza omogućuje prepoznavanje skrivenih obrazaca unutar kompleksnih skupova podataka, pružajući uvide koji mogu unaprijediti poslovne strategije, poboljšati zdravstvenu skrb te produbiti razumijevanje društvenih pojava. Primjenom naprednih analitičkih tehnika i algoritama strojnog učenja, dubinska analiza omogućuje otkrivanje skrivenih povezanosti koje mogu poslužiti kao temelj za oblikovanje učinkovitih politika i intervencija usmjerenih na smanjenje rodne diskriminacije i postizanje rodne ravnoteže.

3.1 Metodologija

Ovo istraživanje slijedi CRISP-DM metodologiju. Obrada podataka i modeliranje provedeni su korištenjem programskog jezika R. Proces je prikazan na slici 1.



Slika 1. Faze istraživačkog procesa

Najprije je provedena deskriptivna analiza kojom se analiziraju postojeći podaci kako bi se otkrili obrasci, odnosi i specifična područja u kojima je rodna nejednakost najizraženija. Deskriptivna analiza je pružila dubinski uvid u osnovne karakteristike podataka i otkrila važne obrasce i trendove koji mogu ukazivati na diskriminaciju.

Regresijska analiza klasična je statistička metoda za kvantificiranje diskriminacije, a odabrana je kao sljedeća tehnika dubinske analize podataka koja će se primijeniti. Za razumijevanje i kvantificiranje odnosa između prediktora i zavisne varijable, korišten je model linearne regresije i procjena metodom najmanjih kvadrata.

Algoritmi grupiranja omogućuju da se vizualiziraju obrasci diskriminacije koji prate geografske, kulturne ili ekonomske linije, što je iznimno važno za ciljano kreiranje politika. U ovom istraživanju se koristi za identifikaciju skupina zemalja sa sličnim obrascima rodne nejednakosti. K-Means algoritam je odabran zbog svoje učinkovitosti i jednostavnosti. Za određivanje

balance between the complexity of the model and its effectiveness in clustering the data. The parameter *random_state* is set to 42.

The Isolation Forest model was used to detect anomalies, with the parameter *contamination* set to 0.05 and *random_state* to 42 to ensure the reproducibility of the results. The parameter *contamination* sets the expected proportion of anomalies in the dataset to about 5%. Anomaly detection enabled the identification of countries that deviate significantly from the usual trends and patterns.

Finally, comparative analysis enabled positioning Croatia within a broader global framework.

3.1 Dataset

The dataset "WORLD DATA OF Gender Inequality Index"[1] was downloaded from the Kaggle website. The dataset contains a total of 195 instances. After pre-processing, i.e. removing instances with missing feature values, 170 instances remained in the dataset. The dataset consists of the features shown in Table 1. The following is a description of individual features and their significance for the conducted analysis. The Human Development Index (*HDI rank*) is a composite measure of performance in three key dimensions of human development: a long and healthy life, access to knowledge, and an adequate standard of living. HDI represents the geometric mean of the normalised indices for each of these three dimensions (UNDP, 2021). Lower numerical ranks indicate higher human development, while higher ranks indicate lower human development. *Maternal_mortality* indicates access to health services for women. *Adolescent_birth_rate* reflects education and health conditions for young girls. *Seats_parliament* (% held by women) measures the political participation and empowerment of women in political structures. *F_secondary_educ* and *M_secondary_educ* indicate the percentage of women and men, respectively, aged 25 and over, who have completed secondary education. *F_secondary_educ* and *M_secondary_educ* are used to assess the gender balance in education. While *F_Labour_force* reflects the economic engagement of women in the country, *M_Labour_force* reflects the engagement of men. These two features are used to assess gender balance in the labour market. The Gender Inequality Index (*GII rank*) reflects gender-specific disadvantages in three dimensions: reproductive health, empowerment and the labour market. The index scores range from 0, indicating full gender equality, to 1, indicating gender inequality in all dimensions measured (UNDP, 2021).

optimalnog broja klastera korištena je metoda „lakat". Ona vizualno prikazuje promjene u ukupnoj varijanci (zbroju kvadrata udaljenosti unutar grupa) u odnosu na broj grupa. Optimalni broj grupa odgovara točki u kojoj dolazi do naglog usporavanja smanjenja varijance i ta točka predstavlja kompromis između složenosti modela i njegove učinkovitosti u grupiranju podataka. Parametar *random_state* je postavljen na 42.

Otkrivanje anomalija pomaže u prepoznavanju zemalja koje značajno odstupaju od uobičajenih trendova i struktura. Za otkrivanje anomalija primijenjen je model izolacijskih šuma. Parametar *contamination* odnosi se na očekivani udio anomalija u skupu podataka, a u ovom istraživanju iznosi 0,05 čime se očekivani udio anomalija postavlja na oko 5%. Parametar *random_state* postavljen je na 42 kako bi se osigurala reproducibilnost rezultata.

Na kraju, komparativna analiza omogućila je smještanje Hrvatske u širi globalni okvir.

3.1 Podatkovni skup

Skup podataka "WORLD DATA OF Gender Inequality Indeks"[1] preuzet je u s web stranice Kaggle. Skup se sastoji od ukupno 195 instanci. Nakon pretprocesiranja, tj. uklanjanja instanci s nedostajućim vrijednostima značajki, u skupu je ostalo 170 instanci. Skup se sastoji od značajki, tj. atributa prikazanih u Tablici 1. U nastavku slijedi opis pojedinih atributa i njihov značaj za provedenu analizu.

Indeks ljudskog razvoja (*HDI rank*) je sažeta mjera dostignuća u tri ključne dimenzije ljudskog razvoja: dug i zdrav život, pristup znanju i životni standard. HDI predstavlja geometrijsku sredinu normaliziranih indeksa za svaku od ove tri dimenzije (UNDP, 2021). Niži brojčani rangovi označavaju visoki ljudski razvoj, a viši brojčani rangovi označavaju niži ljudski razvoj. *Maternal_mortality* pokazuje pristup zdravstvenim uslugama za žene. *Adolescent_birth_rate* odražava obrazovne i zdravstvene uvjete adolescenata. *Seats_parliament (% held by women)* mjeri političku participaciju i osnaživanje žena u političkim strukturama. *F_secondary_educ* i *M_secondary_educ* predstavljaju postotak žena, odnosno muškaraca, u dobi od 25 i više godina koji su stekli srednjoškolsko obrazovanje. Koriste se za procjenu rodne ravnoteže u obrazovanju. Analogno, dok *F_Labour_force* pokazuje ekonomski angažman žena u zemlji, *M_Labour_force* odražava angažman muškaraca. Te dvije značajke koriste se za procjenu rodne ravnoteže na tržištu rada. Indeks rodne nejednakosti (*GII rank*) odražava rodno zasnovane nepovoljnosti u tri dimenzije: reproduktivno zdravlje, osnaženost i tržište rada. Vrijednosti indeksa se kreću od 0, gdje

[1] https://www.kaggle.com/datasets/rajkumarpandey02/world-data-of-gender-inequality-index

Table 1. Features of the dataset

Feature	Feature description
HDI rank	the overall level of human development in a country based on factors such as health, education and standard of living
Country	name of the country
HDI value	level of human development in the country
GII value	GII value for the country
GII rank	ranking of countries based on their GII score
Maternal_mortality	number of women who die from pregnancy-related causes per 100,000 live births
Adolescent_birth_rate	number of births to women aged 15–19 per 1,000 women in this age group
Seats_parliamentt(% held by women)	percentage of seats held by women in the national parliament
F_secondary_educ	percentage of women aged 25 and over who have completed secondary education
M_secondary_educ	percentage of males aged 25 and over who have completed secondary education
F_Labour_force	percentage of women participating in the labour force
M_Labour_force	percentage of men in the labour force

4 Results

The following subsections present the results of the descriptive analysis, the regression analysis, the clustering, the detection of anomalies and, finally, the comparative analysis.

4.1 Descriptive Analysis

The boxplot analysis shows that most GII values are between 0.15 and 0.50 and that in most countries the rate of secondary school completion for both genders is between 40% and 90%. Female labour force participation shows a wider range, with several

postoji potpuna rodna jednakost, do 1, gdje je rodna nejednakost uočena u svim mjerenim dimenzijama (UNDP, 2021).

Tablica 1. Značajke skupa podataka

Značajka	Opis značajke
HDI rank	opća razina ljudskog razvoja u zemlji na temelju faktora kao što su zdravlje, obrazovanje i standard života
Country	naziv zemlje
HDI value	razina ljudskog razvoja u zemlji
GII value	vrijednost GII-ja za zemlju
GII rank	rangiranje zemalja prema njihovom GII-u
Maternal_mortality	broj žena umrlih od uzroka povezanih s trudnoćom na 100 000 živorođene djece
Adolescent_birth_rate	broj poroda u žena u dobi od 15 do 19 godina na 1000 žena u dobi od 15 do 19 godina
Seats_parliamentt(% held by women)	udio žena u parlamentu u ukupnom broju zastupnika/zastupnica
F_secondary_educ	postotak žena u dobi od 25 i više godina koje je steklo srednjoškolsko obrazovanje
M_secondary_educ	postotak muškaraca u dobi od 25 i više godina koje je steklo srednjoškolsko obrazovanje
F_Labour_force	postotak žena koje sudjeluju u radnoj snazi
M_Labour_force	postotak muškaraca koji sudjeluju u radnoj snazi

4 Rezultati

U sljedećim podpoglavljima prikazani su rezultati deskriptivne analize, regresijske analize, grupiranja, otkrivanja anomalija, i zaključno, komparativne analize.

4.1 Deskriptivna analiza

Analizom kutijastih dijagrama zaključeno je da se većina vrijednosti GII-ja nalazi između 0.15 i 0.50 i da većina zemalja ima udio srednjoškolskog obrazovanja između 40% i 90% za oba roda. Sudjelovanje žena u radnoj snazi ima veći raspon, uz

outliers at the lower end, while male labour force participation is higher and more constant (65–75%). In most countries, women hold between 10% and 35% of parliamentary seats. Although most countries have a low maternal mortality rate, there are some exceptions. The situation is similar with the adolescent birth rate, which is below 75 in most countries, but far higher in some. With both features, extreme values are observed at the upper end.

The correlation matrix is important to understand how different dimensions of gender inequality influence each other. There is a strong positive correlation between GII, maternal mortality and adolescent birth rate (Fig. 2) and a significant negative correlation between these variables and the completion of secondary school for both females and males (Fig. 3).

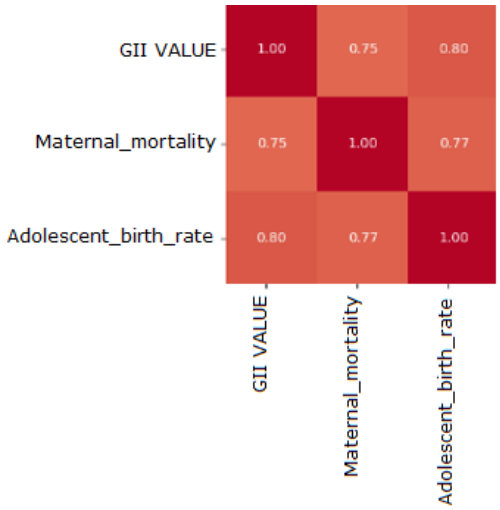


Figure 2. Correlation between GII, maternal mortality and adolescent birth rate

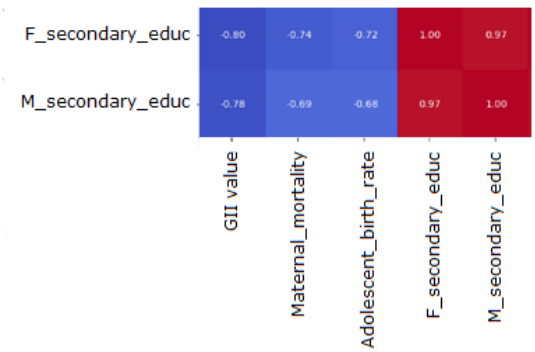


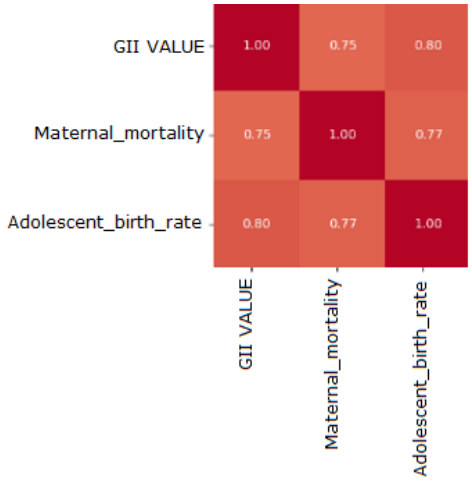
Figure 3. Correlation with secondary education

4.2 Regression Analysis

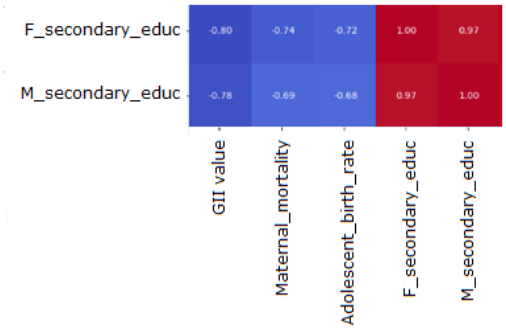
Regression analysis is used to identify key factors that contribute significantly to the variability of GII values in different countries. Scatterplots are generated to provide better understanding of the

nekoliko stršećih podataka prema nižim vrijednostima, dok je kod muškaraca sudjelovanje veće i ujednačenije (65-75%). Većina zemalja imala je između 10% i 35% zastupnica. Uz nekoliko iznimaka, većina zemalja je imala nisku stopu mortaliteta rodilja. Također, iako je većina zemalja imala stopu nataliteta kod adolescenata ispod 75, neke su znatno premašile tu vrijednost. Dakle, u obje značajke uočene su ekstremne vrijednosti na višem kraju.

Korelacijska matrica ključna je za razumijevanje kako pojedine dimenzije rodne nejednakosti utječu jedna na drugu. Utvrđena je visoka pozitivna korelacija između GII, stope mortaliteta rodilja smrtnosti i stope nataliteta kod adolescenata (Slika 2) i značajna negativna korelacija navedenih varijabli sa srednjoškolskim obrazovanjem oba roda (Slika 3).



Slika 2. Korelacija između GII, stope mortaliteta rodilja smrtnosti i stope nataliteta kod adolescenata



Slika 3. Korelacija sa srednjoškolskim obrazovanjem

4.2 Regresijska analiza

Regresijska analiza korištena je za prepoznavanje ključnih čimbenika koji značajno doprinose varijabilnosti GII vrijednosti među različitim zemljama. Za bolje razumijevanje odnosa između

relationships between the independent variables and GII.

The results of the least squares method show that the linear regression model effectively explains the variation in the GII, with a coefficient of determination (R^2) of 0.919, which means that the model explains 91.9% of the total variation. The F-statistic is 258.1, and the extremely low p-value (2.17e-83) confirms the high statistical significance of the model.

The variables HDI rank and adolescent birth rate have a positive and statistically significant impact on GII, indicating that countries with lower levels of human development (i.e. higher HDI rank) and higher adolescent birth rates contribute to greater gender inequality (Table 2). On the other hand, both, female representation in parliament and female labour force participation, have negative and statistically significant impact on GII, suggesting that a higher proportion of women in political decision-making and the labour market contributes to a decrease in gender inequality (Table 2).

Table 2. Coefficients

	Coefficient	p>|t|
Constant	0.2448	0.000
HDI rank	0.0019	0.000
Adolescent_birth_rate	0.0015	0.000
Seats_parliamentt(% held by women)	-0.0032	0.000
F_secondary_educ	-0.0009	0.210
M_secondary_educ	0.0003	0.728
F_Labour_force	-0.0018	0.000
M_Labour_force	0.0009	0.141

4.3 Clustering

Clustering enabled the identification of groups of countries with similar patterns of gender inequality. The K-Means algorithm was selected due to its efficiency and simplicity. The optimal number of clusters is set to 3, as determined by the elbow method (Fig. 4).

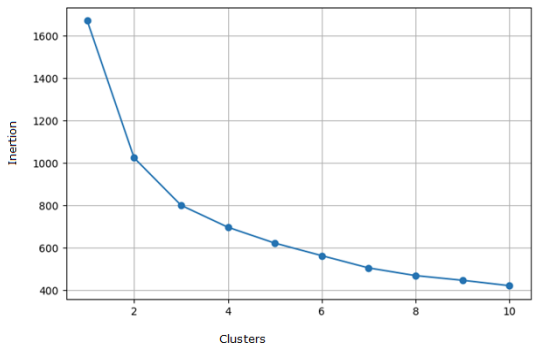


Figure 4. Elbow method

nezavisnih varijabli i GII-ja kreirani su dijagrami raspršenosti.

Rezultati metode najmanjih kvadrata pokazali su da linearni regresijski model vrlo učinkovito objašnjava varijacije u vrijednosti indeksa rodne nejednakosti (GII), s koeficijentom determinacije (R^2) od 0,919, što znači da model objašnjava 91,9 % ukupne varijabilnosti. F-statistika iznosi 258,1, a izuzetno niska p-vrijednost (2,17e-83) potvrđuje visoku statističku značajnost modela.

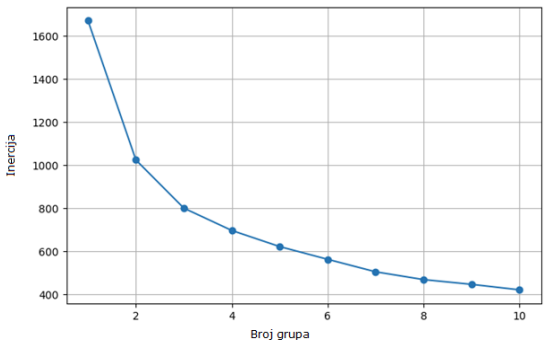
Varijable HDI rang i stopa nataliteta kod adolescenata pokazali su pozitivan i statistički značajan utjecaj na GII, upućujući na to da zemlje s nižim stupnjem ljudskog razvoja, tj. višim HDI rangom te većom stopom nataliteta kod adolescenata doprinose višoj razini rodne nejednakosti (Tablica 2). Varijable zastupljenost žena u parlamentu i udio žena u radnoj snazi su pokazali negativan i statistički značajan utjecaj na vrijednost GII-a, ukazujući na to da veći udio žena u političkom odlučivanju i na tržištu rada doprinosi smanjenju rodne nejednakosti (Tablica 2).

Tablica 2. Koeficijenti

	Koeficijent	p>|t|
Konstanta	0.2448	0.000
HDI rank	0.0019	0.000
Adolescent_birth_rate	0.0015	0.000
Seats_parliamentt(% held by women)	-0.0032	0.000
F_secondary_educ	-0.0009	0.210
M_secondary_educ	0.0003	0.728
F_Labour_force	-0.0018	0.000
M_Labour_force	0.0009	0.141

4.3 Grupiranje

Grupiranje je omogućilo identifikaciju skupina zemalja sa sličnim obrascima rodne nejednakosti. Odabran je K-Means algoritam s 3 grupe. Za određivanje optimalnog broja grupa korištena je metoda „lakat“ (Slika 4).



Slika 4. Metoda “lakat”

The cluster with high GII values represented countries with greater gender inequality, whereas the cluster with low GII values corresponded to countries with lower levels of gender inequality (Fig. 5).

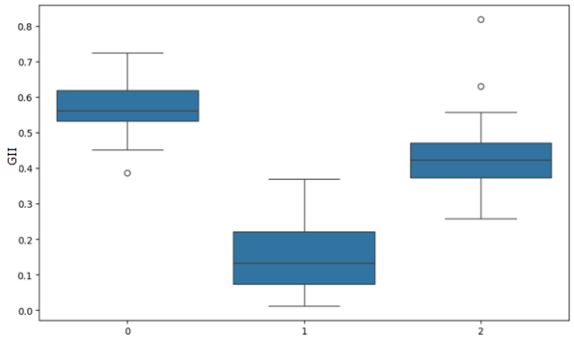


Figure 5. GII distribution per cluster

4.4 Anomaly Detection

The Isolation Forest model was used to detect anomalies, which enabled the identification of countries that represent extreme cases of gender inequality.

Countries such as Tonga, Papua New Guinea, Mauritania, Nigeria, Rwanda, Afghanistan, Yemen, Mozambique and the Central African Republic are characterised by unfavourable indicators and require special attention and targeted measures to reduce discrimination and promote gender equality.

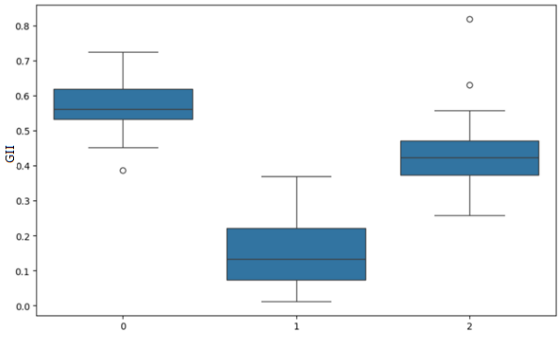
4.5 Comparative Analysis

In order to draw well-founded conclusions about the situation in Croatia, a comparative analysis was carried out. A comparison of Croatia's feature values with global average and the average for countries with very high HDI is given in Table 3.

Table 3. Comparison of Croatia's feature values with global averages

	CRO	World	Countries with high HDI
GII VALUE	0.09	0.3	0.15
Maternal_mortality	8	139.59	13.98
Adolescent_birth_rate	8.6	43.62	13.51
Seats_parliamentt (% held by women)	31.1	25.30	28.40
F_secondary_educ	97	62.78	85.82
M_secondary_educ	100	67.07	87.83
F_Labour_force	45.9	50.22	52.70
M_Labour_force	58.8	70.09	69.68

Grupa s visokim GII vrijednostima predstavlja zemlje s većom rodnom nejednakošću, dok grupa s niskim GII vrijednostima predstavlja one s nižom razinom rodne nejednakosti (Slika 5).



Slika 5. Raspodjela GII po grupama

4.4 Otkrivanje anomalija

Za otkrivanje anomalija u ovom istraživanju je primijenjen model izolacijskih šuma, čime su prepoznate zemlje koje predstavljaju ekstremne slučajeve rodne nejednakosti.

Zemlje poput Tonge, Papue Nove Gvineje, Mauritanije, Nigerije, Ruande, Afganistana, Jemena, Mozambika i Centralnoafričke Republike izdvojile su se po svojim nepovoljnim pokazateljima te su ukazale na potrebu ciljanih intervencija kako bi se smanjila diskriminacija i unaprijedila rodna ravnopravnost.

4.5 Komparativna analiza

Kako bi se donijeli utemeljeni zaključci o situaciji u Hrvatskoj, izvršena je komparativna analiza. Usporedba značajki Hrvatske s globalnim prosjekom i s prosjekom zemalja s vrlo visokim indeksom ljudskog razvoja dana je u Tablici 3.

Tablica 3. Usporedba značajki Hrvatske s globalnim prosjecima

	HR	Svijet	Države visokog HDI-ja
GII VALUE	0.09	0.3	0.15
Maternal_mortality	8	139.59	13.98
Adolescent_birth_rate	8.6	43.62	13.51
Seats_parliamentt (% held by women)	31.1	25.30	28.40
F_secondary_educ	97	62.78	85.82
M_secondary_educ	100	67.07	87.83
F_Labour_force	45.9	50.22	52.70
M_Labour_force	58.8	70.09	69.68

5 Discussion and Conclusion

In this study, the selected techniques were used to comprehensively analyse the causes and manifestations of gender inequality. This resulted in a detailed analysis of gender inequality at the global level and enabled comparing these findings with the Croatian context.

Descriptive analysis revealed the underlying patterns and characteristics of the data. A strong positive correlation was determined between GII, maternal mortality rate, and adolescent birth rate, and a negative correlation was determined between these indicators and secondary education.

Regression analysis identified the key socio-economic factors contributing to inequality. Countries with higher HDI rank (worse human development), more adolescent births, fewer women in parliament, and lower female labour force participation tend to have lower gender equality (higher GII). Education (both male and female secondary education) and male labour force participation did not show significant effects on GII.

Countries were categorised into groups based on the extent of gender discrimination. The visualisation of the partial plots clearly showed that variables such as the maternal mortality rate, the adolescent birth rate and HDI had a strong influence on clustering. In contrast, variables such as the representation of women in parliament and the labour force participation of women showed considerable overlaps between the clusters, which indicated that they are less effective in distinguishing countries based on the characteristics analysed. Cluster 0 had the highest rates of maternal mortality and adolescent pregnancies, indicating countries with greater health risks, cluster 2 included countries with the lowest rates, while cluster 1 fell in-between.

Anomaly detection highlighted countries that deviated significantly from the usual patterns, highlighting the most critical cases.

Finally, comparative analysis enabled the cross-country comparison of key indicators, providing a deeper insight into global and local trends in gender inequality. Significantly lower Croatia's GII indicated relatively higher level of gender equality in Croatia and suggested the effectiveness of measures to reduce gender inequality. Additionally, an exceptionally low maternal mortality rate and a low number of adolescent pregnancies indicated the efficiency of its health and education system. The representation of women in parliament above the global average and above the average of countries with a very high HDI indicated progress in the political participation of women. Croatia demonstrated exceptional strength in the area of education, as the secondary school completion rate for both men and women was well above the global average and the average for industrialised countries.

5 Diskusija i zaključak

U ovom istraživanju primijenjen je niz analitičkih tehnika s ciljem sveobuhvatnog sagledavanja uzroka i manifestacija rodne nejednakosti. To je rezultiralo detaljnom analizom rodne nejednakosti na globalnoj razini i omogućilo pozicioniranje Hrvatske po tom pitanju.

Deskriptivna analiza otkrila je temeljne obrasce i karakteristike podataka. Utvrđena je snažna pozitivna korelacija između GII-ja, stope mortaliteta rodilja i stope nataliteta kod adolescenata, a negativna između navedenih značajki i srednjoškolskog obrazovanja.

Regresijom su identificirani ključni socioekonomski čimbenici koji doprinose nejednakosti. Zemlje s višim rangom prema HDI-ju (što ukazuje na lošiji ljudski razvoj), većom stopom nataliteta kod adolescenata, manjim udjelom žena u parlamentu te nižom stopom sudjelovanja žena u radnoj snazi pokazuju nižu razinu rodne jednakosti (viši GII). Obrazovanje (srednjoškolsko, i žensko i muško) te sudjelovanje muškaraca u radnoj snazi nisu pokazali značajan učinak na GII.

Grupirane su zemlje prema razinama rodne diskriminacije. Vizualizacija parcijalnih dijagrama jasno je pokazala da varijable poput stope mortaliteta rodilja, stope nataliteta kod adolescenata i razine ljudskog razvoja imaju snažan utjecaj na formiranje grupa. Nasuprot tome, varijable poput zastupljenosti žena u parlamentu i udjela žena u radnoj snazi pokazale su značajno preklapanje među grupama, što upućuje na njihovu manju učinkovitost u razlikovanju zemalja prema analiziranim obilježjima. Klaster 0 se istaknuo s najvišim stopama mortaliteta rodilja i adolescentskih trudnoća ukazujući na zemlje s većim zdravstvenim rizicima, klaster 2 je obuhvatio zemlje s najnižim stopama, dok je klaster 1 obuhvatio zemlje sa srednjim rizicima.

Detekcija anomalija je izdvojila države koje značajno odstupaju od uobičajenih obrazaca, ističući najkritičnije slučajeve.

Konačno, komparativna analiza je omogućila usporedbu ključnih pokazatelja među državama, pružajući dublji uvid u globalne i lokalne trendove rodne nejednakosti. Za Hrvatsku je zabilježena značajno niža vrijednost indeksa rodne nejednakosti (GII) u usporedbi sa svjetskim prosjekom, ali i s prosjekom zemalja s vrlo visokim indeksom ljudskog razvoja (HDI). Ova razlika upućuje na relativno višu razinu rodne ravnoteže u Hrvatskoj te sugerira učinkovitost politika usmjerenih na smanjenje rodnih nejednakosti. Dodatno, za Hrvatsku je zabilježena izuzetno niska stopa mortaliteta rodilja i adolescentskih trudnoća, upućujući na učinkovitost zdravstvenog i obrazovnog sustava. Zastupljenost žena u parlamentu je premašila globalni prosjek i prosjek

However, the labour market participation of both genders was still below average, which indicated that further efforts are needed to promote greater integration of the population into the labour market.

Data mining is a powerful tool for uncovering hidden patterns and understanding complex phenomena such as gender discrimination, enabling informed and targeted decision-making. It should be used as the basis for developing effective policies and initiatives to reduce inequality at a global level.

A key limitation of the present study lies in the time-bound nature of the dataset. Access to more recent or feature-rich data would allow for deeper insights and broaden the scope of investigation.

Future research will focus on evaluating model diagnostics to assess the validity of the linear regression approach, as well as applying alternative data mining techniques for comparative analysis.

Acknowledgments

This work was fully supported by the University of Rijeka as part of the scientific project "LinguaGender: Primjena računalne obrade jezika u analizi rodne pristranosti i nasilja u jeziku", uniriiskusni-drustv-23-271.

References

Alatrista-Salas, H., Esposito, B., Nunez-del-Prado, M., & Valdivieso, M. (2017). Measuring the gender discrimination: A machine learning approach. *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 1–6.

Bonaccolto-Töpfer, M., Castagnetti, C., & Rosti, L. (2023). Changes in the gender pay gap over time: the case of West Germany. *Journal for Labour Market Research, 57(*1). https://doi.org/10.1186/s12651-023-00338-7

Edelsztein, V. C., & Waisbrot, S. A. (2023). Breaking down the Gender Pay Gap through a machine learning model. *Convergencia*, 30. https://doi.org/10.29101/crcs.v30i0.20656

Fisher, B., & Naidoo, R. (2016). The geography of gender inequality. *PLoS ONE, 11*(3). https://doi.org/10.1371/journal.pone.0145778

World Economic Forum. (2025). *Global Gender Gap Report*. https://reports.weforum.org/docs/WEF_GGGR_2025.pdf

Grissom, J. A., Timmer, J. D., Nelson, J. L., & Blissett, R. S. L. (2021). Unequal pay for equal work? Unpacking the gender gap in principal

zemalja s vrlo visokim HDI-jem, odražavajući napredak u političkoj participaciji žena. Hrvatska se posebno istaknula u području obrazovanja, s udjelom žena i muškaraca koji završavaju srednju školu znatno iznad svjetskog prosjeka i prosjeka razvijenih zemalja. Ipak, sudjelovanje u radnoj snazi bilo je ispodprosječno za jedne i druge, upućujući na potrebu za dodatnim naporima u poticanju radne aktivnosti stanovništva.

Dubinska analiza podataka predstavlja snažan alat za otkrivanje skrivenih obrazaca i razumijevanje složenih pojava poput rodne diskriminacije, omogućujući donošenje informiranih i ciljanih odluka. Ona treba činiti temelj za oblikovanje učinkovitih politika i inicijativa usmjerenih na smanjenje nejednakosti na globalnoj razini.

Ključno ograničenje ovog istraživanja odnosi se na vremenski ograničenu prirodu skupa podataka. Pristup novijim ili sadržajnijim skupovima podataka omogućio bi dublje uvide i proširio opseg istraživanja.

Buduća istraživanja usmjerit će se na procjenu dijagnostičkih pokazatelja modela kako bi se ocijenila valjanost pristupa linearne regresije, kao i na primjenu alternativnih metoda dubinske analize podataka radi komparativne analize.

Zahvale

Ovaj rad je u potpunosti podržan od strane Sveučilišta u Rijeci kao dio znanstvenog projekta "LinguaGender: Primjena računalne obrade jezika u analizi rodne pristranosti i nasilja u jeziku", uniriiskusni-drustv-23-271.

Reference

Alatrista-Salas, H., Esposito, B., Nunez-del-Prado, M., & Valdivieso, M. (2017). Measuring the gender discrimination: A machine learning approach. *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 1–6.

Bonaccolto-Töpfer, M., Castagnetti, C., & Rosti, L. (2023). Changes in the gender pay gap over time: the case of West Germany. *Journal for Labour Market Research, 57(*1). https://doi.org/10.1186/s12651-023-00338-7

Edelsztein, V. C., & Waisbrot, S. A. (2023). Breaking down the Gender Pay Gap through a machine learning model. *Convergencia*, 30. https://doi.org/10.29101/crcs.v30i0.20656

Fisher, B., & Naidoo, R. (2016). The geography of gender inequality. *PLoS ONE, 11*(3). https://doi.org/10.1371/journal.pone.0145778

compensation. *Economics of Education Review, 82*, 102114.

Hangartner, D., Kopp, D., & Siegenthaler, M. (2021). Monitoring hiring discrimination through online recruitment platforms. *Nature, 589*(7843), 572–576. https://doi.org/10.1038/s41586-020-03136-0

Heinz, A., Catunda, C., van Duin, C., Torsheim, T., & Willems, H. (2020). Patterns of Health-Related Gender Inequalities—A Cluster Analysis of 45 Countries. *Journal of Adolescent Health, 66*(6), S29–S39. https://doi.org/10.1016/j.jadohealth.2020.02.011

Kabeer, N. (2016). Gender Equality, Economic Growth, and Women's Agency: the "Endless Variety" and "Monotonous Similarity" of Patriarchal Constraints. *Feminist Economics*, 22(1), 295–321. https://doi.org/10.1080/13545701.2015.1090009

Leythienne, D., & Pérez-Julián, M. (2022). *Gender pay gaps in the European Union-a statistical analysis*.

Nacinovic Prskalo, L. (2024). Text Analysis Methods for Identifying and Recognizing Gender Bias, Inequality, and Violence in Language: An Overview. *Proceedings of the Central European Conference on Information and Intelligent Systems*. https://www.researchgate.net/publication/389627282

Poli, D., & Maier, S. (2022). Enforcing`EqualEnforcing`Equal Pay for Equal Work' in the EU: what would it take? In *JRC Working Papers on Taxation and Structural Reforms* (11; Issue 11). https://joint-research-centre.ec.europa.eu

Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing, 5(*4), 13–22. https://mineracaodedados.wordpress.com/wp-content/uploads/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf

Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). *Introduction to Data Mining (2nd global ed.)*. Pearson, New York.

UNDP. (2021). *Technical Notes. Human Development Reports.* https://hdr.undp.org/sites/default/files/2021-22_HDR/hdr2021-22_technical_notes.pdf

United Nations. (2025). *The Gender Snapshot 2025*. https://reports.weforum.org/docs/WEF_GGGR_2025.pdf

World Economic Forum. (2025). *Global Gender Gap Report*. https://reports.weforum.org/docs/WEF_GGGR_2025.pdf

Grissom, J. A., Timmer, J. D., Nelson, J. L., & Blissett, R. S. L. (2021). Unequal pay for equal work? Unpacking the gender gap in principal compensation. *Economics of Education Review, 82*, 102114.

Hangartner, D., Kopp, D., & Siegenthaler, M. (2021). Monitoring hiring discrimination through online recruitment platforms. *Nature, 589*(7843), 572–576. https://doi.org/10.1038/s41586-020-03136-0

Heinz, A., Catunda, C., van Duin, C., Torsheim, T., & Willems, H. (2020). Patterns of Health-Related Gender Inequalities—A Cluster Analysis of 45 Countries. *Journal of Adolescent Health, 66*(6), S29–S39. https://doi.org/10.1016/j.jadohealth.2020.02.011

Kabeer, N. (2016). Gender Equality, Economic Growth, and Women's Agency: the "Endless Variety" and "Monotonous Similarity" of Patriarchal Constraints. *Feminist Economics*, 22(1), 295–321. https://doi.org/10.1080/13545701.2015.1090009

Leythienne, D., & Pérez-Julián, M. (2022). *Gender pay gaps in the European Union-a statistical analysis*.

Nacinovic Prskalo, L. (2024). Text Analysis Methods for Identifying and Recognizing Gender Bias, Inequality, and Violence in Language: An Overview. *Proceedings of the Central European Conference on Information and Intelligent Systems*. https://www.researchgate.net/publication/389627282

Poli, D., & Maier, S. (2022). Enforcing`EqualEnforcing`Equal Pay for Equal Work' in the EU: what would it take? In *JRC Working Papers on Taxation and Structural Reforms* (11; Issue 11). https://joint-research-centre.ec.europa.eu

Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing, 5(*4), 13–22. https://mineracaodedados.wordpress.com/wp-content/uploads/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf

Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). *Introduction to Data Mining (2nd global ed.)*. Pearson, New York.

Grant Thornton International. (2025). *Women in Business 2025 – Impacting the missed generation.*

Torres-Berru, Y., Lopez-Batista, V. F., & Zhingre, L. C. (2023). A Data Mining Approach to Detecting Bias and Favoritism in Public Procurement. *Intelligent Automation and Soft Computing, 36*(3), 3501–3516. https://doi.org/10.32604/iasc.2023.035367

UN Women. (2025). *Gender mainstreaming: strategy for promoting gender equality*. https://www.un.org/womenwatch/osagi/pdf/factsheet1.pdf

United Nations. (2025). *The Sustainable Development Goals Report*. https://unstats.un.org/sdgs/report/2025/The-Sustainable-Development-Goals-Report-2025.pdf

Vinska, O., & Tokar, V. (2021). Cluster analysis of the european union gender equality and economic development. *Business, Management and Economics Engineering, 19*(2), 373–388. https://doi.org/10.3846/bmee.2021.15382

UNDP. (2021). *Technical Notes. Human Development Reports.* https://hdr.undp.org/sites/default/files/2021-22_HDR/hdr2021-22_technical_notes.pdf

United Nations. (2025). *The Gender Snapshot 2025*. https://reports.weforum.org/docs/WEF_GGGR_2025.pdf

Grant Thornton International. (2025). *Women in Business 2025 – Impacting the missed generation.*

Torres-Berru, Y., Lopez-Batista, V. F., & Zhingre, L. C. (2023). A Data Mining Approach to Detecting Bias and Favoritism in Public Procurement. *Intelligent Automation and Soft Computing, 36*(3), 3501–3516. https://doi.org/10.32604/iasc.2023.035367

UN Women. (2025). *Gender mainstreaming: strategy for promoting gender equality*. https://www.un.org/womenwatch/osagi/pdf/factsheet1.pdf

United Nations. (2025). *The Sustainable Development Goals Report*. https://unstats.un.org/sdgs/report/2025/The-Sustainable-Development-Goals-Report-2025.pdf

Vinska, O., & Tokar, V. (2021). Cluster analysis of the european union gender equality and economic development. *Business, Management and Economics Engineering, 19*(2), 373–388. https://doi.org/10.3846/bmee.2021.15382

Methodology for the comparative analysis of PCA and Autoencoders in dimensionality reduction: impact on classification accuracy and computational efficiency

Saša Mitrović, Neven Vrček
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{smitrovic, nvrcek}@foi.hr

Abstract. *This paper discusses an adapted CRISP-DM methodological framework for an automated comparison of the performance of dimensionality reduction methods — PCA, LDA, Factor Analysis and AE, t-SNE, Kernel PCA — in the context of multiclass and binary classification tasks using different datasets. The use of the adapted CRISP-DM methodology ensures an automated and systematic analysis in all phases of the CRISP-DM process, including data preparation, modelling and evaluation. The automation is achieved by dynamically determining the number of components according to the dimensional structure of the datasets by integrating dimensionality reduction methods and machine learning classification methods. The results within this study indicate that PCA and Autoencoder are effective dimensionality reduction techniques in linear and nonlinear space, respectively. Furthermore, the results show that, on average, PCA achieves higher classification accuracy with significantly lower computational effort, while AE shows advantages in higher classification accuracy of the nonlinear domain with the trade-off in low computational efficiency. The proposed approach enables a replicable, modular and computationally efficient evaluation of algorithms for analysing high-dimensional data and thus contributes to the improvement of automated systems in the field of intelligent data analysis. A Python implementation of the adapted CRISP-DM methodology for the automated comparison of the performance of two dimensionality reduction methods can be found at the following GitHub link: https://github.com/smitroviefos/amf_dim_red_paper2025* .

Keywords. CRISP-DM, PCA, AutoEncoders, Dimensionality reduction, classification performance, computational efficiency

1 Introduction

Given the growing importance of data-intensive intelligent systems in today's environment, especially in the field of analysing and processing big data, the high dimensionality of input features poses a significant challenge to the efficiency and interpretability of machine learning algorithms.(Baratchi et al., 2024) Dimensionality reduction as a key stage of data processing reduces model complexity, speeds up execution and improves the robustness of classification and regression systems. The most common techniques include principal component analysis (PCA)(Bartal et al., 2019; Cohen, 2017; Geron, 2019; Jolliffe, 2002; Van Der Maaten et al., 2009) and autoencoders (AE)(Chandra, 2024; Cohen, 2017; Goodfellow et al., 2016), which approach the problem from two different perspectives — linear and nonlinear.

The PCA method represents a projection of the data into a space with lower dimensionality, preserving the highest possible variance, while autoencoders, which are based on multilayer neural networks, allow the modelling of non-linear relationships between features, thus ensuring greater informativeness, but require more computational resources. Although there is scientific literature comparing these approaches applied to different tasks (Fournier & Aloise, 2019; Mayur Prakashrao Gore, 2024), a systematic analysis combining quantitative and qualitative metrics within a single methodological framework is still relatively limited in the SCOPUS database to the best of the authors' knowledge based on the literature review presented below. Therefore, this paper aims to explore the advantages and disadvantages of PCA and AE as representative techniques of linear and nonlinear dimensionality reduction, respectively, within the context of common machine learning classification methods. The primary objective is to analyse the strengths, limitations, and applicability of these methods across different data domains by embedding them into a fully automated analytical pipeline based on an adapted CRISP-DM (Cross-Industry Standard Process for Data Mining) methodological framework. Additionally, research paper compares linear methods (PCA, Linear Discriminant Analysis – LDA, and Factor Analysis) separately from nonlinear methods (Autoencoders, t-Distributed Stochastic Neighbour Embedding – t-SNE, and Kernel PCA), the study incorporates an extended evaluation that organises

dimensionality reduction algorithms into these two conceptual categories – linear and nonlinear. While the primary analytical emphasis remains on PCA and AE due to their widespread use, business domain-independent datasets and widely adopted classifiers are utilised to evaluate model performance through multiple quantitative metrics, including accuracy, precision, recall, F1 score, and ROC/AUC. This approach enables the assessment of generalisation capability, information preservation, and execution efficiency of each method, with the ultimate goal of contributing to a more structured and comparative understanding of dimensionality reduction techniques in automated machine learning pipelines.

2 Literature review

In analysing the existing scientific literature, the present paper focuses on the research gap identified by a number of authors discussing dimensionality reduction, PCA and AE methods. Specifically, Mendes Junior et al. (2020) point out that in previous research they found that the comparison of feature selection and dimensionality reduction methods was not sufficiently systematic and quantitatively sound. In particular, Mendes Junior et al. (2020) found that the application of dimensionality reduction techniques after feature selection often did not lead to statistically significant differences in the accuracy of machine learning classification methods. Furthermore, Mendes Junior et al. (2020) point out that the research gap also includes a limited number of comparative analyses of dimensionality reduction techniques and the lack of comprehensive evaluations in different problem domains, indicating the need for more methodologically rigorous approaches that cover a broader range of data, models and performance metrics.(Mendes Junior et al., 2020) Vantuch et al. (2016) point out that future research should focus on the problem of reconstruction methods that inadequately reconstruct the original datasets after applying dimensionality reduction, which can lead to information loss. Similarly, Vantuch et al. (2016) emphasise that the analysis of information loss and uncertainties that occur during transformation is insufficiently discussed in the literature, although these aspects can have a significant impact on the reliability of the model. Vantuch et al. (2016) therefore point out that there is a need for empirical comparisons that quantitatively evaluate the advantages of different dimensionality reduction algorithms in terms of maintaining data representativeness and the interpretability of the results.(Vantuch et al., 2016) In their research, Ghobadi et al. (2023) point out the need to adjust hyperparameters in dimensionality reduction methods to achieve optimal results in the unique data domain contexts. Since different datasets can have markedly heterogenous features — in terms of noise levels, correlations between features or structural complexity — fixed and manual approaches without automation of parameter adaptation often lead to suboptimal transformations. Ghobadi et al. (2023) therefore emphasise the importance of dynamic and context-sensitive parameter adjustments to increase model accuracy and maintain the semantic representation of the data in low-dimensional space.(Ghobadi et al., 2023)

In the present work, the above shortcoming is addressed using a systematic and methodologically structured approach that integrates a dynamic selection of the number of components, automated model evaluation and statistical validation of performance differences, thus contributing to the understanding of the effectiveness and applicability of modern dimensionality reduction techniques in scientific and applied contexts.

3 Methodology

In order to conduct a systematic analysis, the present authors used an approach based on the adaptation of the CRISP-DM methodology (Cross-Industry Standard Process for Data Mining), which provides a comprehensive framework for conducting data analyses in six phases: problem understanding, data understanding, data preparation, modelling, evaluation, and implementation.(Bratkovsky, 2024; Chapman, 2000; Costa, 2022; Shearer, 2000)

3.1 Understanding the research problem

The main objective of this study is to compare two dimensionality reduction methods: as a linear method and autoencoders (AE) as a nonlinear, neural method. Dimensionality reduction is a crucial step in the processing and analysis of high-dimensional data, i.e. in the context of this work for classification tasks where the high complexity of the input space can negatively affect the generalisation capability of machine learning models. The comparison of PCA and AE is made in terms of classification accuracy, execution time, computational efficiency and preservation of the semantic structure of the data in the latent space. In addition, other dimensionality reduction methods are used, structured into two conceptual groups - linear and nonlinear. The linear methods include PCA, Linear Discriminant Analysis (LDA), and Factor Analysis (FA), all of which operate under the assumption of linear transformations and rely on variance maximisation or class separability criteria (Pedregosa et al., 2011). In contrast, the nonlinear methods comprise Autoencoders, t-Distributed Stochastic Neighbor Embedding (t-SNE), and Kernel PCA, each capable of modelling non-linear manifolds and capturing more complex intrinsic data structures through techniques such as deep representation learning or kernel-based transformations(Chandra, 2024; Cohen, 2017; Schölkopf et al., 1997). While the

core analytical focus remains on PCA and AE due to their conceptual contrast and practical relevance in both academic research papers and various business domain settings, the inclusion of other methods enables a more comprehensive and structured benchmarking process. The comparison is made in terms of classification accuracy, execution time, computational efficiency and preservation of the semantic structure of the data in the latent space.

3.2 Data understanding

The Phoneme (OpenML ID: 1489), Vehicle Silhouettes (OpenML ID: 54), Blood Transfusion Service Center (OpenML ID: 1464) and the HighDim (Gina agnostic, OpenML ID: 40978) datasets from the OpenML repository were used in this study as they represent different application domains and different feature structures, thus enabling an evaluation of the generalisation capability of dimensionality reduction algorithms. The Phoneme dataset consists of 5,404 instances and 5 acoustic features, containing the sound features of speech phonemes and is primarily used for speech recognition tasks; Vehicle dataset includes 846 instances with 18 geometrical features, containing geometric attributes of vehicles classified by type; Blood Transfusion Service Center dataset consists of 748 instances and 4 features containing demographic and behavioural data of blood donors with the aim of predicting their propensity to re-donate, while HighDim dataset is a high-dimensional binary classification handwritten digit as image recognition dataset with 1,400 instances and 970 features, constructed to evaluate algorithmic scalability and behaviour in high-dimensional spaces. This diversity provides a relevant experimental context to test the effectiveness of the methods within an automated, adapted CRISP-DM methodological framework.

3.3 Data preparation

The data were first normalised to the interval 0 to 1 to reduce numerical differences between features and ensure model stability. The data were stratified to divide them into training and test datasets in a ratio of 80:20. As dimensionality reduction methods can be sensitive to the size and distribution of the data, special care is taken to maintain class balance when splitting a dataset with an extremely high-class imbalance. In all cases - PCA, LDA, Factor Analysis and AE, t-SNE, Kernel PCA - the number of components and the dimension of the layer are dynamically determined as a minimum between the preset threshold (in the Python code the threshold is set to 10), the number of features and the number of instances in the dataset. Such an approach ensures the methodological validity and flexibility of the algorithms when applied to heterogenous datasets, while maintaining the structural consistency of the model. This enables the automated application of dimensionality reduction techniques in the CRISP-DM modelling and evaluation phase, regardless of the unique data domain context.

3.4 Modelling

Following dimensionality reduction methods were used in the modelling phase: (1) PCA, LDA, and Factor Analysis are the linear techniques methods. Methods were implemented from the scikit-learn library (Pedregosa et al., 2011).(Bartal et al., 2019; Cohen, 2017; Geron, 2019; Jolliffe, 2002; Van Der Maaten et al., 2009) (2) The autoencoder is an neural network with a hidden layer of dimension 64 implemented in the TensorFlow (Abadi et al., 2016) environment. The mean squared error (MSE)(Santosh et al., 2022), is used for the error function, while the Adam algorithm with an initial learning parameter of 0.001 is used for optimisation.(Chandra, 2024; Cohen, 2017; Goodfellow et al., 2016). Other nonlinear techniques such as t-SNE(van der Maaten & Hinton, 2008) and Kernel PCA (Schölkopf et al., 1997) are also used in the modelling phase.

After reduction, identical machine learning classification methods LogisticRegression (s.-l. developers, 2024a), RandomForest(s.-l. developers, 2024b), SVM(s.-l. developers, 2024c) and XGBoost(x. developers, 2024), are trained on the obtained latent representations to isolate the effects of the reduction method on the classification performance.

3.5 Evaluation

The evaluation of the model was based on several criteria: classification accuracy, execution time and visualisation of the reduced dimensionality using bar charts. The qualitative analysis of the visual representations provides information about the separability of the classes in the latent space. The quantitative results were analysed using McNemar's statistical test for dependent samples of binary classification to determine the significance of the differences between the accuracies of the machine learning methods (Perktold et al., 2025b). The quantitative results for dependent samples of multiclassification were analysed using Cochran's Q test (Perktold et al., 2025a). In addition, the data reconstruction ability of the autoencoder was compared to evaluate the information loss.

3.6 Implementation

The entire experiment was implemented in the Python programming language. The code is available at the GitHub link and uses the externally referenced libraries scikit-learn [17], tensorflow (Abadi et al., 2016), matplotlib(Barrett et al., 2005) and numpy(Harris et al., 2020). The programme code is modular and reproducible, and it is possible to extend the analysis to additional datasets and dimensionality reduction methods such as UMAP (Healy & McInnes, 2024).

The results show that PCA retains its advantage in terms of computational efficiency, while AE shows superiority in preserving complex structures and higher classification accuracy at lower dimensions. The conclusions obtained make a scientific contribution to the understanding of the trade-off between linearity and non-linearity in dimensionality reduction in the context of application to different domains of business data.

4 Results and discussion

The experiment uses an adapted CRISP-DM methodology to automate the process of intelligent data analysis and compares the effects of linear and nonlinear dimensionality reduction methods – PCA, LDA, Factor Analysis and AE, t-SNE, Kernel PCA – in combination with four classification methods LogisticRegression (s.-l. developers, 2024a), RandomForest(s.-l. developers, 2024b), SVM(s.-l. developers, 2024c) and XGBoost(x. developers, 2024) in the modelling phase. The evaluation was carried out with heterogenous datasets that differ in the number of classes, the distribution of features and the semantic domain. The aim was to assess how the choice of dimensionality reduction method affects the classification efficiency and computational complexity of the model.

Table 1. Accuracy per Datasets

DR Method	LogReg	RF	SVM	XGBoost
Dataset **Blood**				
AE	0.7467	**0.7667**	**0.7667**	0.74
FA	0.7467	0.7533	**0.76**	0.7333
KPCA	0.76	0.72	**0.7667**	0.7333
LDA	**0.7533**	0.6533	0.7467	0.74
PCA	0.7467	0.7467	**0.7667**	0.7267
t-SNE	**0.7667**	0.7333	0.7733	0.6867
Dataset **HighDim**				
AE	0.7522	0.7594	**0.7695**	0.7392
FA	0.7954	0.7896	**0.8112**	0.7954
KPCA	0.7896	0.7997	**0.8228**	0.7925
LDA	0.8228	0.8069	**0.8228**	0.8156
PCA	0.8055	**0.8228**	**0.8228**	0.8084
t-SNE	0.6643	**0.7752**	0.745	0.7233
Dataset **Phoneme**				
AE	0.7336	0.8705	0.8437	**0.8751**
FA	0.741	**0.8668**	0.7974	0.8548
KPCA	0.7512	**0.8742**	0.8039	0.8511
LDA	0.7373	0.7391	**0.7623**	0.7586
PCA	0.7364	**0.8788**	0.8326	0.864
t-SNE	0.7188	**0.8696**	0.778	0.8686
Dataset **Vehicle**				
AE	0.6588	0.6059	0.6765	**0.7059**
FA	0.7765	0.7588	**0.7882**	0.7765
KPCA	0.6412	0.6647	**0.6765**	0.6706
LDA	**0.5941**	0.5118	0.5882	0.5529
PCA	0.6765	**0.7353**	0.7176	0.7059
t-SNE	0.4882	**0.6882**	0.6	0.6353

The results, as shown in Table 1, show several important findings. Within the group of linear dimensionality reduction methods, accuracy results have shown that Principal Component Analysis (PCA) is the most consistent and effective dimensionality reduction technique. In datasets Phoneme and HighDim, PCA yielded very high classification accuracy, particularly when used in combination with RandomForest and XGBoost classifiers. Furthermore, PCA demonstrated extremely low computational complexity, making it highly suitable for integration into automated analytical pipelines within the CRISP-DM methodological framework. In contrast, Linear Discriminant Analysis (LDA) and Factor Analysis (FA) exhibited lower robustness across most experimental scenarios. Although they achieved satisfactory results on the HighDim dataset, their average classification accuracy was inferior to that achieved by PCA. Method LDA has limitations due to its restrictions in handling multiclass problems and the requirement for class labels during the reduction phase, which constrains its flexibility of application in automated systems aligned with the CRISP-DM methodology.

Within the group of nonlinear methods, accuracy results have shown that Autoencoder is the most consistent and effective technique to dimensionality reduction, achieving high classification accuracy and consistent performance across all datasets. While the Autoencoder provides high predictive accuracy, such performance is accompanied by longer reduction times compared to linear methods. This trade-off is justified by its ability to preserve nonlinear relationships in the data, thus enabling a better representation of complex structures in the latent space (Chandra, 2024). Given its accuracy, the Autoencoder positions itself as an exceptional tool for automating data processing pipeline within the CRISP-DM cycle, particularly when dealing with datasets exhibiting nonlinear distributions. The t-SNE method showed acceptable accuracy on the Phoneme dataset but also demonstrated high variability and substantial computational overhead. It is important to emphasize that t-SNE was limited to three components due to scalability constraints, attempts with higher dimensions failed to complete even after several hours of execution. Despite its usefulness for visualization, the application of t-SNE in classification-oriented workflows, such as those structured by CRISP-DM, remains methodologically limited. KernelPCA, as a nonlinear extension of PCA, produced solid results, particularly on the HighDim dataset. Its capacity to preserve complex relationships in the data without excessive computational burden makes it a

methodologically balanced choice, although it does not surpass the Autoencoder in overall effectiveness.

The accuracy results within this study indicate that PCA and Autoencoder are effective strategies for dimensionality reduction in linear and nonlinear space, respectively. Method PCA offers an efficient and fast linear transformation suitable for a wide range of problems, while the Autoencoder provides superior quality for complex and nonlinear datasets. Based on the experimental findings and within the context of developing automated systems aligned with the CRISP-DM methodology, the authors of this study recommend the combined use of PCA and Autoencoders, taking into consideration the trade-off between computational efficiency and the quality of data representation in lower-dimensional space.

Table 2. Reduction execution time per Datasets

DR Method	LogReg	RF	SVM	XGBoost
Dataset **Blood**				
AE	4.4284	4.4284	4.4284	4.4284
FA	0.004	0.004	0.004	0.004
KPCA	0.0198	0.0198	0.0198	0.0198
LDA	0.001	0.001	0.001	0.001
PCA	0.0011	0.0011	0.0011	0.0011
t-SNE	5.8834	5.8834	5.8834	5.8834
Dataset **HighDim**				
AE	9.1207	9.1207	9.1207	9.1207
FA	0.1198	0.1198	0.1198	0.1198
KPCA	0.9979	0.9979	0.9979	0.9979
LDA	0.1953	0.1953	0.1953	0.1953
PCA	0.0346	0.0346	0.0346	0.0346
t-SNE	58.4865	58.487	58.487	58.4865
Dataset **Phoneme**				
AE	10.0798	10.08	10.08	10.0798
FA	0.0057	0.0057	0.0057	0.0057
KPCA	1.1999	1.1999	1.1999	1.1999
LDA	0.002	0.002	0.002	0.002
PCA	0.0303	0.0303	0.0303	0.0303
t-SNE	67.0619	67.062	67.062	67.0619
Dataset **Vehicle**				
AE	4.6146	4.6146	4.6146	4.6146
FA	0.2431	0.2431	0.2431	0.2431
KPCA	0.0439	0.0439	0.0439	0.0439
LDA	0.002	0.002	0.002	0.002
PCA	0.002	0.002	0.002	0.002
t-SNE	9.3425	9.3425	9.3425	9.3425

The reduction execution time results, as shown in Table 2, for various dimensionality reduction techniques across four experimental datasets (Blood, HighDim, Phoneme, and Vehicle) provide insight into the computational complexity of each method, which is a crucial consideration when designing automated systems based on the CRISP-DM methodological framework. The linear method PCA (Principal Component Analysis) demonstrates consistently low computational complexity across all datasets (with execution times ranging from approximately 0.001 to 0.03 seconds), making it a methodologically and computationally optimal option for dimensionality reduction in scenarios involving large data volumes and the need for rapid processing. Similarly as PCA, LDA (Linear Discriminant Analysis) and FA (Factor Analysis) also achieve very low reduction execution times, however, it should be noted that their applicability depends on the characteristics of the dataset, such as the classification type problem number of classes and the linearity of relationships among features.

Within the group of nonlinear methods the method Kernel PCA (KPCA) requires slightly more computational time (from 0.04 seconds on the Vehicle dataset to 1.20 seconds on the Phoneme dataset), but still remains within acceptable thresholds for integration into automated data pipelines, particularly when there is a need to preserve nonlinear relationships without excessive computational cost.

The Autoencoder (AE) method consistently shows slower execution times across all datasets, ranging from approximately 4.4 seconds (Blood and Vehicle) to over 10 seconds (Phoneme), and nearly 9.1 seconds on the high-dimensional HighDim dataset. This level of low computational efficiency reflects the fact that AE involves multiple data passes during training, including parameter optimization and input reconstruction, making it more suitable for scenarios where latency is not the primary constraint, but where preserving complex nonlinear data structures is critical.

The most computationally inefficient method is t-SNE (t-distributed Stochastic Neighbor Embedding) with only three components, with reduction times exceeding several seconds in all datasets, and reaching up to 67 and 58 seconds in the Phoneme and HighDim datasets, respectively.

The reduction execution runtime per Datasets results emphasizes linear methods (PCA, LDA, FA) as computationally efficient dimensionality reduction methods and nonlinear methods (AE, KPCA, t-SNE) as more resource-intensive dimensionality reduction methods. Autoencoders are better in preserving the semantic structure of the data, with the cost in their execution time since it is considerably longer compared to PCA. This implies the need to balance accuracy and computational cost, preferably with objective function, when designing an optimal automated machine learning CRISP-DM-based process. Method PCA should be applied in time-sensitive applications, while AE and KPCA may be more suitable for applications involving complex nonlinearities and sufficient computational resources.

The results comparing linear and nonlinear methods by analysing the average values of performance metrics—accuracy, precision, recall, F1 score, and the area under the ROC curve (ROC/AUC)—it is observed that linear methods (PCA,

LDA, and FA) consistently outperform nonlinear approaches (Autoencoder, t-SNE, and KernelPCA) across all evaluated metrics.

Table 3. Group Averages: Linear and Nonlinear methods

Metrics	Group	
	Linear	Nonlinear
Accuracy	0.7572	0.7426
Precision	0.747	0.7296
Recall	0.7572	0.7426
F1 Score	0.7467	0.73
ROC/AUC	0.8909	0.8755

Linear methods achieved an average accuracy of 0.7572, whereas nonlinear methods yielded a slightly lower value of 0.7426. This pattern is consistent across other metrics as well: precision (0.747 vs. 0.7296), recall (0.7572 vs. 0.7426), F1 score (0.7467 vs. 0.73), and ROC/AUC (0.8909 vs. 0.8755). These results indicate a higher level of consistency and robustness of the linear approach, especially in structured classification tasks where relationships among features are predominantly linear or can be effectively approximated through linear transformations.

From a methodological perspective, linear methods benefit from low computational complexity, stable behavior across datasets of varying sizes, and transparent interpretability of the transformations. Although nonlinear methods are better in preserving complex relationships in the data, their higher variability and computational inefficiency efficiency may limit their applicability in automated machine learning systems. Therefore, the authors findings suggest that linear methods—particularly PCA—may be considered both methodologically and operationally optimal for most conventional classification problems within the CRISP-DM framework, whereas nonlinear methods are more suitable in scenarios where nonlinear structures are known to exist or where representation quality outweighs computational efficiency constraints.

The results of the statistical significance tests performed provide additional information on the differences between the classification models in connection with the different methods of dimensionality reduction. McNemar's test for binary classification tasks and Cochran's Q test for multiclass datasets are applied. Particularly when the PCA method was used with RandomForest, SVM, and XGBoost classifiers, a sizable number of statistically significant differences ($p < 0.05$) were found within the Phoneme dataset between models based on logistic regression (LogReg) and other classifiers. With a p-value of 0.0000 from the comparison between LogReg PCA and RandomForest PCA, the distribution of predictions was clearly highly significantly changed. Conversely, the matching p-values were 1.0000 and 0.9011 respectively when LogReg PCA was compared with LogReg LDA or RandomForest LDA. This suggests that models using different dimensionality reduction methods yet belonging to the same classifier family have not much variation. These findings imply that rather than the single impact of any one of these components, the combination of the classifier and the reduction technique determines the classification performance in major importance. Methodologically, these results validate the need of formal statistical validation inside the evaluation stage of the CRISP-DM framework. In this regard, McNemar's test helps to identify consistent prediction variations at the level of individual instances, so improving the dependability of choosing the most suitable combination of classifier and dimensionality reduction technique for application in automated analytical pipelines. Finally, the author’s findings imply that combinations like RandomForest PCA and XGBoost PCA differ greatly from simpler models like LogReg PCA, which methodologically supports their usage in more difficult classification tasks. All results, together with the Python code, are also available at the GitHub link.

5 Conclusion

This paper provides a comparative analysis of methodologically different dimensionality reduction techniques — PCA, LDA, Factor Analysis and AE, t-SNE, Kernel PCA — for multiclass and binary classification tasks on different datasets. The use of the adapted CRISP-DM methodology ensured a transparent implementation of all phases of the data processing and analysis process, from problem understanding to model implementation.

The results obtained confirm that the choice of dimensionality reduction method significantly influences the classification efficiency and computational feasibility of the model. The PCA method proves to be robust, stable and computationally efficient and is particularly suitable for classification tasks, while the AE method may have advantages in certain cases, e.g. for non-linear data structures. The combination of metrics and execution time provides a multidimensional evaluation of the model and thus contributes to a more comprehensive understanding of the application of dimensionality reduction methods in a production environment. The proposed automated approach, based on the CRISP-DM methodology, has been shown to be suitable and extensible for complex evaluations of machine learning models related to dimensionality reduction. Future research could aim to extend the analysis to other dimensionality reduction methods such as UMAP, as well as to develop and integrate new dimensionality reduction approaches and methods into the AutoML system to further extend the

applicability and scientific contribution of the proposed methodology.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., & Zhang, X. (2016). TensorFlow: A system for large-scale machine learning. https://doi.org/10.48550/arXiv.1605.08695

Baratchi, M., Wang, C., Limmer, S., van Rijn, J. N., Hoos, H., Bäck, T., & Olhofer, M. (2024). Automated machine learning: past, present and future [Article]. Artificial Intelligence Review, 57(5), Article 122. https://doi.org/10.1007/s10462-024-10726-1

Barrett, P., Hunter, J., Miller, J. T., Hsu, J. C., & Greenfield, P. (2005). matplotlib -- A Portable Python Plotting Package.

Bartal, Y., Fandina, N., & Neiman, O. (2019). Dimensionality reduction: theoretical perspective on practical measures. Neural Information Processing Systems,

Bratkovsky, E. (2024, 10.12.2024). A simple explanation of CRISP-ML for beginners. https://www.kaggle.com/discussions/general/487787

Chandra, S. B. (2024). Introduction to Autoencoders [Theory and Implementation]. Medium. Retrieved 2.7.2024. from https://medium.com/@c17hawke/introduction-to-autoencoders-theory-and-implementation-72fed7cf2f70

Chapman, P. (2000). CRISP-DM 1.0: Step-by-step Data Mining Guide. SPSS. https://books.google.hr/books?id=po7FtgAACAAJ

Cohen, E. (2017). Reducing Dimensionality from Dimensionality Reduction Techniques. Medium. Retrieved 2.7.2024. from https://medium.com/towards-data-science/reducing-dimensionality-from-dimensionality-reduction-techniques-f658aec24dfe

Costa, R. (2022). The CRISP-ML Methodology: A Step-by-Step Approach to Real-World Machine Learning Projects. Independently published.

developers, s.-l. (2024a). LogisticRegression. scikit learn. Retrieved 6.6.2024. from https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

developers, s.-l. (2024b). RandomForestClassifier. scikit learn. Retrieved 6.6.2024. from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

developers, s.-l. (2024c). Support Vector Machines. scikit learn. Retrieved 6.6.2024. from https://scikit-learn.org/stable/modules/svm.html

developers, x. (2024). XGBoost. xgboost developers. Retrieved 6.6.2024. from https://xgboost.readthedocs.io/en/stable/

Fournier, Q., & Aloise, D. (2019, 3-5 June 2019). Empirical Comparison between Autoencoders and Traditional Dimensionality Reduction Methods. 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE),

Geron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc.

Ghobadi, F., Tayerani Charmchi, A. S., & Kang, D. (2023). Feature Extraction from Satellite-Derived Hydroclimate Data: Assessing Impacts on Various Neural Networks for Multi-Step Ahead Streamflow Prediction [Article]. Sustainability (Switzerland), 15(22), Article 15761. https://doi.org/10.3390/su152215761

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://books.google.hr/books?id=omivDQAAQBAJ

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

Healy, J., & McInnes, L. (2024). Uniform manifold approximation and projection. Nature Reviews Methods Primers, 4(1), 82. https://doi.org/10.1038/s43586-024-00363-x

Jolliffe, I. T. (2002). Principal Component Analysis. Springer. https://books.google.hr/books?id=_olByCrhjwIC

Mayur Prakashrao Gore, A. C. A. A. S. (2024). Applying-Principal-Component-Analysis-and-Autoencoders-for-Dimensionality-Reduction-in-Data-Stream. International Journal of Innovative Research in Engineering & Management (IJIREM), 11(5), 121-126. https://doi.org/10.55524/ijirem.2024.11.5.17

Mendes Junior, J. J. A., Freitas, M. L. B., Siqueira, H. V., Lazzaretti, A. E., Pichorim, S. F., & Stevan, S. L. (2020). Feature selection and dimensionality reduction: An extensive comparison in hand gesture classification by sEMG in eight channels armband approach [Article]. Biomedical Signal Processing and Control, 59, Article 101920. https://doi.org/10.1016/j.bspc.2020.101920

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825--2830.

Perktold, J., Seabold, S., Taylor, J., & statsmodels-developers. (2025a). statsmodels.stats.contingency_tables.cochrans_q. statsmodels. Retrieved 6.5.2025. from https://www.statsmodels.org/dev/generated/statsmodels.stats.contingency_tables.cochrans_q.html

Perktold, J., Seabold, S., Taylor, J., & statsmodels-developers. (2025b). statsmodels.stats.contingency_tables.mcnemar. statsmodels. Retrieved 6.5.2025. from https://www.statsmodels.org/dev/generated/statsmodels.stats.contingency_tables.mcnemar.html

Santosh, K. C., Das, N., & Ghosh, S. (2022). Chapter 2 - Deep learning: a review. In K. C. Santosh, N. Das, & S. Ghosh (Eds.), Deep Learning Models for Medical Imaging (pp. 29-63). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-823504-1.00012-X

Schölkopf, B., Smola, A., & Müller, K.-R. (1997). Kernel principal component analysis. In W. Gerstner, A. Germond, M. Hasler, & J.-D. Nicoud, Artificial Neural Networks — ICANN'97 Berlin, Heidelberg.

Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. Journal of Data Warehousing, 5(4).

van der Maaten, L., & Hinton, G. (2008). Viualizing data using t-SNE. Journal of Machine Learning Research, 9, 2579-2605.

Van Der Maaten, L., Postma, E., & Van den Herik, J. (2009). Dimensionality reduction: a comparative review. J Mach Learn Res, 10, 66--71.

Vantuch, T., Snasel, V., & Zelinka, I. (2016). Dimensionality Reduction Method's Comparison Based on Statistical Dependencies. Procedia Computer Science.

From Black Box to User Insight: Explainability and Usability in Toxicity Detection

Oliver Lohaj
Department of Cybernetics and Artificial Intelligence
Faculty of Electrical Engineering and Informatics
Technical university of Košice, Letná 9, 040 01, Košice, Slovakia
oliver.lohaj@tuke.sk

Anastasiia Radishevska
Department of Cybernetics and Artificial Intelligence
Faculty of Electrical Engineering and Informatics
Technical university of Košice, Letná 9, 040 01, Košice, Slovakia
anastasiia.radishevska@student.tuke.sk

Abstract*. This study investigates the detection of toxic content in text data by integrating deep learning models with explainable artificial intelligence (XAI) techniques, with a particular focus on model transparency and usability. We evaluate three widely used neural architectures—CNN, LSTM, and BERT—on a labeled Twitter dataset, comparing their classification performance and the interpretability of their outputs. To enhance model explainability, we apply SHAP and Layer-wise Relevance Propagation (LRP) methods, visualizing word-level contributions to each prediction. The usability of these models is assessed through the clarity and reliability of their explanations. Our results show that while LSTM achieved the best overall classification performance, the combination of SHAP with LSTM provided the most interpretable and actionable insights. This work highlights the trade-offs between accuracy, explainability, and usability in toxicity detection, offering practical guidance for deploying trustworthy AI systems in content moderation.*

Keywords. BERT, classification, CNN, detection of toxicity, explainable AI, LRP, LSTM, SHAP

1 Introduction

This work addresses the detection of toxic content in text data, combining deep learning with explainability techniques. It begins with a theoretical overview and analysis of existing approaches. The main goal is to apply and evaluate selected models and explainability methods to better understand the decision-making of neural networks. Finally, all methods are assessed and compared using standard evaluation metrics.

This work also builds on previous research presented at CECIIS 2024, where we explored the usability challenges of integrating multiple data sources for toxic behavior detection in social media (Lohaj et al., 2024). While that study focused on the broader context of combining heterogeneous data inputs to support detection models, the present work narrows the focus to the explainability and usability of the models themselves. Specifically, we aim to make the decision-making processes of individual deep learning models more transparent and actionable for end users. This progression reflects a shift from system-level integration concerns to model-level interpretability, aligning with the growing need for trustworthy and user-centered AI systems in content moderation. While our dataset originates from sentiment analysis, we frame negative sentiment as indicative of toxic content for the purposes of model training and evaluation. This framing enables us to assess toxicity detection methods using sentiment-labelled social media data, while maintaining focus on the broader societal relevance of mitigating harmful online interactions.

Toxicity detection plays a critical role in protecting individuals from online harassment, hate speech, and other harmful behaviors. As online communication increasingly shapes public discourse, detecting and mitigating toxic content is essential for maintaining respectful dialogue, preventing psychological harm, and supporting the work of moderators on social media platforms. By focusing on explainability and usability, this study addresses not only the accuracy of detection but also the trust and accountability required for societal adoption of AI moderation tools.

2 Related Works

2.1 Text classification

Several studies have addressed toxic content classification in online environments. One notable work (Grine, 2021) analyzed various methods, including SVM (Support Vector Machine), CNN (Convolutional Neural Network), and LSTM (Long-Short-Term Memory Network), to identify toxic comments and evaluate their performance. The main

challenges tackled in the study were the lack of multi-label support and significant class imbalance.

The models were trained and tested on the Conversation AI "Wikipedia Talk" dataset, containing nearly 160,000 comments labelled across six categories: toxic, severe toxic, obscene, threat, insult, and identity hate. The study revealed a substantial imbalance in label distribution, which negatively influenced the models' performance on minority classes. A correlation matrix also showed strong co-occurrence between classes such as toxic and insult or obscene (e.g., r = 0.74 for insult–obscene).

Evaluation based on Precision, Recall, and F1-score metrics demonstrated that CNN outperformed others in most categories, especially for the minority class threat. SVM achieved the best precision, while LSTM showed slightly better recall. All models performed better on majority classes, with F1-scores above 80%, and worse on minority ones (47–73%).

Learning curve analysis further showed that CNN reached strong performance with only 2% of training data, outperforming both SVM and LSTM early in training. LSTM required more data (up to 25%) to surpass SVM, with diminishing returns observed beyond 80% of the dataset.

Another study by authors (Anand & Enswari, 2019) focused on CNN and LSTM models, comparing their training behavior using two key metrics: training accuracy and training loss. The models were trained on a dataset from Wikipedia's talk page edits. In this dataset, there are almost 160.000 comments and labelled with different categories some of the comments belong to more than one category. The visualized results showed that CNN achieved rapid accuracy gains during training, consistent with previously reported trends for convolutional architectures in text classification. CNN achieved a rapid increase in accuracy, reaching 97.8%, with a low training loss of 5.42%. In contrast, LSTM demonstrated lower accuracy and higher loss, indicating slower convergence and potentially less effective learning dynamics.

In another publication by (Maslej-Krešňáková et al., 2020), the authors examined and compared traditional deep learning models (FFNN (Feedforward Neural Network), CNN, GRU (Gated Recurrent Unit), BiGRU (Bidirectional Gated Recurrent Unit), BiLSTM-CNN (Bidirectional Long Short-Term Memory)) and transformer-based language models (BERT, DistilBERT, XLNet) under various preprocessing and text representation techniques. The experiments were conducted on the Kaggle Toxic Comment Classification dataset, which training data contains almost 160.000 Wikipedia comments labelled across six toxicity categories. The dataset exhibits a strong imbalance toward non-toxic classes, making it a challenging benchmark for classification models. The experiments revealed that the combined BiLSTM-CNN model achieved the best results among the traditional architectures, with an F1 score of approximately 0.67. When TF-IDF (Term Frequency-Inverse Document Frequency) text representation was used along with standard preprocessing techniques—tokenization, lowercasing, punctuation removal, and stop word elimination—performance significantly improved. The F1 score increased from 0.09 to 0.35. Furthermore, in the comparison of transformer-based models, the BERT-base (uncased) variant achieved the highest performance, with an F1 score of 0.69 and an AUC score of 0.984, outperforming both other transformer variants and traditional models. Study by (Ansar et al., 2024) reviews new strategies for optimizing transformer-based NLP models for faster inference and lower resource use, including pruning, quantization, and low-rank adaptation and supports the transformer advancement and efficiency. Recent work by (Wu et al., 2025) highlights how transformer-based architectures, such as BERT and GPT, have redefined state-of-the-art performance in text understanding by effectively handling long-range dependencies and complex contextual relationships. Their methodological framework and insights into efficiency optimization could inform the integration of advanced NLP techniques into similar AI-driven applications discussed in the present study.

2.2 Using explainable artificial intelligence in text classification

Authors (Nguyen et al., 2024) compared three popular XAI techniques: LIME (Local Interpretable Model Agnostic Explanations), SHAP (SHapley Additive exPlanations), and CAM (Class Activation Mapping), and proposed enhancements for their evaluation and application. Their stability analysis showed that LIME and SHAP produce consistent results with 1,000 samples, but become less reliable with 500 or 200 samples, indicating that a higher sample count improves result stability and accuracy.

Each method had distinct advantages: CAM offered the fastest computation, while SHAP uniquely identified both positive and negative feature contributions. However, LIME and CAM are limited to classification tasks, and CAM is image-specific. In contrast, SHAP demonstrated broader applicability across data types and tasks.

In another paper by (Gholizadeh & Zhou, 2021) the authors applied LRP (Layer-wise Relevance Propagation) to explain SVM-based review classification. They used a dataset of over 229,000 customer reviews, reformulated as binary classification. To analyze the SVM predictions, a CNN model was trained to replicate SVM outputs, using pre-trained word embeddings and multiple convolutional filter sizes. After five epochs (batch size 30), the CNN achieved high agreement with SVM predictions (F1 score of 0.93).

LRP highlighted the most relevant tokens influencing classifications and visualized them for interpretability. This approach demonstrated that LRP + CNN can

effectively interpret machine learning models and offer actionable insights for improving business decisions based on customer feedback.

3 Deep Learning Methods

Deep learning has become a dominant approach in many natural language processing (NLP) tasks due to its ability to automatically extract complex features from large datasets. Among the wide variety of deep learning architectures, certain models have demonstrated exceptional performance in text classification and interpretability research. This section provides a concise overview of three models: CNN, LSTM, and BERT.

3.1 Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) (O'Shea, 2015), originally designed for image processing, have been adapted to text classification to capture local n-gram patterns using convolutional filters. A typical CNN for text includes convolutional and activation layers (e.g., ReLU) to detect patterns, pooling layers to reduce dimensionality, and fully connected layers for classification. CNNs are valued for their computational efficiency and ability to capture short-range dependencies in text.

3.2 Long Short-Term Memory Network (LSTM)

LSTM networks (DiPietro, 2020) are designed to capture long-term dependencies in sequential data by using a memory cell and three gates (forget, input, output) to regulate information flow. This architecture overcomes the vanishing gradient problem in traditional RNNs and is particularly effective for tasks requiring contextual understanding, such as sentiment analysis and speech recognition.

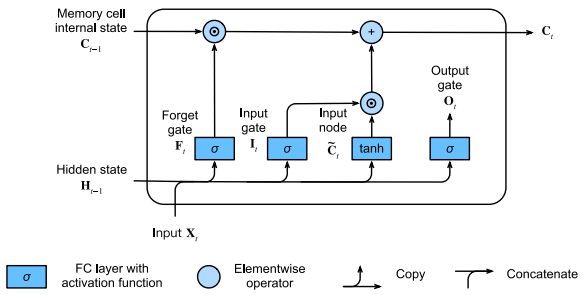


Figure 1. LSTM architecture.[1]

3.3 Bidirectional Encoder Representations from Transformers (BERT)

BERT is a transformer-based model that utilizes a bidirectional attention mechanism to understand the context of a word based on its surrounding words, regardless of their position (Sun et al., 2019). Unlike traditional left-to-right models, BERT processes text in both directions simultaneously.

The training process includes:

- Masked Language Modeling (MLM): predicting masked words in a sentence.
- Next Sentence Prediction (NSP): determining whether two sentences follow each other.

After pre-training, BERT is fine-tuned on specific NLP tasks using smaller, domain-specific datasets. Its strong contextual understanding has made it state-of-the-art in various classification and interpretation tasks.

4 Explainable AI Methods

Modern machine learning models, especially deep learning architectures, are often referred to as black box (Awati & Yasar, 2024) due to their complex internal structures that make it difficult to understand how input data leads to a particular output. This lack of transparency poses significant challenges in domains where trust, accountability, or legal compliance are crucial. To address this issue, XAI (Ryo, 2022) methods have been developed to provide insights into model behavior and decision-making. We can count the following methods as frequently used:

- SHapley Additive exPlanations (SHAP)

This approach is based on game theory and uses Shapley values to evaluate the contribution of individual factors to the model's outcome (Atesli, 2023). Since computing exact Shapley values is computationally expensive, the method approximates them to provide an understanding of how each factor contributes to the prediction.

- Layer-wise Relevance Propagation (LRP)

LRP is a model-specific method that provides detailed explanations of predictions, especially for deep neural networks (Praveen, 2021). This process analyzes which parts of the input data—such as image pixels or words in text—contributed the most to a given prediction. Relevance is propagated backward through the layers of the neural network, helping to identify the inputs that had the greatest impact on the model's result.

[1] https://d2l.ai/chapter_recurrent-modern/lstm.html

5 Data Understanding and Preparation

5.1 Data Description

This section provides an overview of the dataset used in this study, focusing on its structure, quality, and potential issues that may impact model accuracy. Understanding the data is crucial for optimizing preprocessing and designing effective models.

The dataset consists of tweets from Twitter\Hugging Face in English, divided into three subsets:

- Training set: 12.9k tweets
- Test set: 3.7k tweets
- Validation set: 1.85k tweets

The dataset has two attributes:

- Text – the tweet content
- Sentiment – numeric attribute (1 for Positive, 0 for Negative) representing the sentiment of the tweet.

Each row represents a tweet with its assigned sentiment. The sentiment labels in this dataset were provided by the original dataset creators on the Hugging Face platform. Annotation was performed through a combination of manual review and automated classification heuristics, with 0 representing a negative sentiment and 1 representing a positive sentiment. The labeling guidelines ensured that tweets containing expressions of dissatisfaction, criticism, or negative tone were marked as negative, while those with expressions of approval, gratitude, or positive tone were marked as positive. Data quality checks revealed no missing values or duplicates, and the sentiment attribute contains only unique values (0, 1). In this study, tweets labeled with negative sentiment polarity (0) are treated as toxic content, while those with positive sentiment polarity (1) are considered non-toxic. This mapping allows us to apply toxicity detection techniques to a sentiment-labeled dataset, reflecting the assumption that negative sentiment often corresponds to toxic or harmful language in online discourse. Initial data analysis was performed through visualizations to better understand the dataset's structure. The training set contains approximately 7,000 positive and 6,000 negative tweets, showing a slight imbalance (see Fig. 1).

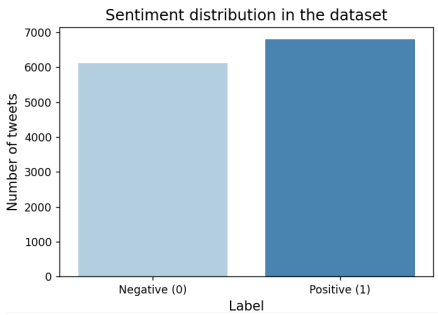


Figure 2. Visualizing sentiment distribution in the Twitter dataset

The analysis of the tweet lengths revealed that the average tweet contains 70.17 characters (see Fig. 2). Interestingly, there was little difference in length between positive and negative tweets, with the average lengths being 70.13 characters for positive tweets and 70.21 characters for negative ones. This suggests that there is no significant correlation between tweet length and sentiment in the dataset.

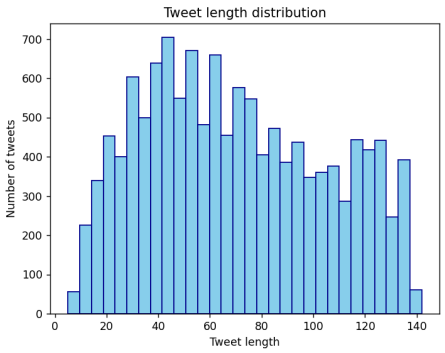


Figure 3. Tweet length distribution in the Twitter dataset

Additionally, the average word count per tweet was 13.24 (see Fig. 3), with slightly higher counts for negative tweets (13.43 words) compared to positive ones (13.07 words). This indicates that tweet content does not appear to vary significantly in terms of word count between the two sentiment classes.

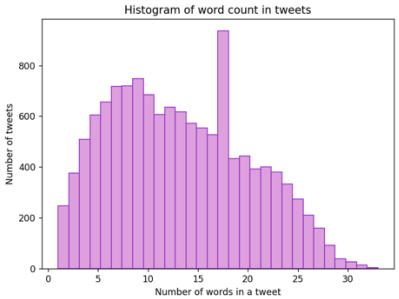


Figure 4. Number of words in tweets in the Twitter dataset

One particularly insightful visualization was the word cloud in Fig. 4, which was used to display the most frequent words in the tweets. This graphical representation highlighted words like "love", "thank", and "good", which were used frequently in the dataset. In addition to these common words, the word cloud also revealed recurring fragments such as "S", "m", "ll", "im", and "t", which likely represent truncated words or abbreviations. These patterns provide valuable insights for the next steps in preprocessing, where handling these truncated forms and abbreviations could improve model performance and ensure better text representation.



Figure 5. Most used words in tweets

5.2 Data Pre-Processing

This stage focuses on transforming raw tweets into a clean, standardized format suitable for analysis and modeling. Tweets are often noisy — containing inconsistent casing, numbers, excessive punctuation, links, usernames, and other irrelevant elements — which can affect model performance.

To address this, a custom preprocess function was implemented. It performs the following key steps:

- Normalization – converts all text to lowercase (e.g., "HAPPY DAY" → "happy day").
- Noise removal – eliminates URLs, emails, HTML tags, usernames (e.g., _user), numbers, punctuation, and extra spaces.
- Tokenization – splits text into individual words.
- Stop-word removal – removes common, uninformative words like "is", "a", "the".
- Lemmatization - converts words to their base forms (e.g., "took" → "take").

This process uses libraries such as *re* for pattern matching, *string* for punctuation handling, and *nltk* for tokenization, stop-word filtering, and lemmatization. As a result, the text becomes cleaner and better suited for machine learning models (see Table 1).

Table 1 Comparison of input and processed tweets

The initial form of tweets	Processed tweets
Just took my IC photo! Looks good - http://tweet.sg	just take ic photo look good
I'm so mad I won't be there!!!!!!!!! UGHHH!!!	mad will ughhh
Oh yes! Level 40	yes level
thank you! Marc Jacobs thou ....love limited too	thank marc jacob thou love limit

6 Modeling

6.1 Setting up models

To evaluate the effectiveness of different deep learning architectures in toxic content classification, three representative models were implemented: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and BERT (Bidirectional Encoder Representations from Transformers). Each model was configured with tailored architectures and training strategies, considering the unique characteristics of text data and model-specific requirements.

CNN Configuration

The CNN model was designed to capture local textual patterns through convolutional operations. It began with an Embedding layer of dimension 50, which transformed input tokens into dense vector representations. Tweets were padded to a fixed length to ensure consistent input shape across the dataset. A 1D Convolutional layer with 100 filters and a kernel size of 3 was used to extract tri-gram level features. Instead of a pooling layer, a Flatten layer was employed to retain the full spatial structure of the learned features. This was followed by a Dense layer with 128 units and ReLU activation, along with a Dropout layer (rate = 0.3) to mitigate overfitting. The final layer was a single-unit output layer with sigmoid activation, suitable for binary classification. The model was compiled with binary cross-entropy loss, Adam optimizer, and trained over 2 epochs with a batch size of 32.

LSTM Configuration

The LSTM model was employed to capture the sequential dependencies and contextual flow in the text. It also began with an Embedding layer (dimension = 50). A single-layer LSTM unit with 64 memory cells processed the input sequences, capturing both short- and long-term dependencies. The output of the LSTM layer was passed through a Dropout layer (rate = 0.4) to enhance generalization. A fully connected Dense layer with a sigmoid output completed the model architecture. The model was trained using the Adam optimizer and binary cross-entropy loss, with a batch size of 32 over 2 epochs. Token sequences were padded to a uniform length, and text preprocessing ensured consistent input representation.

BERT Configuration

The transformer-based BERT model was leveraged to take advantage of its powerful contextual language understanding. Specifically, the bert-base-uncased variant from Hugging Face's Transformers library was used. Tokenization was performed using BertTokenizerFast, which ensured compatibility with the model's vocabulary and token structure. The base model outputs were passed through a custom classification head comprising three Dense layers (768 → 512 → 256 → 2), with GELU activations and Dropout layers (rate = 0.2) in between for regularization. The final output was normalized using LogSoftmax to produce log probabilities for each class.

The model was fine-tuned for 9 epochs using the AdamW optimizer with a learning rate of 2e-5, and training was stabilized using a linear learning rate scheduler with warm-up steps. The batch size was set to 32.

The number of training epochs for CNN and LSTM was limited to 2 based on preliminary experiments in which both models reached near-optimal validation performance very quickly. Extending training beyond this point led to early signs of overfitting, with increased validation loss and minimal accuracy gains. This rapid convergence is partly due to the relatively small dataset size and the extensive preprocessing, which reduced noise and simplified the learning task. The chosen setting balances performance with computational efficiency.

The selected hyperparameters for CNN and LSTM (embedding size = 50, 100 filters with kernel size = 3, dropout rates of 0.3 and 0.4) follow commonly used configurations in toxic content and sentiment classification tasks on short text datasets (e.g., Anand & Eswari, 2019; Maslej-Krešňáková et al., 2020). These settings balance model complexity and computational efficiency, ensuring stable training without overfitting. The BERT learning rate (2e-5) was chosen in line with recommendations from Sun et al. (2019) for fine-tuning transformer-based models. Preliminary tests with slightly higher embedding dimensions and lower dropout rates did not yield notable performance improvements, so the original configuration was retained.

Each model was trained and evaluated under the same data conditions to ensure a fair comparison. Evaluation metrics included Accuracy, Precision, Recall, F1 score, and AUC, allowing for both predictive and probabilistic performance analysis.

6.2 Evaluation

After completing the modeling phase, the next critical step is the evaluation of the developed models. This phase focuses on analyzing their performance using selected metrics to assess their effectiveness in toxic text classification. The main objective is to determine which of the implemented models achieves the best results and is therefore most suitable for the task.

Three previously mentioned models were evaluated and compared in this study: CNN, LSTM, and BERT — each representing a different approach to text processing. Their evaluation results are summarized in Table 2.

Table 2 Comparison of the performance metrics of CNN, LSTM and BERT

	Accuracy	Precision	Recall	F1 score	AUC score
CNN	0.8492	0.8712	0.8376	0.8541	0.92
LSTM	**0.8568**	0.8580	**0.8726**	**0.8652**	**0.93**
BERT	0.8106	0.8100	0.8106	0.8102	0.89

Based on the comparison, the LSTM model achieved the highest overall performance and is recommended as the most suitable for the given classification task. The CNN model also showed strong results and can be considered an effective alternative. Although BERT performed slightly lower in terms of accuracy, it offers stable outputs and has the potential for improvement through further training optimization.

6.3 Application of the LRP XAI method

This section presents the application of the LRP method for explaining the decision-making processes of CNN and BERT models. To enable LRP, the final sigmoid layer was removed from the CNN model, as LRP requires access to pre-activation outputs. The *innvestigate* library was used, specifically the *"lrp.alpha_2_beta_1"* rule, which balances positive and negative contributions of neurons.

The method was applied to selected test examples. For each word in the input, LRP computed relevance scores, indicating its impact on the model's prediction. These scores were normalized using min-max scaling (0 to 1) and visualized using a blue gradient — darker blue indicating higher influence, white indicating none.

For example, in a correctly classified positive tweet, CNN assigned relevance scores such as:

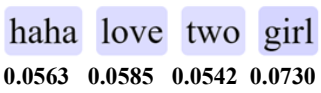


Figure 6. Relevance scores by CNN

In a negative example, LRP highlighted key influencing words with scores like:

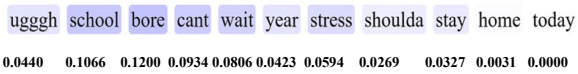


Figure 7. Negative example scores

For the BERT model, a modified version was created by removing the final LogSoftmax activation to enable analysis of raw model outputs. A custom *simple_lrp* implementation was used to compute relevance scores by backpropagating gradients through input embeddings.

Relevance scores were calculated for each token in a selected test example, indicating their contribution to the final classification. These scores were normalized to a [0,1] range using min-max scaling. A visualization was created using a colour scale — blue representing strong contributions, and red weak or no influence.

In a positive example, the token-level relevance scores were:



Figure 8. Token-level relevance scores

For a negative (toxic) input, the scores were:



Figure 9. Negative input scores

This visualization made it possible to identify the key tokens that influenced the classification decision. The method demonstrated that BERT relies on semantically relevant words, confirming that the model's decision-making is largely interpretable and transparent.

6.4 Application of the SHAP XAI method

Another explainability technique used in this work is SHAP, applied to interpret the predictions of the LSTM neural network model. The *KernelExplainer* from the SHAP library was used, enabling SHAP value computation for each word in the input tweet. Positive SHAP values indicated contribution toward the non-toxic class, while negative values supported the toxic class.

The explanations were visualized as horizontal bar plots, where each word was color-coded — blue for non-toxic influence and red for toxic influence. These visualizations provided clear insight into which words influenced the model's decisions.

Several examples from the test set were analyzed. In the first case, the LSTM correctly classified a toxic input. The SHAP values highlighted the words that strongly contributed to the toxic label.

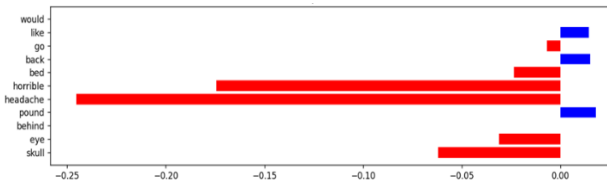


Figure 10. SHAP analysis for LSTM– 1st example

Another example showed a correctly classified non-toxic input:

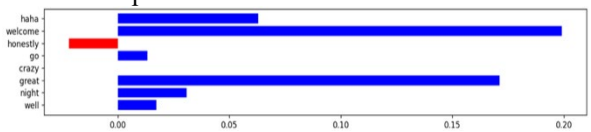


Figure 11. SHAP analysis for LSTM – 2nd example

A third example involved a misclassified input: the model predicted the toxic class with a probability of 0.2897, but the true label was non-toxic. SHAP values revealed which specific words misled the model, highlighting how even well-performing models can struggle with certain inputs due to misinterpretation of key terms.

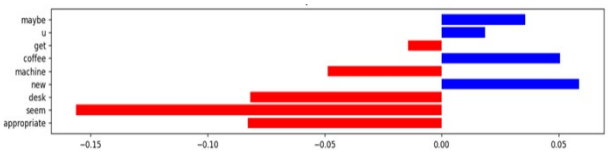


Figure 12. SHAP analysis for LSTM – 3rd example

This analysis demonstrated how SHAP helps uncover the reasoning behind both correct and incorrect classifications, enhancing the interpretability of the model.

In the context of this study, "usability" is used as a broader term that encompasses explainability, interpretability, and the clarity of the model outputs. Here, it refers to how easily a human moderator could understand and act upon the explanations provided by SHAP and LRP visualizations. While these analyses demonstrate interpretability potential, no formal user study or quantitative usability metrics were collected in this work.

7 Conclusion

This study demonstrates that integrating deep learning with explainable artificial intelligence (XAI) methods significantly enhances the transparency and usability of toxicity detection systems. By aligning model performance with transparency and usability, this work contributes to the development of AI systems that are not only effective but also trusted and responsible, supporting healthier online communities and reducing the impact of toxic discourse on individuals and society. Through comparative analysis of CNN, LSTM, and BERT architectures, we found that LSTM achieved the highest classification performance, while the application of SHAP and LRP offered valuable insights into model decision-making processes. These results underscore the necessity of balancing predictive accuracy with interpretability in applications where trust, fairness, and accountability are critical.

From a practical perspective, we recommend that developers and platform operators prioritize explainability when selecting models for content moderation tasks. Incorporating visual explanation tools—such as SHAP plots or LRP-based heatmaps—can support moderation teams in understanding and validating AI-driven decisions. Additionally, the early integration of user-centered evaluation into the model development process ensures that explanations are accessible and actionable for non-expert users. To maintain fairness and performance over time, toxicity detection systems should also incorporate continuous feedback mechanisms from end-users and moderators.

By aligning model performance with transparency and usability—understood here as a broader concept encompassing explainability and interpretability—this work contributes to the development of AI systems that are not only effective but also trusted and responsible. The usability evaluation in this study was based solely

on visual inspection of model explanations, without direct input from end users. As usability often depends on users' subjective perceptions and experiences, future work could benefit from incorporating structured user feedback to better assess how explanation visualizations support moderation tasks in practice.

Acknowledgments

This research was funded by the Slovak Research and Development Agency under the contract No. APVV-22-0414 and by the Scientific Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic and the Slovak Academy of Sciences under grant No. 1/0259/24.

References

Anand, M., & Eswari, R. (2019). *Classification of Abusive Comments in Social Media using Deep Learning*. IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8819734

Ansar, W., Goswami, S., & Chakrabarti, A. (2024). A survey on transformers in NLP with focus on efficiency. *arXiv:2406.16893*. https://doi.org/10.48550/arXiv.2406.16893

Atesli, H. (2023). *Explainable AI With SHAP*. Medium. https://medium.com/hepsiburada-data-science/explainable-ai-with-shap-6f629dfa6eef

Awati, R., & Yasar, K. *What is black box AI?* TechTarget. https://www.techtarget.com/whatis/definition/black-box-AI

DiPietro, R., & Hager, G. D. (2020). Deep learning: RNNs and LSTM. In *Handbook of medical image computing and computer assisted intervention* (pp. 503-519). Academic Press

Grine, A. E. (2023). *Toxic Comment Classification*. Medium. https://medium.com/@alaeddine.grine/toxic-comment-classification-317628632336

Lohaj, O., Paralič, J., & Buinytska, R. (2024). Exploring usability in combining data sources for detecting toxic behavior in social media. In *Central European Conference on Information and Intelligent Systems* (pp. 1-8). Faculty of Organization and Informatics Varazdin.

Maslej-Krešňáková, V., Sarnovský, M., Butka, P., & Machová, K. (2020). *Comparison of Deep Learning Models and Various Text Pre-Processing Techniques for the Toxic Comments Classification*. MDPI. https://www.mdpi.com/2076-3417/10/23/8631#B21-applsci-10-08631

Nguyen, H. T. T., Cao, H., Nguyen, V. T. K., & Pham, D. K. N. (2021). *Evaluation of Explainable Artificial Intelligence: SHAP, LIME, and CAM*. ResearchGate. https://www.researchgate.net/publication/362165633_Evaluation_of_Explainable_Artificial_Intelligence_SHAP_LIME_and_CAM

O'shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Praveen. (2021, March 1). *Overview of Explainable AI and Layer-wise relevance propagation (LRP)*. Medium. https://praveenkumar2909.medium.com/overview-of-explainable-ai-and-layer-wise-relevance-propagation-lrp-cb2d008fec57

Ryo, M. (2022). *Explainable artificial intelligence and interpretable machine learning for agricultural data analysis*. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2589721722000216

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019, October). How to fine-tune bert for text classification?. In *China national conference on Chinese computational linguistics* (pp. 194-206). Cham: Springer International Publishing.

Wu, T., Wang, Y., & Quach, N. (2025). Advancements in natural language processing: Exploring transformer-based architectures for text understanding. In *2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)* (pp. 1384-1388). IEEE.

Zhou, N., & Gholizadeh, S. (2021). *Model Explainability in Deep Learning Based Natural Language Processing*. arXiv. https://arxiv.org/abs/2106.07410

Smart Cuvette: An Integrated Optical System for Ethanol Measurement in Liquid Samples for Enological Applications

Ivana Kovačević, Ivan Aleksi, Tomislav Matić, Josip Zidar
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Department of Computer Engineering and Automation
Kneza Trpimira 2B, 31000 Osijek, Croatia
{ivana.kovacevic, ivan.aleksi, tomislav.matic1, josip.zidar}@ferit.hr

Abstract. *This paper presents the development and testing of a compact measurement system prototype, called Smart Cuvette, used to determine ethanol concentration in liquid samples. The device uses light-emitting diodes (LEDs) with different wavelengths as selective light sources, and photodetectors to monitor changes in light transmission through a cuvette containing the sample. The calibration was performed using ethanol-water mixtures over a wide concentration range. The developed sensor response showed a generally linear trend with increasing ethanol concentration, still accompanied by noticeable oscillations. Although no validation with real wine samples has been conducted at this stage, the system demonstrated consistent performance under controlled conditions. The results suggest that the prototype offers a modular and accessible approach for rapid ethanol estimation, with the potential for further applications in enology.*

Keywords. wine ethanol measurement, spectroscopy, optoelectronic sensor, NIR, transmittance measurement

1 Introduction

Winemaking is a complicated process, where every step is important for maintaining the quality of the final product. The process is divided into several phases where the fermentation is the most significant one and should be carefully monitored. During the fermentation process sugars, fructose, and glucose, are converted into ethanol, carbon dioxide, and glycerol by the action of yeast metabolism. Ethanol is the main product of wine fermentation, and its measuring is not only important for monitoring fermentation kinetics and progression but also for legal classification, product labeling, and quality control [1].

Various methods for measuring ethanol in liquid samples have been developed and applied both in research and in industry. Enzymatic methods are based on specific biochemical reactions, such as alcohol dehydrogenase (ADH), which converts ethanol into acetaldehyde that can be quantified spectrophotometrically. These methods are selective and relatively easy to perform but require expensive reagents and are sensitive to temperature and pH fluctuations [14]. Raman spectroscopy has been proposed as a non-invasive and reagent-free method for ethanol analysis due to its ability to identify the molecular vibrations characteristic of ethanol. However, its application in complex solutions, like wine can be challenging because of weak Raman scattering signals and significant fluorescence interference [17]. Refractive index analysis (RI) is another traditional method that takes advantage of the lower refractive index of ethanol compared to water. Although RI methods are simple and cost-effective, their precision decreases in the presence of other solutions and with temperature fluctuations [15]. Gas chromatography (GC) and high-performance liquid chromatography (HPLC) are still the standard methods due to their precision and ability to separate ethanol from other volatile or non-volatile compounds. However, both methods require expensive equipment, trained personnel and a considerable amount of time, which makes them impractical for continuous monitoring [13]. Densimetry, which is based on the change in liquid density with ethanol content, is widely used in commercial alcohol analyzers. However, it lacks specificity and can be inaccurate in complex mixtures due to interference from sugars and other dissolved substances [16]. An overview of ethanol measurement methods, along with their main advantages and disadvantages, is presented in Table 1.

In response to the limitations of conventional and laboratory-based techniques, authors in [2] proposed an optoelectronic sensor device that utilizes light absorption measurements to monitor ethanol concentration during grape must fermentation. Their system is based on near-infrared (NIR) spectroscopy with LEDs and photodetectors at specific wavelengths, demonstrating the possibility of real-time monitoring through absorbance changes. However, this proposed system lacks modularity and practical accessibility for small-scale producers. Authors in [18] have explored the potential of optoelectronic systems for monitoring key

Table 1. Overview of ethanol measurement methods with advantages and limitations

Method	Advantages	Limitations
Enzymatic (ADH-based)	Selective, relatively easy to perform	Requires expensive reagents, sensitive to temperature and pH fluctuations [14]
Raman Spectroscopy	Non-invasive, reagent-free, detects molecular vibrations	Limited applicability in complex mixtures [17]
Refractive Index (RI) Analysis	Simple, cost-effective	Low precision in mixed solutions and affected by temperature variations [15]
Gas Chromatography (GC)	High precision, capable of separating volatile compounds	Requires expensive equipment, trained personnel, and is time-consuming [13]
High-Performance Liquid Chromatography (HPLC)	High precision, separates non-volatile compounds	Requires expensive equipment, trained personnel, and is time-consuming [13]
Densimetry	Widely used, fast and straightforward	Lacks specificity, interference from sugars and other dissolved substances [16]

processes in winemaking. In their study, the authors developed an optoelectronic device for monitoring device for monitoring the maceration process of red wine. This sensor focused on capturing optical changes related to phenolic and color compound extraction, relying on NIR LED photodiode configurations and validating its performance through comparative testing. Additionally, the same authors outlined the system architecture and design considerations for their NIR-based ethanol monitoring prototype [19]. Their work highlights emitter selection, photodetector pairing, and signal processing required for accurate optical sensing in a winemaking context. These papers confirm the possibility of NIR-based systems for noninvasive ethanol monitoring, and possibilities for further enhancements in terms of open hardware design, cost-effectiveness, and scalability in different production environments.

In this paper, we present a prototype of the device for the determination of ethanol concentration in liquid samples. The proposed prototype is a compact optoelectronic device based on LED light transmission through a liquid sample, measured by a photodetector and processed via a microcontroller. Its modular design enables simple integration, portability, and real-time data acquisition. In this work, several tests are carried out with different levels of ethanol-water samples. The obtained results validate the purpose of the system and provide a foundation for further development and application of the device for the specific application in enology and process monitoring.

The rest of the paper is organized as follows. Section 2 describes the proposed prototype of a device for ethanol measurement. Section 3 presents the experimental analysis setup and discusses the results of testing. Section 4 concludes the paper and gives remarks for future work.

2 System development

2.1 Device prototype architecture

This paper proposes a prototype of an ethanol detection device that uses light transmission through a liquid sample to estimate the ethanol concentration. The system is based on the ESP32 microcontroller, which features integrated Wi-Fi, Bluetooth, and multiple digital interfaces for peripheral control. Three pairs of light-emitting diodes (LEDs) with wavelengths of $980\,nm$ (MTE9730CP), $1200\,nm$ (MTSM1200MT2-BK), and $1310\,nm$ (MTE1301C1) are used in the suggested prototype system. These wavelengths are selected due to their sensitivity to changes in the optical properties of ethanol-water mixtures in the near-infrared (NIR) region. Selected diodes provide reliable output within the NIR range, making them suitable for detecting subtle changes in transmission caused by varying ethanol concentrations [8] [9] [10]. The transmitted light is measured by photo diodes, which are sensitive to changes in light intensity and enable accurate detection of sample absorbance variations.

The measured sample is placed in a quartz cuvette, housed in a custom holder that ensures stable alignment between the light sources and the photodetector.

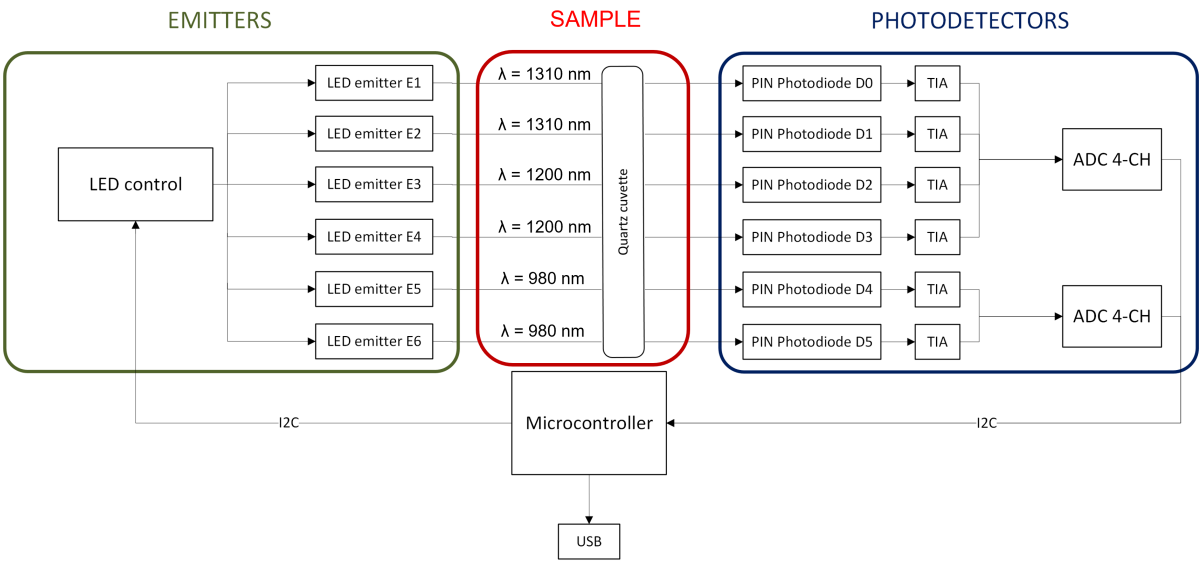


Figure 1. Architecture of the proposed prototype for ethanol measurement

A quartz cuvette is selected because of its high optical transparency in the near-infrared range, $190\,nm - 2500\,nm$, which enables accurate and repetitive measurements of ethanol at the wavelengths mentioned above [4].

The proposed prototype bases its measurements on the correlation between ethanol concentration and the optical transmission properties of ethanol-water samples in the NIR range. As the light emitted by LEDs passes through the sample contained in a quartz cuvette, it is partially absorbed and scattered depending on the chemical composition and concentration of ethanol. The part of the light that is not absorbed is transmitted through the sample and reaches the photodetector, an MTPD1309D3-030 [11] InGAs PIN photodiode, which provides high responsivity from 900 nm to 1700 nm and is placed directly opposite the light source.

The photodiode converts the transmitted light into a current through an internal photoelectric effect [3]. The magnitude of the generated photocurrent I_{ph} in (1) is directly proportional to the intensity of transmitted light P_{opt}, as expressed in (1):

$$I_{ph} = R \cdot P_{opt}, \tag{1}$$

where R is the responsivity of the photo diode. A higher light intensity results in a greater current, whereas increased absorption by ethanol reduces the transmitted light, leading to a lower value of I_{ph}, cf. (1). Because the photocurrent generated by the photodiode is typically in the microampere range, it must be converted into a voltage signal suitable for digital acquisition. This is achieved using a transimpedance amplifier (TIA) configuration, where the AD8601 precision operational amplifier converts a small current into a proportional voltage. The TIA ensures low noise amplification and maintains signal integrity [12]. The resulting analog voltage is then digitized by the ADS1115, a 16-bit analog-to-digital converter (ADC) which communicates with the microcontroller over an I^{2}C interface [20]. Diode switching is handled via the ULN2803 transistor array, which enables stable and sequential LED activation. Emitters E1 to E6 correspond to LED Emitters 1 to 6, while D0 to D5 denote the corresponding photodetectors aligned with each light source. The overall system architecture is illustrated in Fig. 1.

The sensor system transmits data via a serial communication interface using a UART module. During measurements, the microcontroller continuously sends sensor readings, one set for all six diodes per sampling cycle, over a USB connection. On the receiving side, a Python-based graphical interface was developed to handle data acquisition and visualization. The Python script establishes a serial connection to the microcontroller, reads incoming data lines, decodes them, and parses six values corresponding to the six photodetectors (D0-D5). The data collection loop was configured to capture 100 measurements per sample for each diode. Once this number was reached, the program automatically closed the serial connection and finalized the CSV file. This enables synchronized, consistent, and identifiable data logging across all sensor channels. This USB-UART-Python pipeline provides a reliable method for real-time analog sensor data acquisition from analog sensors, allowing easy recording and visualization of measurements.

The measured voltage corresponds to the amount of light that has passed through the sample, and it is typically expected to be inversely proportional to ethanol concentration due to increased absorbance. However, in our measurements, an opposite trend was observed at certain wavelengths. In our experiment, increased

ethanol concentration resulted in higher transmitted light and consequently higher voltage. The relationship between ethanol content and the signal is nonlinear, affected by a combination of refractive index changes, changes in the operating wavelength-dependent absorbance, and light dispersion within the sample. The elemental causes of this atypical response will be discussed in more detail in the following sections.

2.2 Data sampling

The developed prototype uses a microcontroller in combination with an external 16-bit analog-to-digital converter (ADC) to acquire photo diode signals corresponding to transmitted light intensity at each LED wavelength. The microcontroller collects raw digital values from the photo diodes without any on-board signal processing and transmits them directly to the PC via UART for further analysis. Sampling was performed at a fixed frequency of approximately $0.5\,Hz$, which was sufficient for capturing gradual changes in ethanol concentration under controlled conditions. This rate is limited by intentional delays for LED switching and signal stabilization. However, it can be increased to $5\,Hz - 10\,Hz$ by reducing these delays and optimizing the acquisition loop, depending on system noise and measurement stability.

The experimental procedure involved preparing a series of samples with ethanol concentration ranging from $0\,\%$ to $96\,\%$ (v/v), in $10\,\%$ increments. Each sample was prepared with $96\,\%$ pure ethanol with distilled water to achieve the appropriate solution concentration. The volume of ethanol and water required for each concentration using the standard dilution formula:

$$C_1 \cdot V_1 = C_2 \cdot V_2 \tag{2}$$

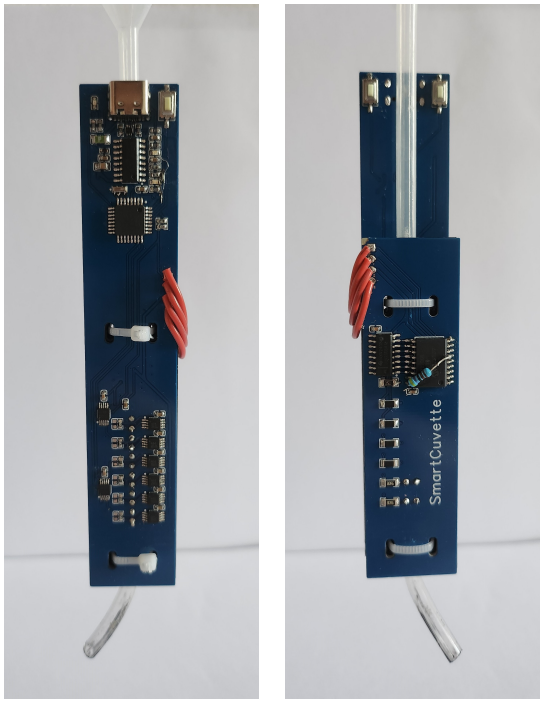
where C_1 is the initial ethanol concentration (96%), V_1 is the volume of ethanol used, C_2 is the target concentration, and V_2 is the final total volume of the prepared sample. This approach ensures consistency and the possibility of repetition across all tested concentrations, as recommended in standard analytical procedures for solution preparation [5].

All measurements were carried out inside a closed, black enclosure designed to eliminate the influence of external light. This type of repetitive sampling combined with averaging is frequently used in low-cost sensing systems to reduce the impact of electrical noise, mechanical disturbances, and ambient light fluctuations [6]. A detailed view of the prototype device, including the integrated measuring section with aligned LED and photo diode components, is shown in Fig. 2, illustrating both the front (a) and back (b) sides of the proposed system.

3 Data analysis and results

To evaluate the system's sensitivity to ethanol concentration, voltage signals from six photodetectors (D0-D5) each paired with a specific infrared LED source, were recorded for each prepared sample. The sensors were grouped by the wavelength of the light source, with D0 and D1 operating at $1310\,nm$, D2 and D3 at $1200\,nm$, and D4 and D5 at $980\,nm$. Fig. 3 compares the mean voltage values of all sensors, D0 to D5, and shows how the response varies depending on the wavelength. Significant variations in output can be observed between sensor pairings that operate at the same wavelength, such as D0 and D1. Although these sensors use identical components and are exposed to the same sample, the signal variation is likely caused by mechanical or optical asymmetries in the current prototype. This limitation affects the comparability of signals between certain channels and highlights the need for improved alignment and calibration in future iterations of the device. The measurement showed a clear relationship between ethanol concentration and output voltage, most notably in detectors operating at $1200\,nm$. Sensors D2 and D3 demonstrated a consistent increase in voltage with rising ethanol concentration up to $80\,\%$, followed by a slight decrease at the highest concentrations, which is shown in Fig. 4.

The nonlinear behavior observed in D2 and D3 sensors, especially the initial rise in voltage up to $80\,\%$ ethanol, followed by a slight drop, may be closely related to changes in hydrogen bonding and molec-

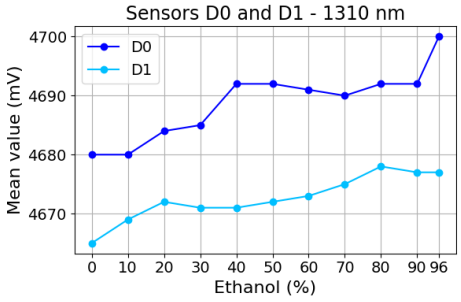


(a) **(b)**

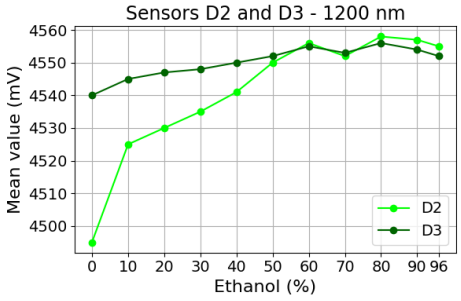
Figure 2. Front (a) and back (b) views of the device prototype.

ular structure within the liquid. According to [7], the strength of inter-molecular hydrogen bonds between ethanol and water peaks around $40\,\%$ ethanol concentration. Beyond this point, ethanol tends to self-associate, reducing the number of hydrogen bonds formed with water molecules.

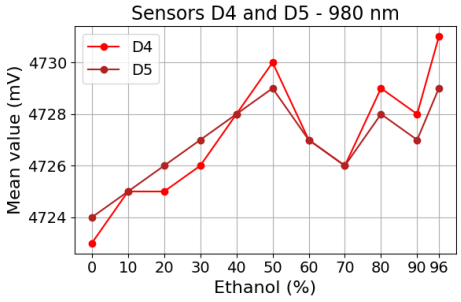
The shift in molecular interaction alters the NIR absorption profile of the sample. As ethanol concentration increases and ethanol-water hydrogen bonding decreases, the overall absorbance in specific NIR bands decreases. The reduced absorbance allows more light to pass through the sample, resulting in increased photodetector voltage, especially at wavelengths like 1200 nm, where ethanol displays relatively low absorbance. At very high concentrations, ethanol structures may absorb less effectively in the NIR range, contributing to signal variations. These results are promising for tested wavelengths, showing a sensitivity to ethanol concentrations. However, the response remains nonlinear in certain concentration ranges. Additional testing and analysis through a broader range of wavelengths and concentration intervals are needed to refine the calibration and better understand the system's behavior.



(a)



(b)



(c)

Figure 3. Mean voltage responses for all sensor pairs (a) D0 and D1 (b) D2 and D3 (c) D4 and D5 at different ethanol concentrations

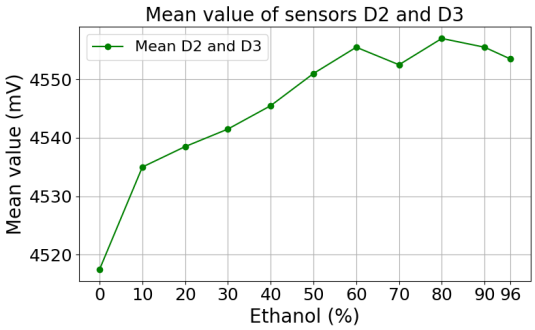


Figure 4. Mean output voltage of sensors D2 and D3

In addition to the previously mentioned nonlinear response patterns, the unexpected signal levels observed for sensors D4 and D5 at high ethanol concentrations may also be influenced by saturation effects or possible measurement errors in the acquisition system. These effects require further investigation to determine whether they stem from hardware limitations or optical phenomena at extreme concentration values.

This interpretation is by the obtained results and explains the observed signal rise as ethanol content increases. They confirm that optical sensor response is not driven only by ethanol concentration, but also by underlying molecular structure and interaction dynamics of the liquid.

4 Conclusion and Future work

In this paper, a prototype of a modular optoelectronic system was developed for determining ethanol concentration in liquid sampled using infrared LED photodetector pairs operating at $980\,nm$, $1200\,nm$, and $1310\,nm$. The system integrates analog signal acquisition through a microcontroller with real-time data visualization and storage via a custom Python interface. Measurements were performed on a series of ethanol samples, $0\,\% - 96\,\%$ v/v, prepared by dilution with distilled water, with each sample measured 100 times per sensor under controlled ambient conditions in the closed black enclosure.

Analysis of the recorded data showed that the sensors operating at $1200\,nm$ (D2) and (D3) provided the most stable response, exhibiting an increasing voltage trend with rising ethanol content. This behavior was attributed to refractive index changes, molecular changes, and light transmission that depends on wavelength. At high ethanol concentrations, a slight decrease in signal was observed, consistent with literature findings related to reduced hydrogen bonding and self-association of ethanol molecules.

The results confirm the sustainability of the proposed approach for ethanol quantification using low-complexity hardware and NIR sensor principles.

Future work will follow two main directions. First, a more comprehensive calibration model will be developed using nonlinear fitting or machine learning methods to better interpret sensor responses. Second, the system will be validated using real wine samples and benchmarked with reference methods. In addition, further experimental work will explore the device's performance at other NIR wavelengths. Extending the optical spectrum and concentration range will help reduce nonlinearity and broaden the application of the system in different ethanol detection scenarios.

Acknowledgments

This work is supported by the project Development of an expert system for food production and processing management (ESIA-Expert System for Intelligent Agriculture KK.01.1.1.07.0036) through the European Regional Development Fund as part of the Operational Programme for Competitiveness and Cohesion (OPCC) 2014–2020.

References

Jackson, R. S. (2020). Fermentation. In *Wine Science* (5th ed., pp. 461–572). Academic Press. https://doi.org/10.1016/B978-0-12-816118-0.00007-6

Jiménez-Márquez, F., Vázquez, J., Úbeda, J., & Sánchez-Rojas, J. L. (2016). Optoelectronic sensor for measuring ethanol content during grape must fermentation using NIR spectroscopy. *Microsystem Technologies, 22*(7), 1799–1809. https://doi.org/10.1007/s00542-016-2835-1

Sze, S. M., & Ng, K. K. (2021). Photodiodes – Photodetectors. In *Physics of Semiconductor Devices* (pp. 281–310). Wiley. https://doi.org/10.1002/9781119769958.ch5

Shimadzu Corporation. (2023). *Cuvette selection for UV-Vis and NIR measurements* (Application Note AN_01-00321-EN). Retrieved May 2025, from https://www.shimadzu.com/an/sites/shimadzu.com.an/files/pim/pim_document_file/applications/application_note/16675/an_01-00321-en.pdf

Skoog, D. A., West, D. M., Holler, F. J., & Crouch, S. R. (2013). *Fundamentals of Analytical Chemistry* (9th ed.). Cengage Learning.

Kedenburg, S., Vieweg, M., Gissibl, T., & Giessen, H. (2012). Linear refractive index and absorption measurements of nonlinear optical liquids in the visible and near-infrared spectral region. *Optical Materials Express, 2*, 1588–1611. https://doi.org/10.1364/OME.2.001588

Dong, Q., Yu, C., Li, L., Nie, L., Li, D., & Zang, H. (2019). Near-infrared spectroscopic study of molecular interaction in ethanol-water mixtures. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 222*, 117183. https://doi.org/10.1016/j.saa.2019.117183

Marktech Optoelectronics. (2019). *MTE9730CP 970nm Infrared Emitter Datasheet*. https://specs.marktechopto.com/pdf/products/datasheet/MTE9730CP.pdf

Marktech Optoelectronics. (2020). *MTSM1200MT2-BK 1200nm SWIR Emitter Datasheet*. https://specs.marktechopto.com/pdf/products/datasheet/MTSM1200MT2-BK.pdf

Marktech Optoelectronics. (2023). *MTE1301C1 1310nm SWIR Emitter Datasheet*. https://specs.marktechopto.com/pdf/products/datasheet/MTE1301C1_v103023_prelim-3368670.pdf

Marktech Optoelectronics. (2024). *MTPD1309D3-030 High-Speed InGaAs PIN Photodiode Datasheet*. https://www.mouser.com/datasheet/2/1094/MTPD1309D3_030_v040324_prelim-3483285.pdf

Analog Devices. (2019). *AD8601/AD8602/AD8604 Precision CMOS Rail-to-Rail Operational Amplifiers Datasheet*. https://www.analog.com/media/en/technical-documentation/data-sheets/ad8601_8602_8604.pdf

Rocchia, M., Rossi, A. M., & Zeppa, G. (2007). Determination of ethanol content in wine through a porous silicon oxide microcavity. *Sensors and Actuators B: Chemical, 123*(1), 89–93. https://doi.org/10.1016/j.snb.2006.07.025

Fernandes, E. N., & Reis, B. F. (2019). Automatic flow procedure for the determination of ethanol in wine exploiting multicommutation and enzymatic reaction with detection by chemiluminescence. *Journal of AOAC INTERNATIONAL, 87*(4), 920–926. https://doi.org/10.1093/jaoac/87.4.920

Gerogiannaki-Christopoulou, M., Kyriakidis, N. V., & Athanasopoulos, P. E. (2003). New refractive index method for measurement of alcoholic strength of small volume samples. *Journal of AOAC International, 86*(6), 1232–1235.

Brereton, P., Hasnip, S., Bertrand, A., Wittkowski, R., & Guillou, C. (2003). Analytical methods for the determination of spirit drinks. *TrAC Trends in Analytical Chemistry, 22*(1), 19–25. https://doi.org/10.1016/S0165-9936(03)00103-1

RDRS. (2017). *Advantages and disadvantages of Raman Spectroscopy*. https://rdrs.ro/blog/articles/advantages-disadvantages-raman-spectroscopy/(Accessed: May15,2025)

Jiménez-Márquez, F., Vázquez, J., & Sánchez-Rojas, J. L. (2015). Optoelectronic sensor device for monitoring the maceration of red wine: Design issues and validation. *Measurement, 63*, 128–136. https://doi.org/10.1016/j.measurement.2014.12.009

Jiménez-Márquez, F., Vázquez, J., Úbeda, J., Rodríguez-Rey, J., & Sánchez-Rojas, J. L. (2015). Optoelectronic sensor device for monitoring ethanol concentration in winemaking applications. In J. L. Sánchez-Rojas & R. Brama (Eds.), *Smart Sensors, Actuators, and MEMS VII; and Cyber Physical Systems* (Vol. 9517, p. 95171P). SPIE. https://doi.org/10.1117/12.2179341

Texas Instruments. (2020). *ADS1113, ADS1114, ADS1115 Ultra-Small, Low-Power, 16-Bit Analog-to-Digital Converters*. https://www.ti.com/lit/ds/symlink/ads1115.pdf(Accessed:May2025)

Impact of Imputation Methods on the Performance of Classification Algorithms

Bojan Radišić
Faculty of Tourism and Rural Development in Požega, Josip Juraj Strossmayer University of Osijek
Vukovarska 17, 34000 Požega, Croatia
bradisic@ftrr.hr

Ivan Dunđer, Sanja Seljan
Faculty of Humanities and Social Sciences, University of Zagreb, Department of Information and Communication Sciences
Ivana Lučića 3, 10000 Zagreb, Croatia
{idundjer, sseljan}@ffzg.unizg.hr

Abstract*. This paper analyses the impact of different imputation methods on the performance of classification algorithms. This was done to predict students' academic success. The original dataset consisted of student data from the Faculty of Tourism and Rural Development. Within this dataset, 46 missing values (around 10%) related to study duration were identified. Imputation methods, such as Random Forest, Gradient Boosted Trees, K-Nearest Neighbors, Multiple Imputation by Chained Equations and arithmetic mean imputation, were applied. Using these methods resulted in five distinct datasets that were used to train and evaluate the classification algorithms using Monte Carlo validation to ensure model assessment stability. For each iteration, classification performance metrics were calculated. A comparative analysis of all models provides insight into how imputation methods affect the performance of classification algorithms.*

Keywords. Machine learning, Data imputation, Missing data, Monte Carlo validation, Accuracy

1 Introduction

Educational Data Mining (EDM) has become an increasingly prominent area of research since its appearance in the literature in 2007 (Romero & Ventura, 2007). In Croatia, higher education institutions systematically collect data on students through the ISVU system (Cro. Informacijski sustav visokih učilišta, Eng. Information System of Higher Education Institutions), which serves as the central database for academic records.

The ISVU system supports the management of various types of academic information, including enrolment history, ECTS credit points, exam results and student progress indicators. When such data is analysed using machine learning techniques, it can yield valuable insights into patterns of academic achievement and student behaviour.

These insights are often used to improve educational strategies, identify students at risk of dropping out, and monitor trends in student retention and academic success.

In this research, data was obtained from ISVU for the period from 2010 to 2018. The dataset included instances of missing information on study duration. Such omissions are usually attributed to factors such as manual data entry errors, technical malfunctions, student non-response, cancellation of study programs, or merging of heterogeneous datasets (Emmanuel et al., 2021). The primary objective of this research is to evaluate how different imputation methods affect the predictive performance of machine learning classifiers in the context of student academic success.

Unlike earlier work (Radišić et al., 2023), which was limited to simple statistical imputations such as mean, median and geometric mean, this study extends the analysis to machine learning-based methods (Random Forest, Gradient Boosted Trees, K-Nearest Neighbors) and a hybrid approach (MICE). In addition, the models were evaluated using Monte Carlo cross-validation with multiple classification algorithms (Random Forest, Support Vector Machine, Naïve Bayes, K-Nearest Neighbors). This broader methodological scope provides new insights into how different imputation techniques interact with classifiers, demonstrating that the Random Forest algorithm remains the most robust and reliable across diverse imputation strategies.

The paper is organized as follows: after the Introduction, the second section provides a literature review. The third section, Methodology, contains a description of the dataset, an analysis of implemented machine learning algorithms, three main imputation strategies, and the testing process. In the fourth section, the results are presented, comparing the models of each individual algorithm and comparing all models among themselves, using accuracy and F1-score. The last section provides a conclusion and suggestions for future research.

2 Literature review

Research on predicting student success using machine learning has grown rapidly in the past decade, reflecting the increasing availability of educational data and the demand for evidence-based decision-making in higher education. Numerous studies have examined how student performance can be modelled, which algorithms are most effective, and what factors contribute to academic outcomes. A total of 438 articles were initially identified, and after a rigorous selection and quality assessment process, 70 studies published between 2018 and 2023 were included in the final review (Pelima et al., 2024). Their analysis showed that most studies focus on student performance prediction in higher education using machine learning techniques, but do not explicitly address how missing data affects predictive performance. This limitation highlights the importance of this particular study, which directly examines the role of imputation techniques in educational data mining.

Several machine learning algorithms, including Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Logistic Regression (LR), have proven to be effective in predicting academic success among students (Yağcı, 2022). In the study, performance on midterms emerged as a key predictor of outcome, i.e. academic success.

Similarly, (Campanilla, 2024) used a Naïve Bayes classifier to predict study completion. The results showed that out of 272 enrolled students, 200 of them (73.6%) had a higher grade point average (GPA) and a higher probability of completing their studies, while the remaining 72 (26.4%) were considered at risk for completing their studies. Based on these insights, the study advocated for a revision of institutional strategies to better address and reduce student attrition.

One model has been developed for the early identification of at-risk students using supervised machine learning techniques, including Support Vector Machine (Martinez et al., 2024). Trained on engagement, demographic and academic performance data, the model demonstrated high accuracy, with the SVM achieving 94% precision, confirming its effectiveness in early detection of academic risk.

A systematic review of 33 peer-reviewed studies published between 2010 and 2023 focused on the methods, datasets and frameworks used to handle missing values (Setiawan et al., 2023). The authors concluded that although no single imputation method is universally superior, statistical methods, K-Nearest Neighbors, Random Forest and hybrid frameworks are among the most commonly used, and their effectiveness is greatly influenced by the nature of the dataset and the objectives of the analysis.

The impact of four imputation techniques – Multivariate Imputation by Chained Equations (MICE), HMISC, Amelia and MissForest – on the predictive performance of eight supervised machine learning algorithms was investigated in a study focusing on the F1-score as the primary evaluation metric (Maale et al., 2025). The authors found that all imputation methods were sensitive to both proportion and mechanism of missingness (Missing Completely at Random – MCAR, Missing at Random – MAR, Missing Not at Random – MNAR), with MissForest achieving the highest average F1-score, while MICE performed consistently well, especially in MAR conditions.

Recent developments show that state-of-the-art (SOTA) imputation methods, especially those based on deep learning and adversarial networks, are being applied in data mining. A thorough review of GAN-based algorithms for imputation of missing data, has shown that they outperform traditional statistical and machine learning methods in accuracy (Shahbazian & Greco, 2023). Through experiments and analysis, the authors have demonstrated that GANs represent a state-of-the-art approach for handling incomplete datasets across various domains.

This particular study therefore contributes by systematically comparing classical, machine learning-based and hybrid imputations, providing a benchmark that can guide future integration of SOTA techniques in educational contexts.

3 Methodology

The data was collected from students of the Professional Study of Commerce at the Faculty of Tourism and Rural Development in Požega, Juraj Strossmayer University of Osijek, Croatia, from 2010 to 2018. Missing values were imputed using five methods: Random Forest (RF), Gradient Boosted Trees (GBT), N-Nearest Neighbors (KNN), MICE and arithmetic mean (Average).

After imputation, classification was performed using four machine learning algorithms: Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). Each algorithm with different imputation combination was evaluated using the Monte Carlo cross-validation method with 30 random and 70/30 train-test splits. All obtained models were compared based on accuracy and F1-score.

3.1. Dataset

The original dataset used in this research was collected from 461 students of the Professional Study of Commerce at the Faculty of Tourism and Rural Development in Požega, Juraj Strossmayer University of Osijek, Croatia.

The study program lasts 6 semesters over three academic years, and has a total of 180 ECTS credit points. The dataset includes all students enrolled from the academic years 2010/2011 to 2018/2019. Out of the total 461 students, there were 195 male students (42%) and 266 female students (58%). There were 264 full-

time students (57%) and 197 part-time students (43%). Eleven input variables, so-called data features, were selected for this study, as listed and described in Table 1.

The output variable focuses on the accumulated, i.e. achieved *ECTS credit points*, and is divided into three categories:

1. Full ECTS credit points (180 ECTS),
2. Achieved ECTS credit points (1-179 ECTS),
3. No ECTS credit points (0 ECTS).

Table 1. Data features

Features	Description and possible values
Special status	Admission based on special status
Enrolled	Part-time or Full-time students
Gender	Male or Female
County of residence	One of the 21 counties in Croatia
Residence in the city of Požega	Residence in Požega or not
Residence in Požega-Slavonia County	Residence in Požega-Slavonia County or not
Age	Student's age at the time of enrolment
High school	Type of completed high school
Duration of study	Time from enrolment to completion (graduation) or dropout
Points upon enrolment	Points when enrolling in a study program
Grade point average	Grade point average during study

The dataset includes three student categories: the first is comprised of 172 students who have earned all 180 ECTS credit points and completed their studies; the second group consists of 199 students who are still active, have passed certain courses, but have not accumulated all 180 ECTS credit points; and the third category includes 90 students classified as passive, as they have not earned any ECTS credit points.

The dataset was missing 46 values related to the *duration of study*. Therefore, they were imputed using five different methods for handling missing data: arithmetic mean, K-Nearest Neighbors (KNN), Multiple Imputation by Chained Equations (MICE), Random Forest (RF), and Gradient Boosted Trees (GBT).

To assess the consistency of the different techniques for imputing missing values, a correlation analysis was performed between four advanced imputation methods: KNN, MICE, RF and GBT. Pearson correlation was calculated for rows that originally contained missing values in the variable *duration of study*.

According to Table 2, it is evident that the highest correlation was observed between Random Forest and Gradient Boosted Trees (r = 0.969), which is expected given their similar architectures based on ensemble trees. The MICE method showed moderate correlation with both GBT (r = 0.554) and Random Forest (r = 0.515).

Table 2. Correlation matrix

		RF	GBT	MICE	KNN
RF	Pearson's r	—			
	df	—			
	p-value	—			
GBT	Pearson's r	0.969***	—		
	df	44	—		
	p-value	<.001	—		
MICE	Pearson's r	0.515***	0.554***	—	
	df	44	44	—	
	p-value	<.001	<.001	—	
KNN	Pearson's r	0.332*	0.364*	0.594***	—
	df	44	44	44	—
	p-value	0.024	0.013	<.001	—

Note: * p < .05, ** p < .01, *** p < .001

KNN showed a slightly higher correlation with MICE (r = 0.594), but relatively low correlations with Random Forest (r = 0.332) and GBT (r = 0.364), indicating that it relies on local proximity between samples, unlike tree-based methods.

The correlation analysis highlights strong agreement between tree-based machine learning models, particularly Random Forest and GBT. MICE and KNN, although algorithmically different, still exhibit moderate consistency with ensemble methods.

3.2. Machine learning

Machine learning techniques are often applied to predict academic success (achievements) among students. Key factors such as exam scores, prior educational performance, demographic details and class attendance play a significant role in determining study outcomes. Various machine learning models can be used for this predictive task, such as Decision Trees, Random Forest, K-Nearest Neighbors, Naïve Bayes and Neural Networks.

3.2.1. Random Forest (RF)

The Random Forest (RF) algorithm is widely applied in both classification and regression tasks due to its robustness and predictive power. It operates by generating multiple decision trees from various random subsets of the original dataset, and aggregates

their outputs, typically through averaging or majority voting, to improve overall prediction accuracy. Among its key strengths are the ability to automatically handle missing data, and its efficiency when applied to large and complex datasets.

Nevertheless, its use can be computationally demanding, requiring significant processing power and memory resources, especially when the number of trees or the amount of data increases (Kovač et al., 2022).

3.2.2. Naïve Bayes (NB)

Naïve Bayes (NB) is a type of classification algorithm based on Bayes' theorem, as stated below:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Naïve Bayes is a probabilistic classifier that assumes independence between features. Despite its simplicity, it is often competitive with more complex algorithms, especially in high-dimensional settings. In the educational context, Naïve Bayes has been effectively applied to predict student performance and academic success (Nakhipova et al., 2024).

3.2.3. K-Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm is a widely used nonparametric method primarily designed for classification, although it can also be applied to regression problems. It assigns a class to a new data point based on the majority votes of its k nearest (closest) neighbours within the training data, where proximity is typically measured by Euclidean distance or other distance metrics.

KNN is particularly valued for its intuitive logic and minimal model training requirements, making it simple to implement and adapt to different domains. One of the main advantages of KNN is that it does not require the construction of an explicit model or strong assumptions about the data distribution. This characteristic allows for flexible application, especially in scenarios involving proximity-based decision-making or pattern recognition (García et al., 2016).

3.2.4. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a widely used supervised machine learning algorithm designed for binary classification tasks. It works by constructing an optimal hyperplane that maximizes the margin between data points of different classes, which improves its generalization performance. One of its key strengths lies in its ability to handle both linearly and non-linearly separable data using kernel functions, which map the data into higher-dimensional spaces.

SVM is built on the principle of structural risk minimization, which contributes to its robust performance in various applications such as text classification, biomedical data analysis, financial modelling and others. However, the limitations of the method limitations include the high computational cost associated with solving large-scale quadratic programming problems, and sensitivity to kernel choice and parameter tuning (Tanveer et al., 2024).

3.3. Imputation of missing values

Addressing the problem of missing data is a key aspect of data preparation, especially in studies that rely on predictive modelling or statistical inference. Setiawan et al. (2023) provide a structured overview of imputation strategies, categorizing them into three core groups: statistical methods, machine learning-based methods and hybrid or ensemble approaches.

Statistical imputation methods are the simplest and most widely adopted, mainly due to their ease of use and minimal computational demands. These include methods such as replacing missing entries (values) with the mean, median or mode of the observed data.

Machine learning-based methods offer greater flexibility, especially when dealing with nonlinear interactions or multidimensional data. Algorithms such as KNN impute missing values based on the similarity between observations, while Random Forest-based imputation leverages ensemble decision trees to generate plausible (probabilistic) estimates. Similarly, GBT have demonstrated the ability to internally handle missing data during training, making them particularly useful in automated machine learning pipelines. These methods typically yield more accurate imputations, especially when the data structure is complex and not supported well by simple statistical heuristics.

Hybrid and ensemble approaches combine the strengths of different methods to increase robustness. One of the most well-known among them is MICE, which iteratively models each variable with missing values based on the others, thus capturing more nuanced dependencies. These approaches are particularly promising in large datasets where traditional imputation methods fall short.

3.4. Testing

From the original, i.e. initial dataset that contained missing values, five new datasets were created. In each dataset, missing values were imputed using different imputation strategies:

- statistical: arithmetic mean,
- based on machine learning: KNN, RF and GBT,
- hybrid and ensemble: MICE.

Four machine learning algorithms were applied to predict student academic success: RF, NB, SVM and KNN using all five datasets. Each algorithm was evaluated through the Monte Carlo cross-validation method with 30 random and 70/30 train-test splits.

There were 322 entries in the training set and 132 entries in the test set per iteration. A total of 20 different models were created.

4 Research results and discussion

All models were analysed both horizontally, by comparing models within each individual algorithm, and vertically, by comparing all models with each other. The comparisons focused on model accuracy and F1-score.

4.1. Accuracy

This section presents a comparative evaluation of the accuracy of four machine learning algorithms: KNN, NB, RF and SVM, based on their performance in different missing value imputation methods.

Table 3. KNN algorithm – accuracy metric

Imputation Method	Mean	Median	Standard Deviation	Minimum	Maximum
RF	0.491	0.489	0.0344	0.424	0.568
Average	0.493	0.493	0.0389	0.403	0.561
GBT	0.501	0.496	0.037	0.417	0.576
MICE	0.475	0.478	0.0318	0.396	0.532
KNN	0.489	0.489	0.0319	0.439	0.554

Table 3 shows that the highest average accuracy for the KNN algorithm was achieved using GBT imputation (Mean = 0.501), followed by Average (0.493) and RF (0.491). This suggests that GBT provides a small performance advantage for the KNN classifier. The lowest average accuracy was observed with MICE imputation (0.475), indicating that this method may not be as effective in preserving class-relevant information for the KNN algorithm.

In terms of variability, the lowest standard deviations were recorded for MICE (0.0318) and KNN imputation (0.0319), suggesting that these imputations yielded more consistent performance results, albeit with lower mean accuracy. For KNN algorithm, GBT imputation seems to achieve the best balance between performance and variability.

Table 4 shows that the RF algorithm demonstrated consistently high performance across all imputation methods. The highest average accuracy was recorded for KNN imputation (Mean = 0.810), followed by Average (0.806) and GBT (0.804). These results indicate that RF maintains robust predictive power regardless of the imputation method. In terms of consistency, GBT imputation yielded the lowest standard deviation (0.0208), indicating the most stable performance across iterations. The minimum and maximum values further highlight the effectiveness of RF. All methods achieved high minimum accuracy values ranged from 0.734 (Average) to 0.77 (GBT and KNN), and maximum values ranged from 0.842 (GBT) to 0.885 (Average), indicating the potential for near-optimal classification across different imputations.

Table 4. RF algorithm – accuracy metric

Imputation Method	Mean	Median	Standard Deviation	Minimum	Maximum
RF	0.8	0.799	0.0278	0.741	0.863
Average	0.806	0.809	0.035	0.734	0.885
GBT	0.804	0.802	0.0208	0.77	0.842
MICE	0.799	0.806	0.0242	0.755	0.871
KNN	0.81	0.809	0.025	0.77	0.863

Overall, RF combined with KNN imputation provided the best trade-off between high accuracy and acceptable variability. GBT offered the most consistent results, while Average achieved the highest peak performance.

The highest mean accuracy was observed with the Average imputation method (0.675), followed closely by GBT (0.674) and RF (0.671) for the NB algorithm, as shown in Table 5.

Table 5. NB algorithm – accuracy metric

Imputation Method	Mean	Median	Standard Deviation	Minimum	Maximum
RF	0.671	0.673	0.0422	0.583	0.748
Average	0.675	0.68	0.0476	0.59	0.763
GBT	0.674	0.68	0.053	0.496	0.763
MICE	0.659	0.669	0.0454	0.532	0.741
KNN	0.663	0.662	0.0541	0.561	0.763

This suggests that these three methods offer comparable overall predictive accuracy when used with the NB classifier. The highest performance variability was noted for GBT (SD = 0.0530) and KNN (SD = 0.0541), despite their relatively strong mean scores, implying that their effectiveness and performance may be more context-dependent or sensitive to data variation. Among the evaluated methods, Average imputation emerges as the most balanced option for the NB algorithm.

Table 6 shows that the highest average accuracy of the SVM algorithm was achieved with the RF imputation method (Mean = 0.713), indicating a slightly superior overall performance compared to other imputation methods.

Table 6. SVM algorithm – accuracy metric

Imputation Method	Mean	Median	Standard Deviation	Minimum	Maximum
RF	0.713	0.712	0.0347	0.633	0.777
Average	0.71	0.719	0.0355	0.626	0.755
GBT	0.706	0.709	0.029	0.655	0.77
MICE	0.695	0.701	0.0387	0.604	0.763
KNN	0.706	0.712	0.0441	0.583	0.791

This was followed closely by Average (0.710) and GBT/KNN (both 0.706), suggesting that all methods yielded relatively similar average results. Average imputation recorded the highest median value (0.719), although its mean was slightly lower than RF. This may indicate a more consistent central performance, with fewer lower outliers compared to other imputation methods. In terms of variability, GBT imputation had the lowest standard deviation (0.0290), indicating the most stable performance. The widest range of results was observed with KNN imputation, where the minimum accuracy dropped to 0.583, the lowest among all methods, but also achieved the highest maximum (0.791). This reinforces the interpretation that although KNN imputations may occasionally exhibit excellent performance, they also pose a risk of instability. For the most part, the SVM algorithm performs well across all imputation methods, with RF and Average imputations providing the best balance between accuracy and consistency.

Overall, the comparative analysis of the four classification algorithms (RF, SVM, NB and KNN) reveals clear differences in their classification accuracy and stability when applied to datasets with imputed values. RF emerged as the most accurate and stable algorithm, achieving average accuracy above 0.80 across all imputation methods. Support Vector Machine ranked second, with average accuracy scores ranging from 0.695 to 0.713, and showed strong consistency. Naïve Bayes showed moderate performance, with average accuracy results between 0.659 and 0.675, depending on the imputation method. Finally, K-Nearest Neighbors consistently yielded the lowest accuracy, with average scores below 0.50. Despite occasional improvements (e.g. with GBT), the algorithm exhibited high sensitivity to data variations and instability, making it the least suitable option among the evaluated models.

4.2. F1-score

This section presents a comparative evaluation of four machine learning algorithms (KNN, NB, RF and SVM) for the F1-score, based on their performance across different missing value imputation methods.

Table 7. KNN algorithm – F1-score metric

Imputation Method	Mean	Median	Standard Deviation	Minimum	Maximum
RF	0.444	0.500	0.134	0.111	0.621
Average	0.448	0.494	0.127	0.111	0.632
GBT	0.454	0.495	0.133	0.093	0.623
MICE	0.431	0.475	0.123	0.167	0.602
KNN	0.448	0.492	0.126	0.093	0.641

Table 7 shows that the highest mean F1-score (0.454) was obtained using GBT imputation, suggesting a modest performance benefit of this approach when applied with the KNN classification algorithm. In contrast, MICE imputation resulted in the lowest average F1-score (0.431), indicating a relatively less favourable outcome. The greatest variability in performance was observed with KNN imputation (Min = 0.093, Max = 0.641), which may reflect increased sensitivity of the model to specific data characteristics or imputation methods. In general, GBT imputation appears to provide the most favourable trade-off between predictive accuracy and consistency for the KNN classifier.

Table 8 demonstrates that the highest average F1-score was achieved using KNN imputation (0.839), which may indicate a slight advantage in the performance of the RF algorithm when combined with this method. Notably, all imputation methods reached a maximum F1-score of 1.000, suggesting that, under certain conditions, perfect classification was attainable regardless of the method applied. Furthermore, KNN

imputation produced the lowest standard deviation (0.0999), indicating the most consistent model performance.

Across all imputation methods, standard deviations remained relatively low (≤ 0.107), supporting the conclusion that the RF algorithm maintained a high level of classification stability regardless of the chosen imputation method. These results highlight the robustness and reliability of the RF classifier across different data preprocessing scenarios.

Table 8. RF algorithm – F1-score metric

Imputation Method	Mean	Median	Standard Deviation	Minimum	Maximum
RF	0.828	0.786	0.107	0.638	1.000
Average	0.832	0.792	0.105	0.621	1.000
GBT	0.831	0.789	0.106	0.667	1.000
MICE	0.828	0.780	0.107	0.674	1.000
KNN	0.839	0.790	0.0999	0.695	1.000

The analysis results presented in Table 9 show that the highest average F1-score for the NB algorithm was achieved using the KNN imputation method (0.700), followed closely by the GBT imputation method, which recorded a score of 0.697. This indicates strong classification performance associated with these two imputation methods.

Table 9. NB algorithm – F1-score metric

Imputation Method	Mean	Median	Standard Deviation	Minimum	Maximum
RF	0.689	0.689	0.114	0.457	0.984
Average	0.674	0.687	0.145	0.0377	0.984
GBT	0.697	0.686	0.109	0.453	0.958
MICE	0.680	0.675	0.153	0.0465	1.000
KNN	0.700	0.680	0.124	0.424	0.984

In contrast, the NB_Average method yielded the lowest minimum F1-score (0.0377), indicating potential instability or poor performance in certain iterations. The MICE imputation method reached a perfect maximum F1-score of 1.000, demonstrating its potential for optimal classification under certain conditions. However, this method also exhibited the highest standard deviation (0.153), indicating substantial variability and less consistent results. GBT imputation again demonstrated a favourable balance, achieving high performance with a low standard deviation (0.109), reinforcing its consistency across multiple iterations.

Based on these findings, it can be concluded that the NB algorithm generally exhibits stable performance when combined with different imputation methods. Among the tested methods, KNN and GBT imputations produced the most favourable outcomes in terms of average F1-score, offering a desirable combination of predictive accuracy and reliability. In contrast, Average and MICE imputations introduced greater variability, which may pose a risk of reduced performance consistency.

Table 10. SVM algorithm – F1-score metric

Imputation Method	Mean	Median	Standard Deviation	Minimum	Maximum
RF	0.754	0.678	0.154	0.520	1.000
Average	0.753	0.682	0.157	0.529	1.000
GBT	0.749	0.667	0.149	0.536	1.000
MICE	0.740	0.669	0.162	0.509	1.000
KNN	0.748	0.681	0.162	0.420	1.000

Table 10 shows that the highest average F1-score for the SVM algorithm was observed with RF imputation (0.754), indicating slightly better overall performance compared to other imputation methods. Notably, all methods reached a maximum F1-score of 1.000, suggesting that perfect classification was achieved in at least some iterations across all imputations. However, KNN imputation produced the lowest minimum value (0.420), pointing to potential sensitivity of this configuration to input data variability or specific data patterns. Among all evaluated methods, GBT imputation exhibited the lowest standard deviation (0.149), signifying the most stable performance across repeated testing. Overall, the choice of imputation method should align with the primary analytical objective, whether it is maximizing classification accuracy or ensuring robust and stable results across diverse data conditions.

A comparison of the four classification algorithms shows that Random Forest consistently produced the best results, with average F1-scores exceeding 0.82 across all imputation methods, and a peak value of

0.839 using KNN imputation. SVM ranked second, with mean F1-scores around 0.75 and good consistency, particularly when paired with GBT imputation, which provided the most stable results for this algorithm.

Naïve Bayes demonstrated moderate but more variable performance with KNN and GBT imputations, achieving relatively high mean scores around 0.70. The lowest performance was observed with the KNN algorithm, with average F1-scores generally below 0.46. Although GBT imputation slightly improved its results in the case of the KNN algorithm, KNN remained the least accurate and most sensitive to data variations.

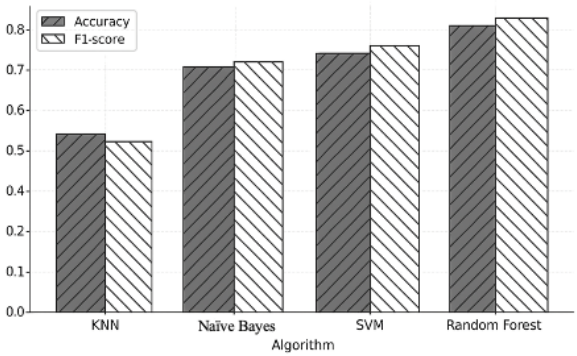


Figure 1. Highest mean values for accuracy and F1-score for each algoritm

Finally, Fig. 1 shows that Random Forest (RF) demonstrated the highest accuracy (up to 0.810) and F1-score (up to 0.839), making it the most robust and reliable classifier among the imputation methods. Support Vector Machine (SVM) showed slightly lower performance, but high stability, especially with GBT imputation. Naïve Bayes (NB) achieved moderate results, and was more sensitive to imputation variability, while K-Nearest Neighbors (KNN) yielded the weakest predictive performance among all evaluated classifiers. RF is recommended as the primary algorithm, with SVM as a strong alternative, while KNN should be avoided due to low accuracy and instability.

5 Conclusion and future research

The comparative analysis of imputation methods and machine learning algorithms clearly demonstrates the significant impact that handling missing values has on classification results in educational data mining. Among the evaluated models, Random Forest consistently achieved the highest accuracy and F1-score, demonstrating both robustness and reliability across all imputation methods. Support Vector Machine followed closely, offering stable and competitive results, particularly with ensemble imputations like GBT. Naïve Bayes, although less accurate, remained a computationally efficient alternative, especially when paired with GBT or KNN imputations. Conversely, K-Nearest Neighbors exhibited the weakest performance and the highest sensitivity to data variations, limiting its applicability in this context. These findings reinforce the importance of carefully selecting both the imputation method and classification algorithm in predictive analytics, especially in domains with limited and incomplete educational data.

A limitation of this study is that the evaluation was conducted on a single institutional dataset with approximately 10% missing values. Future research should replicate the experiments on more varied datasets from different universities and educational contexts to enhance generalizability. In addition, testing the robustness of imputation methods under higher proportions of missingness (e.g. 20-30%) would provide deeper insights into their stability. Finally, while this paper focused on statistical, machine learning-based and hybrid imputations, future work should extend the analysis to state-of-the-art approaches such as deep learning-based and GAN-based imputations, which have shown promising results in other domains.

Acknowledgments

This research was supported by the Faculty of Humanities and Social Sciences, University of Zagreb, Croatia [Institutional research project, 2025].

References

Campanilla, B. S. (2024). Forecasting Degree Completion: A Naïve Bayes Predictive Model for Students' Success. *2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 1–4. https://doi.org/10.1109/ICAECT60202.2024.10469398

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, *8*(1), 140. https://doi.org/10.1186/s40537-021-00516-9

García, S., Luengo, J., & Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. *Knowledge-Based Systems*, *98*, 1–29. https://doi.org/10.1016/j.knosys.2015.12.006

Kovač, A., Dunđer, I., & Seljan, S. (2022). An overview of machine learning algorithms for detecting phishing attacks on electronic messaging services. *2022 45th Jubilee International*

Convention on Information, Communication and Electronic Technology (MIPRO), 954–961. https://doi.org/10.23919/mipro55190.2022.9803517

Maale, F. D., Okyere, G. A., & Awe, O. O. (2025). Effects of Imputation Techniques on Predictive Performance of Supervised Machine Learning Algorithms. In O. O. Awe & E. A. Vance (Eds.), *Practical Statistical Learning and Data Science Methods* (pp. 29–48). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72215-8_2

Martinez, A. L. J., Sood, K., & Mahto, R. (2024). *Early Detection of At-Risk Students Using Machine Learning* (arXiv:2412.09483). arXiv. https://doi.org/10.48550/arXiv.2412.09483

Nakhipova, V., Kerimbekov, Y., Umarova, Z., Suleimenova, L., Botayeva, S., Ibashova, A., & Zhumatayev, N. (2024). Use of the Naive Bayes Classifier Algorithm in Machine Learning for Student Performance Prediction. *International Journal of Information and Education Technology*, *14*(1), 92–98. https://doi.org/10.18178/ijiet.2024.14.1.2028

Pelima, L. R., Sukmana, Y., & Rosmansyah, Y. (2024). Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review. *IEEE Access*, *12*, 23451–23465. https://doi.org/10.1109/ACCESS.2024.3361479

Radišić, B., Seljan, S., & Dunđer, I. (2023). *Impact of missing values on the performance of machine learning algorithms*. 54–62. https://urn.nsk.hr/urn:nbn:hr:277:483201

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, *33*(1), 135–146. https://doi.org/10.1016/j.eswa.2006.04.005

Setiawan, I., Gernowo, R., & Warsito, B. (2023). A Systematic Literature Review On Missing Values: Research Trends, Datasets, Methods and Frameworks. *E3S Web of Conferences*, *448*, 02020. https://doi.org/10.1051/e3sconf/202344802020

Shahbazian, R., & Greco, S. (2023). Generative Adversarial Networks Assist Missing Data Imputation: A Comprehensive Survey and Evaluation. *IEEE Access*, *11*, 88908–88928. https://doi.org/10.1109/ACCESS.2023.3306721

Tanveer, M., Rajani, T., Rastogi, R., Shao, Y. H., & Ganaie, M. A. (2024). Comprehensive review on twin support vector machines. *Annals of Operations Research*, *339*(3), 1223–1268. https://doi.org/10.1007/s10479-022-04575-w

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11. https://doi.org/10.1186/s40561-022-00192-z

Performance study of a bidirectional vanilla spiking neural network for signal noise suppression

Dominik Čambál, Martin Pukanec, Monika Herchlová, Juraj Ďuďák
Slovak University of Technology
Faculty of Materials Science and Technology in Trnava
Jána Bottu 25, Trnava, Slovakia
{dominik.cambal, martin.pukanec, monika.herchlova, juraj.dudak}@stuba.sk

Matúš Nečas
University of Žilina
Research Centre
Univerzitná 8215/1, Žilina, Slovakia
{matus.necas}@feit.uniza.sk

Abstract. *The paper focuses on the design and performance evaluation of a basic "vanilla" spiking neural network (SNN) with a simulated bidirectional architecture, intended for signal noise suppression. The testing was conducted on a synthetic dataset composed of periodic signals with various types of additive noise. The proposed solution was implemented using Leaky Integrate-and-Fire (LIF) neurons within the SNNTorch framework. The procedure consisted of generating noise scenarios, implementing comparative classical filters and neural architectures: a recurrent neural network (RNN), a convolutional neural network (CNN) and a spiking neural network (SNN), designing the bidirectional model, and training it on a defined sample of signals. The paper also included the definition of evaluation metrics and timing parameters for comparing the performance of the individual approaches. The paper describes the entire development process, from architectural design to functional testing at the software simulation level.*

Keywords. Spiking neural network, Recurrent neural network, Convolutional neural network, Denoising

1 Introduction

Noise reduction in signals represents a fundamental challenge in the processing of time-dependent data. In engineering practice, various methods are employed to address this issue. Noise suppression is commonly applied in the medical and biomedical fields. For example, it is used to mitigate distortions caused by patient movement or possible electromagnetic interference from power lines. It also plays a critical role in processing weak brain signals or suppressing other physiological signals. In telecommunications, noise removal is essential to ensure reliable and high-quality communication. In radar systems, noise suppression is used for accurate object detection and improved target resolution. Within industrial automation, denoising sensor signals is necessary to maintain correct process control and regulation, as well as to detect faults and anomalies. Noise reduction is also applied in audio technology to enhance the quality of recordings and sound transmission. In the financial sector, denoising techniques are primarily used to filter market signals in order to uncover underlying trends and predict price movements of various assets.

Conventional noise suppression methods are traditionally classified (Martínez et al., 2022) based on their functional principles. The most widely used, thanks to their computational simplicity and sufficient accuracy, are the *linear filters*, which satisfy the superposition principle. Martínez et al., 2022 use the term *traditional filtering methods*, for a subset of linear filters, albeit the largest one. These methods are based on the assumption that noise occupies different, usually higher frequencies than the original (clean) signal; therefore, frequencies above a chosen cutoff need to be blocked by a low-pass filter (LPF). Besides these traditional methods, the superposition principle is also satisfied by other filters such as the Savitzky-Golay filter. *Adaptive filters* require a reference signal with which is the noisy signal compared. Such a reference signal is often difficult to obtain; in practice it is either provided by the manufacturer, or derived from the same noisy signal, temporarily shifted. The filtering is based on minimization of the error between the reference signal and the filtered signal using a specified statistical metric.

Artificial neural networks (ANNs) approximate the clean signal by being trained to minimize error either between noisy and clean signals (supervised learning),

or between noisy signals alone (unsupervised learning).

Other methods not covered by these categories include signal transforms (wavelet transform), statistical methods (Kalman filter), nonlinear and regularization-based methods (e.g., total variation (TV) denoising, non-local means), and non-neural-model-based approaches (e.g., autoregressive models). The four methods (Butterworth LPF, Savitzky-Golay filter, LMS, and RLS filters) that were chosen for the comparison are covered in more detail in the chapter 2.2.

Unlike the aforementioned methods, artificial neural networks adapt to model nonlinear relationships between noisy and clean signals without needing prior information about the signal shape or the noise model (Maas et al., 2012). If a clean or reference signal–or signals–are available, ANNs can be trained by minimizing the loss function (typically MSE or L1) between the noisy and clean versions (supervised training). This is usually achieved using ANNs designed for temporal data processing, thanks to their inherent statefulness–examples include RNNs, transformers, and SNNs. On the other hand, ANNs, and especially autoencoders, can be trained for signal denoising in an unsupervised manner.

Recurrent neural networks (RNNs) are a common approach, successfully used for noise suppression in audio processing (Maas et al., 2012), biosignal (ECG, EMG) analysis (Antczak, 2019) and others (e.g., industrial or scientific data) (Chen et al., 2020, Piccolomini et al., 2019). While the traditional unidirectional variant is mostly considered sufficient, it has several drawbacks in comparison with the so-called bi-directional RNN (BiRNN), particularly in the first few time steps of the denoised signal.

A BiRNN processes each input sequence twice, once from start to finish and then in reverse, so that each hidden state is influenced not only by its previous value, but also by its future value.

In this paper, we investigate whether the bi-directionality traditionally used in RNNs can also be applied to spiking neural networks (SNNs). Spiking neurons are also described by their internal state, and since they are biologically inspired, the state is known as the membrane potential. The literature on SNNs for signal noise suppression is limited, with most existing studies focusing on SNNs designed as autoencoders (Ren et al., 2025), hybridized with CNNs (Dorzhigulov and Saxena, 2024) or enhanced with mathematical algorithms such as active tuning (Ciurletti et al., 2021). In our work, we propose a vanilla SNN (using only LIF neurons) that simulates bidirectionality, which to our knowledge has been discussed only by (Xiao et al., 2023). Using a set of synthetic signals, we compare, using statistical metrics and processing times, the use of conventional noise suppression with RNNs (forward and bi-directional), CNN AE and SNNs (forward and bi-directional).

2 Methodology

This section describes the signal dataset, noise suppression methods, and metrics used for the performance analysis.

2.1 Signal dataset

For evaluation, we use a synthetic signal dataset of size $n = 1000$. As the study focuses on the proof-of-concept of a bi-directional SNN for noise suppression and its comparison to other methods, we found a single, controlled dataset sufficient. Although we acknowledge the possible limited generalizability, we believe that our approach minimizes confounding variables and allows for a clear comparison. The base signal of our dataset is periodic, composed of two sine waves (Eq. 1).

$$y(t) = A_1 \sin(2\pi f_1 t + \varphi_1) + A_2 \sin(2\pi f_2 t + \varphi_2) \tag{1}$$

with different amplitudes ($A_1 = 1$, $A_2 = 0.5$), frequencies ($f_1 = 30$ Hz, $f_2 = 70$ Hz), and phases ($\varphi_1 = 0$, $\varphi_2 = \pi/4$). The resulting signal is sampled at a frequency $F_s = 1000$ Hz for a duration of 2 s, yielding 2000 samples. The additive noise we use is Gaussian, although we also tested the performance with impulse and quantization noise, which were omitted for brevity.

Gaussian noise is based on the Gaussian (normal) distribution, whose probability density function is given by Eq. 2, where μ is the mean of the distribution and σ is the standard deviation.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2}$$

Gaussian noise is among the most widely used noise types for noise simulations, as it closely approximates real-world noise and has well-understood mathematical properties that make it controllable and predictable. In our project, it is generated using the `numpy` library method, `np.random.normal(mean, std_dev, size)`, with mean $\mu = 0$ and standard deviation $\sigma = 0.5$. The value of the standard deviation was chosen arbitrarily to introduce a significant level of noise to the synthetic signal to create a challenging, yet consistent, denoising problem for all the methods being compared.

2.2 Conventional noise suppression methods

In this section, four conventional noise suppression methods (Butterworth LPF, Savitzky-Golay filter, LMS and RLS filters) are described. These four methods were picked as they represent different paradigms of conventional filtering approaches and are commonly used in practice, as they provide a fair balance between computational complexity and accuracy.

2.2.1 Butterworth LPF

The Butterworth low-pass filter (LPF) was designed to have the frequency response $H(j\omega)$ in the passband as flat as possible (with minimal ripple), which distinguishes it from, for example, Chebyshev filters, which ripple either in the passband or in the stopband (Zhang and Jiang, 2021). The squared frequency response of the filter is given by the Eq. 3.

$$|H(j\omega)|^2 = \frac{1}{1+\left(\frac{\omega}{\omega_C}\right)^{2N}} \tag{3}$$

Based on this formula, two parameters for the design of the filter are crucial. The first one is the order of the filter N: the higher the order, the steeper the frequency response in the transition band. The second one is the cutoff angular frequency ω_C. In our project, after some experimentation, we chose the order $N = 10$ as a good trade-off between computational complexity and accuracy. The cutoff frequency was chosen as $f_c = 100$ Hz, based on the representation of the signal in the frequency domain using Fourier transform. The `butter(order, Wn)` function in Python's `scipy.signal` library calculates internally the cutoff angular frequency using the normalized frequency W_n–the function's second parameter. Its value is usually computed using the formula $W_n = f_c/f_N$, where f_N is the Nyquist frequency ($f_N = F_s/2$, in our case $1000\,\text{Hz}/2 = 500\,\text{Hz}$).

2.2.2 Savitzky-Golay filter

The Savitzky-Golay filter is based on a least-squares polynomial approximation of a window of odd length $N = 2m + 1$ (Liu et al., 2016; Samann and Schanze, 2019). For each window, a polynomial of the order k with coefficients $b = [b_0, \ldots, b_k]$ is fitted using least squares. Accordingly, the `savgol_filter` function in Python's `scipy.signal` library has, in addition to the noisy signal, two other parameters: the window length N and the polynomial order k. In our case, we experimentally found that, for a given polynomial order k, the best window length is approximately given by the Eq. 4.

$$N \approx 7 \cdot \left\lfloor \frac{k}{2} \right\rfloor + 7. \tag{4}$$

If N is even, 1 is added to the result. We use $k = 7$ and $N = 31$.

2.2.3 LMS and RLS filters

LMS (least mean squares) and RLS (recursive least squares) filters both require a reference (or desired) signal for their computations. LMS filter minimizes the difference–the mean squared error–between the noisy and the desired signals using stochastic gradient descent (Hinton, 2003). The filter simplifies the algorithm further by calculating only an estimate of the gradient vector. In general, the k-th sample of a noisy signal x is denoised using a formula given by Eq. 5.

$$\hat{s}_k = y_k - \sum_{i=0}^{N-1} w_k(i) \cdot x_{k-i} \tag{5}$$

In the formula, y is the reference signal, w represents the weights and N is the filter length. In the case of LMS, the weight update rule for the next sample is given by Eq. 6.

$$w_{k+1}(i) = w_k(i) + 2\mu e_k x_{k-i} \tag{6}$$

for $i = 0$ to $N - 1$, where μ is the learning rate, x is the input (noisy) signal vector of size N and e is the error between the desired and actual output.

RLS, in comparison, minimizes the exponentially weighted sum of squared errors; all formulas describing the filter are provided by Tosi et al., 2013.

2.3 Artificial neural networks

In this section, we discuss different artificial neural network (ANN) architectures commonly used for time series analysis.

2.3.1 Recurrent neural networks

Recurrent neural networks (RNNs) were originally developed to address the limitations of feedforward neural networks (FNNs). FNNs were found to be exceptional at static data processing, e.g. measuring datasets or images, but they perform poorly on time-dependent data (time series, signals, audio, video), which must be input one step at a time rather than all at once.

Units in RNNs are enhanced with inner hidden states, neuronal outputs are wired recurrently, i.e. in the next time step, they return to the same neuron or other neurons in the same layer. Hidden states subsequently act act as a kind of memory for each unit, updated based on the previous state and the input. Thanks to this, RNNs are able to capture temporal information and patterns within sequences.

There are three main architectures of RNNs: the vanilla RNN and the RNNs with GRU (gated recurrent unit) or LSTM (long short term memory) units. Vanilla RNN is the default architecture used by the Pytorch `nn.RNN` module, with hyperbolic tangent as the default activation function. In our project, we found the vanilla RNN to provide a good balance of computational complexity, accuracy, and training time. Both versions we used are many-to-many, since both the output and the input of the same size is expected (a denoised signal).

At first, we designed a unidirectional (forward) RNN, where the sequence is processed from start to finish and the hidden state of each neuron is updated based on its direct previous value. On the other hand, in a bi-directional RNN, each time sequence is processed twice, at first from start to finish and then from finish

to start. The single-hidden-layer architecture of a bi-directional RNN is schematically depicted in Fig. 1.

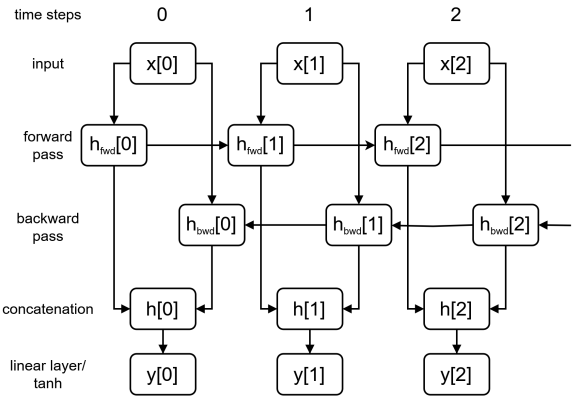


Figure 1. Bi-directional RNN architecture

The forward pass of a bi-directional RNN is described by the Eq. 7, while the backward pass is described by the Eq. 8.

$$h_{fwd}[t] = \tanh(\overrightarrow{W}x[t] + \overrightarrow{U}h_{fwd}[t-1]) \quad (7)$$

$$h_{bwd}[t] = \tanh(\overleftarrow{W}x[t] + \overleftarrow{U}h_{bwd}[t+1]) \quad (8)$$

Denoting n as the input sequence length and d as the hidden size, in eqs. 7 and 8, t is time ($t = 0..n-1$ in the forward pass, $t = n-1..0$ in the backward pass), $x[t]$ is the scalar input, $W \in \mathbb{R}^{d\times 1}$ are the input-to-hidden weight matrices, $U \in \mathbb{R}^{d\times d}$ are the hidden-to-hidden weight matrices and $h[t] \in \mathbb{R}^{d\times 1}$ the hidden state vectors. The weights are shared across time steps.

Both passes yield forward and backward hidden state vectors for each layer for each time step; the hidden state vectors of both passes of the last hidden layer are concatenated as $h[t] \in \mathbb{R}^{2d\times 1}$ (Eq. 9) and serve as inputs to the linear layer (Eq. 10, where $W \in \mathbb{R}^{1\times 2d}$ is the linear layer weight matrix).

$$h[t] = \begin{bmatrix} h_{fwd}[t] \\ h_{bwd}[t] \end{bmatrix} \quad (9)$$

$$y[t] = 2 \cdot \tanh(Wh[t]) \quad (10)$$

The input layer and the output layer each have size 1, and the activation function of the output layer is twice the hyperbolic tangent (eq. 10) as it better captures information on the interval $[-2, 2]$. The interval is slightly larger than the interval of possible amplitudes of our base signal.

2.3.2 Convolutional neural networks

Convolutional neural networks (CNNs), feedforward in nature and therefore widely used for static data processing, can also be applied to time sequences. They learn features in data using a filter, and while in static data (especially 2D data such as images) the filter "moves" across space, in time series (1D data) the filter "moves" across time. Convolutional filters are also translation-invariant, which means that if a temporal pattern is learned in one part of the signal, it can be recognized elsewhere as well. On the other hand, they are usually well-suited for learning short-term features only, and for long-term dependencies are often combined with other neural network architectures (RNN, SNN) or other signal processing techniques (wavelet transforms).

One CNN architecture for denoising is an autoencoder (AE), composed of two parts: an encoder and a decoder. The data in the encoder is compressed, until it reaches a specific layer called the bottleneck (or latent space). From there on, the compressed data is reconstructed. The layers in the decoder symmetrically mirror the layers of the encoder. The number of channels in the encoder increases in each layer to capture finer details, while in the decoder it decreases. Conversely, the data size decreases (the number of time steps in time series) in the encoder by a factor of two and symmetrically increases by the same factor in the decoder. This architecture is self-supervised, meaning that it doesn't need to have the original underlying base signal, but it learns patterns that are consistent throughout the entire dataset.

The CNN AE in our project (architecture shown in Fig. 2) consists of two layers in the encoder and the decoder, with neurons in each layer having ReLU as their activation function. The number of time steps decreases before increasing by a factor of two.

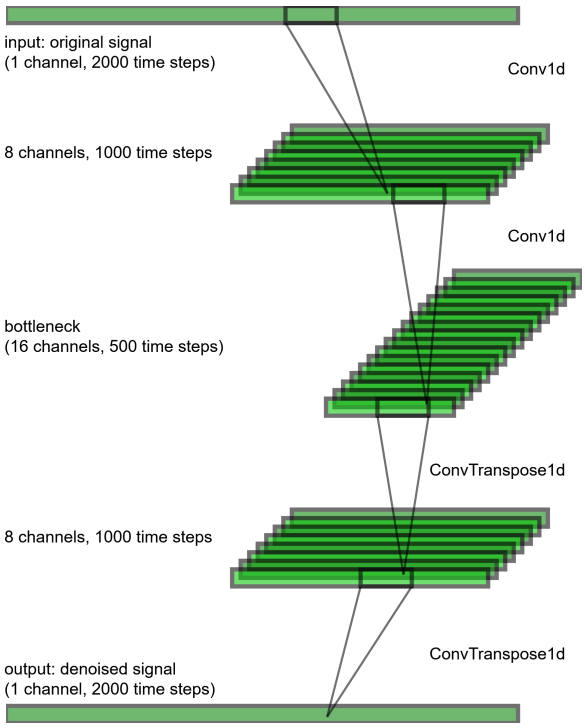


Figure 2. CNN AE architecture

2.3.3 Spiking neural networks

Spiking neural networks (SNNs) (Eshraghian et al., 2023) are biologically inspired, with neuron activation functions that are essentially unit steps: at each time step, a neuron's output is either 0 or 1. This output depends on the neuron's internal state, called the membrane potential, which produces an output called a spike (a 1 value) when a certain threshold θ is reached and then resets (either to zero or decreases by a set value).

SNNs are typically designed using so-called Leaky Integrate-and-Fire (LIF) neurons, which offer a trade-off between computational complexity and biological implausibility. They are described by the formula given by Eq. 11.

$$U[t] = \underbrace{\beta U[t-1]}_{decay} + \underbrace{WX[t]}_{input} - \underbrace{S_{out}[t-1]\theta}_{reset} \quad (11)$$

In the formula, t is time, U is the membrane potential, W is a weight matrix, β is the decay rate, θ is the threshold and S_{out} is the output spike (the activation function) (Eq. 12),

$$S_{out}[t] = \begin{cases} 1 & U[t] \geq \theta \\ 0 & otherwise \end{cases} \quad (12)$$

Spikes then travel across synapses to other neurons, where they accumulate and increase the membrane potential. Unlike units in RNNs, without any external stimuli (spikes), the internal state U of LIF neurons decreases (decays) by each time step. Any information passed through the SNN is therefore not encoded in the value of the output (since it is binary), but typically in the timing of the spikes (latency code) or the spike rate (rate coding).

One of the main problems when training SNNs is the so-called dead neuron problem: since backpropagation relies on stochastic gradient descent, by the chain rule, the Eq. 13 follows.

$$\frac{dL}{dW} = \frac{dL}{dS}\frac{dS}{dU}\frac{dU}{dI}\frac{dI}{dW} \quad (13)$$

In Eq. 13, L is the loss function, W is the weight matrix, S is the activation function and I represents the inputs. Since S is essentially a unit step, its weak derivative $\frac{dS}{dU}$ is the Dirac delta function, which takes on only two values: ∞ when the threshold is reached and 0 otherwise. Consequently, $\frac{dL}{dW}$ will have the same value, which prevents the weights from being updated. In practice, one of the most common ways to resolve this is by using a surrogate function during the backward pass, usually a smoothed version of the unit step (e.g., sigmoid or arctangent).

Nonetheless, the provided architecture results in various advantages of SNNs: they're asynchronous and highly parallelizable, which stems from the fact that most of the time most of the neurons are at rest. With the exception of membrane potentials, most of the information that travels in SNNs is not only non-floating, but also binary; and can be stored in vectors or simply as lists of spike times.

The driving principle behind SNN research is energy reduction: ANNs are known for their enormous energy consumption. In general, SNNs were shown to reduce energy use; the review of various architectures and their effect on energy consumption is provided by (Malcolm and Casco-Rodriguez, 2023). For example, the SNN Brain-Machine Interface, developed by (Liao et al., 2022), achieved reductions in comparison to aconventional ANN in total operations and memory accesses by well over 90%. To achieve this, a special neuromorphic hardware, which can support asynchronicity and parallelization, is needed. Various chips (Intel Loihi) and institutional supercomputers (University of Manchester's SpiNNaker) were built, although there were successful implementations of SNNs on FPGAs as well (Padovano et al., 2024).

That said, it is possible to simulate SNNs on traditional computer architectures using various libraries. In our project, we use SNNTorch, based on PyTorch, that we use in our RNN and CNN modules.

As with RNNs, we used both a forward model, but also a model that simulates bi-directionality. The overall architecture of the bi-directional variant is schematically depicted in the Figure 3.

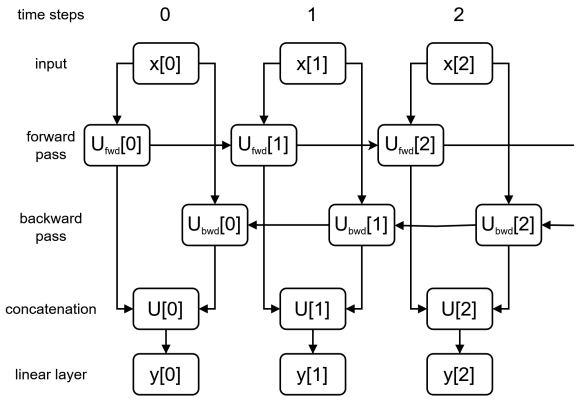


Figure 3. Bi-directional SNN architecture

The architecture of the bi-directional SNN consists of two branches (one for the forward pass and one for the backward pass), with single-neuron input and output layers and multiple hidden layers. Mathematically, and for demonstration purposes with only one hidden layer, in the forward pass for each time step t, the formula given by Eq. 14 is used for each neuron, and for the backward pass, the formula given by Eq. 15 is used.

$$\overrightarrow{U}[t] = \beta\overrightarrow{U}[t-1] + \overrightarrow{W}X[t] - \overrightarrow{S}_{out}[t-1]\theta \quad (14)$$

$$\overleftarrow{U}[t] = \beta\overleftarrow{U}[t+1] + \overleftarrow{W}X[t] - \overleftarrow{S}_{out}[t+1]\theta \quad (15)$$

Eqs. 14 and 15 are based on the Eq. 11. The spike output formula (Eq. 12) remains unchanged. Denoting d as the hidden size, for neurons in the hidden

layer, $X[t]$ is a scalar input, $\overrightarrow{W}[t], \overleftarrow{W}[t] \in \mathbb{R}^{d\times 1}$ are the weight matrices, $\overrightarrow{U}[t], \overleftarrow{U}[t] \in \mathbb{R}^{d\times 1}$ are the membrane potential matrices, and $\overrightarrow{S}_{out}[t], \overleftarrow{S}_{out}[t] \in \{0,1\}^{d\times 1}$ are the spike matrices. For the single output neuron, the weight matrices are of size $1 \times d$ and all other quantities are scalar. As with the RNN (chapter 2.3.1), the resulting membrane potentials are then concatenated and passed through a linear layer.

The main difference between the SNN and the RNN architecture (Figure 1) is in the way the information is propagated between the neurons in the hidden layers.

In our project, bi-directionality was implemented using a class `SNNBranch` with two hidden layers of LIF neurons (100 and 75 neurons, respectively).

We found bi-directional SNN to be a promising architecture for signal denoising, based on initial experiments with shorter, simpler sine wave signals with an amplitude of $A = 1$ and a frequency of $f = 2\pi/60$ Hz, sampled at 500 time steps. Both the forward and the bi-directional SNN were trained on 80% of the 5000 samples. While the forward SNN yielded a non-optimal denoised output, the bi-directional SNN significantly enhanced the denoising quality (Fig. 4). The average MSE between the noisy and BiSNN-denoised signals was 0.2506.

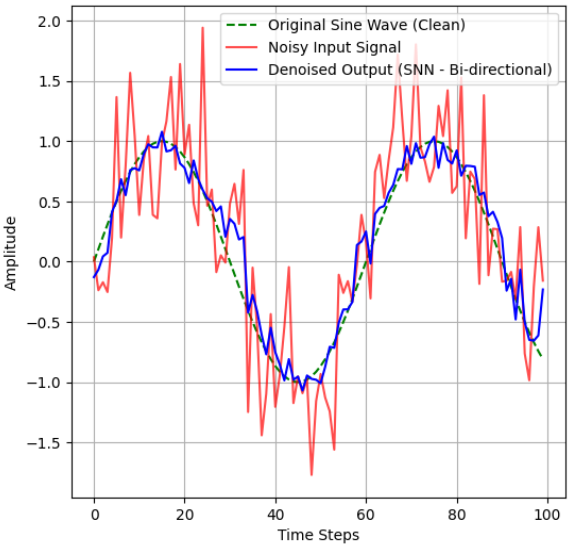


Figure 4. Simple single-sine signal denoising with BiSNN

2.4 Metrics

In our project, we use the mean squared error (MSE), which is based on the squared differences between the clean signal and the denoised signal; the lower the MSE, the better the denoising performance. The MSE values shown in the table are averages computed across all noisy-denoised signal pairs in the testing dataset. Other metrics, such as the root mean squared error (RMSE) and the signal-to-noise ratio (SNR), were considered but were omitted from presented results mostly due to brevity, as their values were roughly concordant with MSE.

We also measured computational time. For the conventional denoising methods, we recorded the time taken to process the entire dataset. For the ANN-based methods, we recorded the time it took for the neural network to be trained. These times were chosen because they represent the most time-consuming operations in the overall computation.

3 Results

The methods discussed in Chapters 2.2 and 2.3 and the metrics described in Chapter 2.4 were implemented in the Python programming language. Existing libraries such as Scikit-learn (for MSE/RMSE), SciPy (for Butterworth LPF and Savitzky-Golay filter), PyTorch (for neural networks), and SNNTorch (for SNNs) were utilized. Custom implementations were developed for LMS and RLS filters, as well as SNR/PSNR metrics. Neural networks were trained on Google Colab using v2-8 TPU processing units.

The results obtained from measurements are provided in Tables 1 and 2.

Table 1. Performance of conventional denoising methods

Denoising Method	MSE	Denoising time [s]
Butterworth LPF	0.0481	0.15
Savitzky-Golay filter	0.0439	0.48
LMS filter	0.0309	12.95
RLS filter	0.0427	138.28

Table 2. Performance of ANNs

Denoising Method	MSE	Training time [s]
Forward RNN	0.0220	146.25
Bi-directional RNN	0.0117	453.41
CNN autoencoder	0.0254	3.07
Forward SNN	0.1672	223.15
Bi-directional SNN (175 neurons)	0.1030	479.72
Bi-directional SNN (650 neurons)	0.0777	3589.26

The conventional methods performed similarly in denoising the datasets, with the LMS filter achieving the lowest MSE, and the RLS, Savitzky-Golay, and Butterworth LP filters performing marginally worse. The main disadvantage of both adaptive filters was the time required to denoise the entire dataset, with

the RLS filter in particular taking more than 2 minutes to denoise the entire dataset. The good performance of the Savitzky-Golay filter can be attributed to the polynomial nature of the sine wave, which, however, may hinder in denoising of non-stationary signals. Both adaptive filters also initialize their weights at zero; therefore, they take some time to learn, which is why the amplitude for the first couple of hundredths of a second is at 0.

Similar problems can be observed in the forward architectures of the RNN and the SNN. The inaccuracies in the first time steps can be attributed to the initialization of weights; in other words, the neural network does not yet have any knowledge about the data. Bi-directionality addresses this problem by having an access to the future context in the early time steps. The bi-directional RNN outperformed all of the conventional methods used, as did the marginally less accurate CNN autoencoder, which also has the advantage of no need for the clean version of the signal. Particularly significant was the superior performance of RNNs over other methods that require a reference signal.

The improved performance of bi-directional SNN compared to the forward SNN described in the chapter 2.3.3 was also observed when training the SNN on the two-sine dataset (Table 2), where the bi-directional SNN outperformed its forward counterpart by almost 40%. Nevertheless, both architectures were outperformed by other methods in terms of accuracy. However, the training time of the bi-directional SNN was comparable with that of bi-directional RNN despite the significant increase in the number of neurons (the RNN having 30 units in the hidden layer, while the SNN having 175 neurons in both hidden layers combined).

Further refinement of the bi-directional SNN included the addition of another hidden layer and an increase in neuron count per layer (250, 200, and 200 neurons, respectively). This upgraded architecture resulted in an MSE of 0.0777, representing nearly a 25% improvement (a sample is shown in the Figure 5).

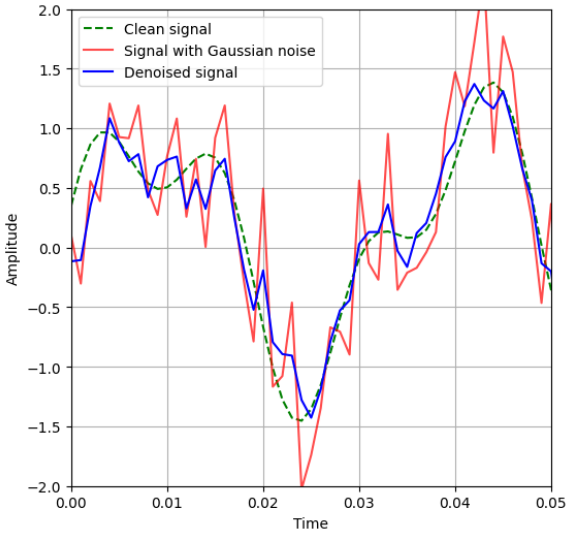


Figure 5. Signal sample with upgraded Bi-SNN

4 Discussion

The initial experiments with SNNs on simpler signals provided valuable insights into their behavior and the benefits of bi-directionality. The quantitative improvement observed with the bi-directional SNN on the actual dataset, despite its lower accuracy compared to other methods, suggests that SNNs hold promise, particularly considering their comparable training times to bi-directional RNNs even with a significantly higher number of neurons. This efficiency, coupled with the potential for more biologically plausible computation and potential energy consumption reduction given a proper architecture, warrants further investigation.

Although the architectural upgrade to the bi-directional SNN yielded a substantial improvement in MSE, its overall accuracy still lags behind other methods. Future research avenues could explore more complex SNN architectures, such as incorporating learnable decay parameters as demonstrated by Hao et al., 2024, or investigating hybrid models combining SNNs with RNNs, CNNs, or autoencoders to leverage the strengths of different neural network paradigms for enhanced denoising performance.

5 Conclusion

The paper presents the design and implementation of a “vanilla” spiking neural network featuring two hidden layers composed of Leaky Integrate-and-Fire neurons, utilizing a simulated bidirectional architecture in the SNNTorch framework. The proposed solution was evaluated on a synthetic dataset of periodic signals corrupted by noise and compared against traditional and neural-based denoising approaches. The development process included the design of the dataset, implementation of baseline filters and networks, construction of the bidirectional SNN processing mechanism, and the selection of metrics for accuracy and computational cost assessment. The results demonstrate that the proposed bidirectional SNN architecture offers clear improvements over its unidirectional counterpart. Although it lags behind BiRNNs and CNN-based autoencoders in certain metrics, its key advantage generally lies in significantly lower energy consumption given a specific hardware, although the training time advantage is also noteworthy. Future research may explore hybrid architectures, e.g., SNN-RNN combinations, or the application of learned parameters for membrane potential decomposition.

Acknowledgments

This work was supported by KEGA through the Teaching and Development of methodology for the use of microcontrollers in automation using practical examples and laboratory exercises for engineering students, under Grant 024STU-4/2024.

References

Antczak, K. (2019). Deep recurrent neural networks for ecg signal denoising. https://arxiv.org/abs/1807.11551

Chen, H., Guo, R., Liu, J., Wang, Y., & Lin, R. (2020). Magnetotelluric data denoising with recurrent neural network. In *Seg 2019 workshop: Mathematical geophysics: Traditional vs learning, beijing, china, 5–7 november 2019* (pp. 116–118). https://doi.org/10.1190/iwmg2019_28.1

Ciurletti, M., Traub, M., Karlbauer, M., Butz, M. V., & Otte, S. (2021). Signal Denoising with Recurrent Spiking Neural Networks and Active Tuning. In I. Farkaš, P. Masulli, S. Otte, & S. Wermter (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2021* (pp. 220–232). Springer International Publishing.

Dorzhigulov, A., & Saxena, V. (2024). Compact Convolutional SNN Architecture for the Neuromorphic Speech Denoising. *2024 IEEE 67th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 1191–1195. https://doi.org/10.1109/MWSCAS60917.2024.10658937

Eshraghian, J. K., Ward, M., Neftci, E. O., Wang, X., Lenz, G., Dwivedi, G., Bennamoun, M., Jeong, D. S., & Lu, W. D. (2023). Training Spiking Neural Networks Using Lessons From Deep Learning. *Proceedings of the IEEE*, *111*(9), 1016–1054. https://doi.org/10.1109/JPROC.2023.3308088

Hao, X., Ma, C., Yang, Q., Tan, K. C., & Wu, J. (2024). When Audio Denoising Meets Spiking Neural Network. *2024 IEEE Conference on Artificial Intelligence (CAI)*, 1524–1527. https://doi.org/10.1109/CAI59869.2024.00275

Hinton, O. (2003). *EEE305 - Digital Signal Processing*. University of Newcastle upon Tyne. https://www.staff.ncl.ac.uk/oliver.hinton/eee305/ (Accessed: 26.06.2025).

Liao, J., Widmer, L., Wang, X., Di Mauro, A., Nason-Tomaszewski, S. R., Chestek, C. A., Benini, L., & Jang, T. (2022). An Energy-Efficient Spiking Neural Network for Finger Velocity Decoding for Implantable Brain-Machine Interface. *2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 134–137. https://doi.org/10.1109/AICAS54282.2022.9869846

Liu, Y., Dang, B., Li, Y., Lin, H., & Ma, H. (2016). Applications of Savitzky-Golay filter for seismic random noise reduction. *Acta Geophysica*, *64*, 101–124. https://doi.org/https://doi.org/10.1515/acgeo-2015-0062

Maas, A. L., Le, Q. V., O'Neil, T. M., Vinyals, O., Nguyen, P., & Ng, A. Y. (2012). Recurrent Neural Networks for Noise Reduction in Robust ASR. *Interspeech 2012*, 22–25. https://doi.org/10.21437/Interspeech.2012-6

Malcolm, K., & Casco-Rodriguez, J. (2023). A comprehensive review of spiking neural networks: Interpretation, optimization, efficiency, and best practices. https://arxiv.org/abs/2303.10780

Martínez, M., Garcia-March, M. A., Enrique, C., & Fernández de Córdoba, P. (2022). *Algorithms for Noise Reduction in Signals: Theory and practical examples based on statistical and convolutional analysis*. Bristol: Institute of Physics Publishing. https://doi.org/10.1088/978-0-7503-3591-1

Padovano, D., Carpegna, A., Savino, A., & Di Carlo, S. (2024). SpikeExplorer: Hardware-Oriented Design Space Exploration for Spiking Neural Networks on FPGA. *Electronics*, *13*(9). https://doi.org/10.3390/electronics13091744

Piccolomini, E. L., Gandolfi, S., Poluzzi, L., Tavasci, L., Cascarano, P., & Pascucci, A. (2019). Recurrent Neural Networks Applied to GNSS Time Series for Denoising and Prediction. In J. Gamper, S. Pinchinat, & G. Sciavicco (Eds.), *26th international symposium on temporal representation and reasoning (time 2019)* (10:1–10:12, Vol. 147). Schloss Dagstuhl – Leibniz-Zentrum für Informatik. https://doi.org/10.4230/LIPIcs.TIME.2019.10

Ren, P., Wang, Y., Wang, Z., Peng, D., Liu, C., & Han, T. (2025). Denoising autoencoder multilayer perceptron spiking neural network for isonicotinic acid yield prediction on real industrial dataset. *Advanced Engineering Informatics*, *65*, 103273.

Samann, F., & Schanze, T. (2019). An efficient ECG denoising method using Discrete Wavelet with Savitzky-Golay filter. *Current Directions in Biomedical Engineering*, *5*(1), 385–387. https://doi.org/doi:10.1515/cdbme-2019-0097

Tosi, D., Poeggel, S., Leen, G., & Lewis, E. (2013). Adaptive filter-based interrogation of high-sensitivity fiber optic Fabry-Perot interferometry sensors. *Sensors and Actuators A: Physical*, *206*. https://doi.org/10.1016/j.sna.2013.12.010

Xiao, R., Wan, Y., Yang, B., Zhang, H., Tang, H., Wong, D. F., & Chen, B. (2023). Towards Energy-Preserving Natural Language Understanding With Spiking Neural Networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *31*, 439–447. https://doi.org/10.1109/TASLP.2022.3221011

Zhang, X., & Jiang, S. (2021). Application of Fourier Transform and Butterworth Filter in Signal Denoising. *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, 1277–1281. https://doi.org/10.1109/ICSP51882.2021.9408933

Compressibility of WGANs in a Distributed Learning Environment

Luka Lukač, Damjan Strnad, Štefan Horvat
University of Maribor
Faculty of Electrical Engineering and Computer Science
Koroška cesta 46, 2000 Maribor, Slovenia
{luka.lukac, damjan.strnad, stefan.horvat}@um.si

Abstract. *This work presents a method for compressing Wasserstein generative adversarial networks (WGANs) in a distributed learning environment to reduce network load and server memory usage. Instead of transferring raw data from edge devices to a central server, a WGAN is trained locally on each edge device. The generator part of a trained WGAN is then pruned according to the lottery ticket hypothesis, and compressed using a general-purpose codec. Experimental results demonstrate that the generator can be compressed by up to 15x and still retain the ability to generate high-quality samples.*

Keywords. machine learning, generative adversarial network, Wasserstein distance, lottery ticket hypothesis, data compression

1 Introduction

In the recent era, a rapid development of low-cost sensor devices and their widespread deployment have led to the collection of vast amounts of data. This presents a significant challenge for centralized machine learning (ML) approaches, as transferring massive datasets from edge devices to the server can be resource intensive (Margara et al., 2023). Furthermore, the data distribution across clients can be non-IID in a real-world environment, therefore, it is essential to capture local data distributions of all clients. To address these issues, a special type of the distributed learning paradigm has been proposed (X. Cao et al., 2023), in which separate clients train their local models and send them to the server where they are aggregated into a global model.

Generative ML models that enable generating synthetic data are often used for reducing the communication between clients and the server within a distributed learning environment (Kasturi and Hota, 2023; Merugu and Ghosh, 2003; Sajjadi Mohammadabadi et al., 2024). The idea behind the approach is that the generative models are trained on edge devices and transferred to the server instead of the raw data. This way, the network load and the server memory consumption can be greatly reduced. Aggregated generative models on the server enable on-the-fly generation of data that resembles the local samples of clients, which can significantly increase the performance of trained models (Wijesinghe et al., 2024).

Modern generative ML models are becoming increasingly more complex (C. Li et al., 2016), which presents a significant burden on the communication channels in a distributed setting. Several model compression techniques for ML models were proposed to reduce the memory consumption and network loads (Z. Li et al., 2023), including model pruning (He et al., 2019; Yang et al., 2017), parameter quantisation (Wu et al., 2016), low-rank decomposition (H. Li et al., 2023), and knowledge distillation (Qin et al., 2021; Tung and Mori, 2019). One of the most promising and efficient model pruning techniques is known as the lottery ticket hypothesis (Frankle and Carbin, 2019). In many cases, this method is able to prune more than 90% of neural network (NN) parameters, and obtain the so called winning tickets (Malach et al., 2020). Pruning by itself does not reduce the model size, as it only sets the redundant parameters to zero (Z. Li et al., 2023). However, this reduces the information entropy of the data, and consequently, enables a more efficient model compression. Therefore, compressing the obtained winning tickets from the locally trained generators using general-purpose data compression algorithms before sending them to the server can significantly reduce network load and improve communication efficiency.

The main contributions of this paper are:

- A novel approach for compression of WGAN generators using the lottery ticket pruning and general-purpose compression methods.
- Experimental work, which indicates the efficiency of the proposed compression scheme for WGANs, having a negligible impact on the generator performance.

The remainder of this paper is structured in the following way: Section 2 describes the existing methods for compression of NN models along with the related generative adversarial networks (GANs). Section 3 explains the methodology behind the proposed Wasserstein GAN (WGAN), the lottery ticket extraction, and

the pruned model compression. Section 4 summarises the results and analyses them, while Section 5 concludes the paper and presents future work.

2 Related Work

This section consists of two parts. Firstly, the existing methods for compression of neural network models are reviewed. After that, the basic principles of GANs are presented.

2.1 Compression of Neural Network Models

The most common and earliest methods for the compression of NNs are known as pruning methods. Their main principle is to remove components inside an NN that have negligible impact on its performance (Z. Li et al., 2023). From the early days of weight decay (Hanson and Pratt, 1988), many different pruning methods have been developed. Structured pruning methods typically remove redundant channels in separate layers of the NN (Luo et al., 2019; Xiang et al., 2021) where the importance of each channel is calculated using different heuristics (Chang et al., 2022; Kuang et al., 2022; Q. Li et al., 2022; Lin et al., 2018). On the other hand, unstructured pruning methods set separate, unrelated NN parameters to zeros (Han et al., 2016; Molchanov et al., 2019). Therefore, such pruned NNs are not compressed by itself after the pruning, and their representation needs to be compressed with data compression algorithms (Frankle and Carbin, 2019).

Parameter quantisation lowers the precision of internal NN parameters (Z. Li et al., 2023), and, consequently, reduce the total NN storage space. Some methods perform the quantisation after the training phase (Cai et al., 2020; Fang et al., 2020; Shomron et al., 2021) while others quantise the NN parameters during each forward and backward propagation of the training (Rastegari et al., 2016; Tailor et al., 2021). The advantage of the first group of methods is simplicity and high-speed quantisation, while the post-training quantisation methods yield more accurate results (Z. Li et al., 2023).

Low-rank decomposition approximates the weight matrix of the NN with the product of lower-rank matrices (H. Li et al., 2023). Although some decomposition methods deal with fully connected layers (Denil et al., 2013; Lu et al., 2017; Yu et al., 2017), most of the procedures for low-rank decomposition operate on convolutional layers (Denton et al., 2014; S. Lee et al., 2021; Lin et al., 2019; Rigamonti et al., 2013). Generally, such form of model compression removes redundancy and is applicable to a wide variety of NN architectures. It may, however, be less effective in state-of-the-art networks due to common occurrences of 1x1 convolution filters, which cannot be efficiently decomposed (Z. Li et al., 2023).

Knowledge distillation is a form of model compression where a smaller student network is trained using the results of a large pre-trained teacher network (Gou et al., 2021). Different approaches for training such teacher-student architectures were proposed. The simplest methods use the output of the last layer of the teacher network and train the student network to mimic such outputs (Ba and Caruana, 2014; Hinton et al., 2015). On the other hand, more advanced distillation models also match outputs of intermediate layers and/or their relations to train the student (S. H. Lee et al., 2018; Yim et al., 2017), which can increase the performance of the student model.

2.2 Generative Adversarial Networks

A generative adversarial network (GAN) is a popular generative ML model that was proposed in 2014 (Goodfellow et al., 2014). It consists of two NNs: generator, which is trained to produce samples that resemble the original data distribution, and discriminator that evaluates the realism of samples (either real or artificially generated). The main goal of the alternating training of both models is to train the generator to the point where the discriminator can no longer reliably distinguish between real and fake samples. The GAN structure is depicted in Figure 1. Conditional GAN (cGAN) was introduced shortly after the vanilla GAN, providing the user with control over generated samples' classes (Mirza and Osindero, 2014).

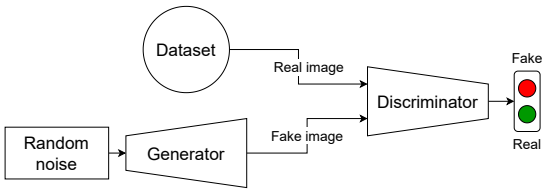


Figure 1. Structure of a GAN.

One of the most common problems that can occur during the training of GANs is mode collapse (Ding et al., 2022). Different ways of tackling this issue were proposed in the past (Bhagyashree et al., 2020; Dai et al., 2024; Durall et al., 2020). A major improvement in training of GANs was the introduction of Wasserstein GAN (WGAN) (Arjovsky et al., 2017), which uses the Wasserstein distance instead of the conventional divergence-based losses. Consequently, the alternating training process of a discriminator (often referred to as a critic) and a generator leads to greater stability and reduces the likelihood of a mode collapse. Due to their great performance and ability to generate a wide variety of data types, WGANs are successfully used in many modern applications (J. Li et al., 2020; Liu et al., 2023; Vo et al., 2024; Wang et al., 2019).

3 Method

This section consists of three parts: firstly, the principles behind the WGAN are described, after that, the lottery ticket pruning is explained, and lastly, the compression techniques of pruned models are presented.

3.1 WGAN

Let G_i be a generator, and let C_i be a critic of a conditional WGAN that belongs to the client $c_i \in \mathcal{C}$. The main goal of G_i is to fool C_i by generating samples that resemble the real data (i.e., critic assigning high scores to fake samples). During the training, G_i is optimised based on the generator loss function $\mathcal{L}_{G_i}$ in Equation 1.

$$\mathcal{L}_{G_i} = -\mathbb{E}_{z\sim\mathcal{N}(0,I)}\left[C_i(G_i(z|y)|y)\right] \tag{1}$$

where z is a latent vector sampled from the standard Gaussian distribution $\mathcal{N}(0, I)$, $\sim$ is the sampling operation, and y is the class label.

The critic C_i estimates the Wasserstein distance between the real and generated data distributions. The main idea is that a well-trained C_i should produce higher scores for real samples and lower values for fake samples. Additionally, the critic loss function $\mathcal{L}_{C_i}$ should impose 1-Lipschitz constraint (bounding the gradient to range $[0, 1]$) (Zhang et al., 2020). This constraint is enforced with a gradient penalty technique (Gulrajani et al., 2017). The critic loss function $\mathcal{L}_{C_i}$ is given in Equation 2.

$$\mathcal{L}_{C_i} = \mathbb{E}_{x\sim\mathbb{P}_r^i}\left[C_i(x|y)\right] - \mathbb{E}_{z\sim\mathcal{N}(0,I)}\left[C_i(G_i(z|y)|y)\right] + \lambda\cdot\mathbb{E}_{\hat{x}}\left[\left(\left\|\nabla_{\hat{x}}C_i(\hat{x}|y)\right\|_2 - 1\right)^2\right] \tag{2}$$

where x represents a real data sample from the i-th client's data distribution $\mathbb{P}_r^i$, λ the strength of the gradient penalty term, and $\hat{x}$ a synthetic sample interpolated between a real and a fake sample.

The training loop of the WGAN with gradient penalty is presented in Algorithm 1. The training is performed for N iterations (Line 8), during which the G_i and C_i are trained in alternating cycles. Usually, C_i is trained for R iterations (Line 10) where $\mathcal{L}_{C_i}$ is calculated in Line 12 using real samples x, latent vectors z, and class labels y. The parameters of C_i are recalculated in Line 13. After that, in Line 16, $\mathcal{L}_{G_i}$ is calculated using the newly sampled latent vectors z. The parameters of G_i are updated in Line 17. After the training is performed for N iterations, the trained G_i and C_i are returned as a result of the function in Line 19.

Algorithm 1 Training loop of the WGAN with gradient penalty on $c_i \in \mathcal{C}$.

1: **function** WGAN-TRAIN(G_i, C_i, N, B, R, η_G, η_C, λ)
2: ▷ N: number of iterations
3: ▷ B: batch size
4: ▷ R: no. of critic iterations per generator iteration
5: ▷ η_G, η_C: generator and critic learning rate
6: ▷ λ: gradient penalty term strength
7:
8: **for** N iterations **do**
9: $x, y \leftarrow$ LoadRealSamples(B)
10: **for** R iterations **do**
11: $z \leftarrow$ SampleLatentVectors(B)
12: $\mathcal{L}_{C_i} \leftarrow$ CriticLoss(x, z, y, λ) ▷ Eq. 2
13: $C_i \leftarrow$ Backpropagation($\mathcal{L}_{C_i}$, η_C)
14: **end for**
15: $z \leftarrow$ SampleLatentVectors(B)
16: $\mathcal{L}_{G_i} \leftarrow$ GeneratorLoss(z, y) ▷ Eq. 1
17: $G_i \leftarrow$ Backpropagation($\mathcal{L}_{G_i}$, η_G)
18: **end for**
19: **return** G_i, C_i
20: **end function**

3.2 Lottery Ticket Pruning

The lottery ticket hypothesis states that a dense, randomly initialised NN contains a subnetwork that can match the accuracy of the original network after training for at most the same number of iterations. To prune models efficiently, non-pruned parameters should be reset to their initial values and retrained after each pruning iteration (Frankle and Carbin, 2019).

Let $f(x, \theta)$ be a NN with initial parameters $\theta \in \mathbb{R}^d$. The network achieves accuracy a after N iterations. There exists a subnetwork $f(x, M \odot \theta)$ with mask M that achieves accuracy a' after at most N' iterations where $\odot$ denotes elementwise multiplication. The lottery ticket hypothesis predicts the existence of a subnetwork:

- $\|M\|_1 \ll d$
  The subnetwork will always have fewer parameters.
- $a' \geq a$
  The subnetwork will have comparable accuracy.
- $N' \leq N$
  The subnetwork will learn in a comparable number of NN training iterations.

The subnetwork $f(x, M \odot \theta)$ is called the winning ticket (Frankle and Carbin, 2019), and is extracted by the iterative pruning procedure shown in Algorithm 2. The default initialisation of the generator and critic parameters, denoted as G_i^0 and C_i^0, is remembered in Line 7. The pruning mask M is initialised to ones in Line 8, and determines which generator parameters have been pruned (i.e., mask value of zero denotes a

pruned parameter). After the initialisation, iterative pruning of G_i is performed for P iterations. The intensity of pruning is determined by the pruning rate δ, which specifies the percentage of remaining parameters that are pruned in each iteration. This means that the total share ρ of pruned parameters after p iterations is given by:

$$\rho = 1 - (1 - \delta)^p \tag{3}$$

In Line 11, the parameters of G_i are reset to the default initialisation according to M, where the previously pruned parameters are set to zero and frozen during subsequent training. The C_i parameters are always reset to the default initialisation and remain unaffected by the pruning mask, since only G_i is being pruned. The models are retrained for a fixed number of training iterations in Line 12. After training, the pruning is performed in Line 13 by sorting the parameters by increasing magnitude, and marking the share ρ of the lowest for pruning by setting the corresponding mask values to zero. After P pruning iterations, the final G_i is declared a winning ticket.

Algorithm 2 Iterative pruning lottery ticket algorithm.

1: **function** ITERATIVELOTTERYTICKET(G_i, C_i, P, δ)
2: ▷ G_i: initialised generator model
3: ▷ C_i: initialised critic model
4: ▷ P: no. of lottery ticket pruning iterations
5: ▷ δ: pruning rate
6:
7: G_i^0, $C_i^0 \leftarrow$ GetInitParams(G_i, C_i)
8: $M \leftarrow$ InitPruningMask()
9: **for** $p \leftarrow 1$ **to** P **do**
10: $\rho \leftarrow 1 - (1 - \delta)^p$ ▷ Eq. 3
11: G_i, $C_i \leftarrow$ ResetToInitParams(G_i^0, C_i^0, M)
12: G_i^e, $C_i^e \leftarrow$ WGAN-Train(G_i, C_i) ▷ Alg. 1
13: G_i, C_i, $M \leftarrow$ Prune(G_i^e, C_i^e, M, ρ)
14: **end for**
15: **return** G_i
16: **end function**

There are two approaches to model pruning. The first approach is local pruning where the proportion ρ of parameters is pruned in each layer separately. This technique is simple and stable, but can prune parameters with larger magnitudes, potentially resulting in reduced model performance. The second approach is global pruning where the parameters are pruned with respect to the global magnitude order. This can lead to higher efficiency and flexibility, but if layers have low numbers of parameters, a layer crash can appear, where all layer parameters are set to zero and the pruned network becomes unusable (Frankle and Carbin, 2019).

3.3 Compression of Pruned Models

The obtained winning ticket model $f(x, M \odot \theta)$ has the same size as the original, non-pruned G_i. However, as it contains many zero parameters (depending on the number of pruning iterations P), the information entropy of $f(x, M \odot \theta)$ is significantly lower. Low information entropy signifies that the pruned $f(x, M \odot \theta)$ is much more compressible than G_i. Consequently, $f(x, M \odot \theta)$ is compressed with an arbitrary general-purpose data compression algorithm, such as zlib (Gailly, Jean-loup and Adler, Mark, 2004), LZMA ("XZ Utils", 2025), bzip2 ("bzip2 and libbzip2", 2025), or paq9a (Mahoney, 2025).

4 Results

The purpose of the experiments was to examine how WGANs can be compressed within a distributed learning environment. The experiments are demonstrated for a single client c_i yet the proposed training and compression pipeline can easily be re-used on multiple clients. In the continuation of this section, the results of the WGAN training, the lottery ticket pruning, and the compression of the pruned model using general-purpose compressors are presented in detail.

The experiments were conducted on our own vector dataset named **H**and**w**ritten **C**haracter **D**ataset (HWCD). As per its name, it consists of 26 handwritten letters of the English, 3 additional letters of the Slovenian alphabet (Č, Š, and Ž), and digits from 0 to 9. Altogether, the HWCD contains 3,314 handwritten character samples across 39 classes. Each sample consists of 10 consecutive vectors with 3 respective components (movement by x-axis, movement by y-axis, and pen down/up). Examples of the character classes that belong to the HWCD are shown in Figure 2.

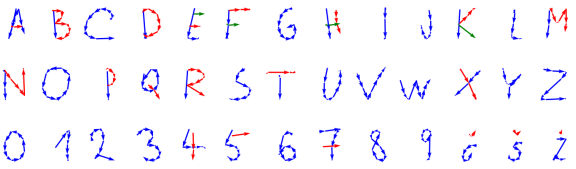


Figure 2. The set of characters in the HWCD. The blue vector color symbolises the first pen stroke, the red color denotes the second stroke, and the green color the third stroke.

The generator G_i that was used in the WGAN training consists of a fully connected layer followed by two 1D convolutional layers. None of the generator layers uses biases. On the other hand, the critic C_i is a multilayer perceptron composed of three fully connected layers. The training hyperparameters of the WGAN are collected in Table 1. The WGAN was implemented in Python 3.11 using the module PyTorch (Paszke et al., 2019).

Table 1. The hyperparameters for the training of WGAN.

Hyperparameter	Symbol	Value
No. of iterations	N	150,000
Batch size	B	256
Critic/generator iterations	R	256
Generator learning rate	η_G	0.0001
Critic learning rate	η_C	0.0001
Gradient penalty term	λ	10

The outputs of the trained WGAN are displayed in Figure 3. It can be seen that all generated samples are recognisable and resemble the real samples from the HWCD, although some samples carry different characteristic that the ones depicted in Figure 2 due to the diverse set of training samples (e.g., number of strokes and distinct local shapes).

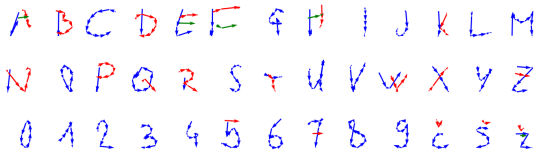


Figure 3. Results of a trained WGAN.

Both pruning approaches yielded similar results, with global pruning achieving marginally superior performance. Therefore, the global pruning method was used throughout the experiments. As critics do not need to be compressed and transferred to the server in a distributed environment, pruning was applied only to G_i. The pruning rate δ of 20% was chosen based on a series of preliminary tests. This means that approximately half of generator weights were pruned after 5 iterations, and approximately 90% of them were frozen after 10 iterations.

In Figure 4, outputs of a pruned WGAN are visualised after different numbers of iterations P. It can be observed that the ability to generate artificial samples that resemble the real samples is pertained even when the pruning rate ρ is high. The generator performance starts to gradually deteriorate after many pruning iterations as some generated characters become unrecognisable (e.g., Q, X, and 8) in Figure 4f. The generator is still able to produce realistic samples even when over 97% of weights are pruned, indicating high efficiency of such approach (Figure 4e).

The generator model is compressed after each iteration of the lottery ticket pruning in order to observe the compressibility of the pruned model. The compression ratios of the generator with the general-purpose compression methods zlib, LZMA, bzip2, and paq9a are collected in Table 2. It can be observed that in the case of a lower number of iterations P, paq9a is the superior compressor, while in the case of a higher number of iterations, LZMA yields better compression ratios. All in all, it can be seen that the generator model can be pruned to the extend where it can be compressed up to 15x without a significant loss of the ability to produce high-quality samples. This can significantly reduce the network load when transferring the model to the server within a distributed learning environment.

Table 2. Compressibility of the pruned generators. The compression algorithm with the highest compression ratio at each is marked bold.

P	% pruned	Compression ratio			
		zlib	LZMA	bzip2	paq9a
/	0.00%	1.08	1.08	1.05	**1.17**
1	20.00%	1.21	1.24	1.23	**1.33**
2	36.00%	1.43	1.49	1.48	**1.59**
3	48.80%	1.70	1.80	1.79	**1.91**
4	59.04%	2.05	2.19	2.17	**2.30**
5	67.23%	2.50	2.68	2.64	**2.79**
6	73.79%	3.08	3.33	3.27	**3.42**
7	79.03%	3.60	3.98	3.92	**4.06**
8	83.22%	4.16	4.68	4.63	**4.75**
9	86.58%	4.79	5.49	5.41	**5.53**
10	89.26%	5.52	**6.40**	6.30	**6.40**
11	91.41%	6.31	**7.46**	7.32	7.35
12	93.13%	7.16	**8.57**	8.40	8.36
13	94.50%	8.11	**9.83**	9.56	9.50
14	95.60%	9.05	**11.12**	10.84	10.64
15	96.48%	9.96	**12.43**	12.02	11.76
16	97.19%	10.86	**13.74**	13.38	12.88
17	97.75%	11.85	**15.17**	14.72	14.14

5 Conclusion

This paper introduces a new method for compression of WGAN generators inside a distributed learning environment. Trained WGANs on separate clients are iteratively pruned with the lottery ticket pruning procedure in the first step. After that, each WGAN is compressed using one of the state-of-the-art general-purpose compression algorithms. The method was extensively tested on vectorised characters of Handwritten Character Dataset. The experiments indicate that the generators can be highly compressed while their ability of generating real-like samples is preserved, which can significantly decrease the network load in a distributed learning environment.

In the future, it would be beneficial for the method to adapt to client data distribution changes with continual learning and on-the-fly pruning of generator models. Generator models could then be periodically sent to the server and aggregated into a global model. Furthermore, using the differential privacy mechanisms, the method could also be modified to operate within a federated learning environment.

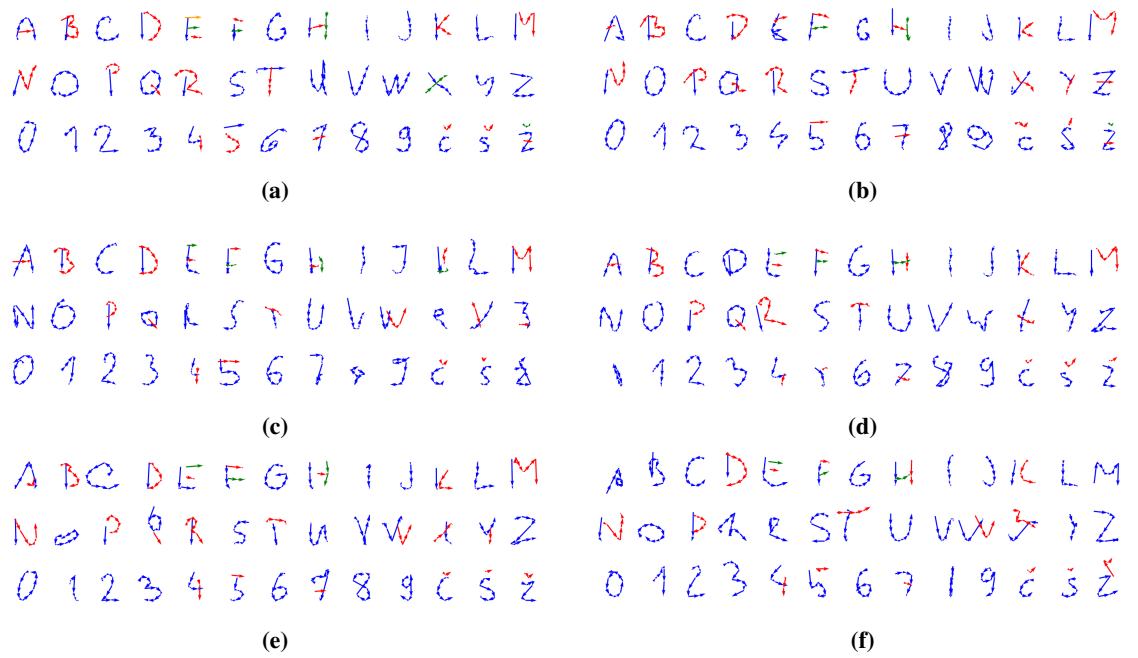


Figure 4. Results of a pruned WGAN: (a) $P = 1$, (b) $P = 5$, (c) $P = 10$, (d) $P = 15$, (e) $P = 16$, (f) $P = 17$.

Acknowledgments

The authors acknowledge the financial support from the Slovenian Research and Innovation Agency under Research Project J2-4458, Young Researcher Funding under Grant 0796-59772, and Research Core Funding P2-0041.

References

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. *ICML'17: Proceedings of the 34th International Conference on Machine Learning*, 214–223.

Ba, L. J., & Caruana, R. (2014). Do deep nets really need to be deep? *NIPS'14: Proceedings of the 28th International Conference on Neural Information Processing Systems*, *2*, 2654–2662.

Bhagyashree, V. Kushwaha, & G. C. Nandi. (2020). Study of prevention of mode collapse in generative adversarial network (GAN). *Proceedings of the 2020 IEEE 4th Conference on Information & Communication Technology (CICT)*, 1–6.

Bzip2 and libbzip2 [Available: https://sourceware.org/bzip2/.]. (2025, May).

Cai, Y., Yao, Z., Dong, Z., Gholami, A., Mahoney, M. W., & Keutzer, K. (2020). ZeroQ: A novel zero shot quantization framework. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13166–13175.

Chang, J., Lu, Y., Xue, P., Xu, Y., & Wei, Z. (2022). Automatic channel pruning via clustering and swarm intelligence optimization for CNN. *Appl. Intell.*, *52*(15), 17751–17771.

Dai, Z., Zhao, L., Wang, K., & Zhou, Y. (2024). Mode standardization: A practical countermeasure against mode collapse of GAN-based signal synthesis. *Appl. Soft Comput.*, *150*, 111089.

Denil, M., Shakibi, B., Dinh, L., Ranzato, M., & de Freitas, N. (2013). Predicting parameters in deep learning. *Proceedings of the 27th International Conference on Neural Information Processing Systems*, *2*, 2148–2156.

Denton, E., Zaremba, W., Bruna, J., LeCun, Y., & Fergus, R. (2014). Exploiting linear structure within convolutional networks for efficient evaluation. *NIPS'14: Proceedings of the 28th International Conference on Neural Information Processing Systems*, *1*, 1269–1277.

Ding, Z., Jiang, S., & Zhao, J. (2022). Take a close look at mode collapse and vanishing gradient in GAN. *Proceedings of the 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*, 597–602.

Durall, R., Chatzimichailidis, A., Labus, P., & Keuper, J. (2020). Combating mode collapse in GAN training: An empirical analysis using Hessian eigenvalues. *arxiv:2012.09673*.

Fang, J., Shafiee, A., Abdel-Aziz, H., Thorsley, D., Georgiadis, G., & Hassoun, J. (2020). Post-training piecewise linear quantization for deep neural networks. *arXiv:2002.00104*.

Frankle, J., & Carbin, M. (2019). The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv:1803.03635*.

Gailly, Jean-loup & Adler, Mark. (2004, December). Zlib compression library [Available: http://www.dspace.cam.ac.uk/handle/1810/3486.].

Goodfellow, I. J., Pouget-Abadie, J., & Mirza, e. a., Mehdi. (2014, December). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems (NIPS 2014)* (Vol. 27). Curran Associates, Inc.

Gou, J., Yu, B., Maybank, S. J., & Tao, D. (2021). Knowledge distillation: A survey. *Int. J. Comput. Vis.*, *129*(6), 1789–1819.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved training of Wasserstein GANs. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 5769–5779.

Han, S., Mao, H., & Dally, W. J. (2016). Deep compression: Compressing deep neural network with pruning, trained quantization and Huffman coding. *Proceedings of the 4th International Conference on Learning Representations (ICLR 2016)*.

Hanson, S., & Pratt, L. Y. (1988, January). Comparing biases for minimal network construction with back-propagation. In D. Touretzky (Ed.), *NIPS'88: Proceedings of the 2nd International Conference on Neural Information Processing Systems* (pp. 177–185, Vol. 1). MIT Press.

He, Y., Liu, P., Wang, Z., Hu, Z., & Yang, Y. (2019). Filter pruning via geometric median for deep convolutional neural networks acceleration. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4335–4344.

Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv:1503.02531*.

Kasturi, A., & Hota, C. (2023). OSGAN: One-shot distributed learning using generative adversarial networks. *J. Supercomput.*, *79*(12), 13620–13640.

Kuang, J., Shao, M., Wang, R., Zuo, W., & Ding, W. (2022). Network pruning via probing the importance of filters. *Int. J. Mach. Learn. Cybern.*, *13*(9), 2403–2414.

Lee, S. H., Kim, D. H., & Song, B. C. (2018). Self-supervised knowledge distillation using singular value decomposition. *Computer Vision – ECCV 2018: 15th European Conference*, 339–354.

Lee, S., Kim, H., Jeong, B., & Yoon, J. (2021). A training method for low rank convolutional neural networks based on alternating tensor compose-decompose method. *Appl. Sci.*, *11*(2), 643.

Li, C., Yang, Y., Feng, M., Chakradhar, S., & Zhou, H. (2016). Optimizing memory efficiency for deep convolutional neural networks on GPUs. *SC '16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 633–644.

Li, H., Wang, Z., Yue, X., Wang, W., Tomiyama, H., & Meng, L. (2023). An architecture-level analysis on deep learning models for low-impact computations. *Artif. Intell. Rev.*, *56*(3), 1971–2010.

Li, J., Niu, K., Liao, L., Wang, L., Liu, J., Lei, Y., & Zhang, M. (2020, July). A generative steganography method based on WGAN-GP. In X. Sun, J. Wang, & E. Bertino (Eds.), *Proceedings of the Artificial Intelligence and Security (ICAIS 2020)* (pp. 386–397). Springer Singapore.

Li, Q., Li, H., & Meng, L. (2022). Feature map analysis-based dynamic CNN pruning and the acceleration on FPGAs. *Electronics*, *11*(18), 2887.

Li, Z., Li, H., & Meng, L. (2023). Model compression for deep neural networks: A survey. *Computers*, *12*(3), 60.

Lin, S., Ji, R., Chen, C., Tao, D., & Luo, J. (2019). Holistic CNN compression via low-rank decomposition with knowledge transfer. *IEEE Trans. Pattern Anal. Mach. Intell.*, *41*(12), 2889–2905.

Lin, S., Ji, R., Li, Y., Wu, Y., Huang, F., & Zhang, B. (2018). Accelerating convolutional networks via global & dynamic filter pruning. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2425–2432.

Liu, M., Wang, Z., Li, H., Wu, P., Alsaadi, F. E., & Zeng, N. (2023). AA-WGAN: Attention augmented Wasserstein generative adversarial network with application to fundus retinal vessel segmentation. *Comput. Biol. Med.*, *158*, 106874.

Lu, Y., Kumar, A., Zhai, S., Cheng, Y., Javidi, T., & Feris, R. (2017). Fully-adaptive feature sharing in multi-task networks with applications in person attribute classification. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1131–1140.

Luo, J. .-., Zhang, H., Zhou, H. .-., Xie, C. .-., Wu, J., & Lin, W. (2019). ThiNet: Pruning CNN filters for a thinner net. *IEEE Trans. Pattern Anal. Mach. Intell.*, *41*(10), 2525–2538.

Mahoney, M. (2025, May). Data Compression Programs [Available: https://mattmahoney.net/dc/].

Malach, E., Yehudai, G., Shalev-Schwartz, S., & Shamir, O. (2020). Proving the lottery ticket hypothesis: Pruning is all you need. *Proceedings of the 37th International Conference on Machine Learning*, *119*, 6682–6691.

Margara, A., Cugola, G., Felicioni, N., & Cilloni, S. (2023). A model and survey of distributed data-intensive systems. *ACM Comput. Surv.*, *56*(1), 1–69.

Merugu, S., & Ghosh, J. (2003, November). Privacy-preserving distributed clustering using generative models. In X. Wu, A. Tuzhilin, & J. Shavlik (Eds.), *Proceedings of the Third IEEE International Conference on Data Mining* (pp. 211–218). IEEE.

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv:1411.1784*.

Molchanov, P., Mallya, A., Tyree, S., Frosio, I., & Kautz, J. (2019). Importance estimation for neural

network pruning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11256–11264.

Paszke, A., Gross, S., Massa, F., Lerer, A., & Bradbury, e. a. (2019). PyTorch: An imperative style, high-performance deep learning library. *arXiv:1912.01703*.

Qin, D., Bu, J.-J., Liu, Z., Shen, X., Zhou, S., Gu, J.-J., Wang, Z.-H., Wu, L., & Dai, H.-F. (2021). Efficient medical image segmentation based on knowledge distillation. *IEEE Trans. Med. Imaging*, *40*(12), 3820–3831.

Rastegari, M., Ordonez, V., Redmon, J., & Farhadi, A. (2016, September). XNOR-Net: ImageNet classification using binary convolutional neural networks. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016* (pp. 525–542). Springer International Publishing.

Rigamonti, R., Sironi, A., Lepetit, V., & Fua, P. (2013). Learning separable filters. *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2754–2761.

Sajjadi Mohammadabadi, S. M., Entezami, M., Karimi Moghaddam, A., Orangian, M., & Nejadshamsi, S. (2024). Generative artificial intelligence for distributed learning to enhance smart grid communication. *Int. J. Intell. Netw.*, *5*, 267–274.

Shomron, G., Gabbay, F., Kurzum, S., & Weiser, U. (2021). Post-training sparsity-aware quantization. *Proceedings of the 35th International Conference on Neural Information Processing Systems*.

Tailor, S. A., Fernandez-Marques, J., & Lane, N. D. (2021). Degree-quant: Quantization-aware training for graph neural networks. *arXiv:2008.05000*.

Tung, F., & Mori, G. (2019). Similarity-preserving knowledge distillation. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*.

Vo, H. V., Du, H. P., & Nguyen, H. N. (2024). APELID: Enhancing real-time intrusion detection with augmented WGAN and parallel ensemble learning. *Comput. Secur.*, *136*, 103567.

Wang, Q., Zhou, X., Wang, C., Liu, Z., Huang, J., Zhou, Y., Li, C., Zhuang, H., & Cheng, J. .-. (2019). WGAN-based synthetic minority over-sampling technique: Improving semantic fine-grained classification for lung nodules in CT images. *IEEE Access*, *7*, 18450–18463.

Wijesinghe, A., Zhang, S., & Ding, Z. (2024). PS-FedGAN: An efficient federated learning framework with strong data privacy. *IEEE Internet Things J.*, *11*(16), 27584–27596.

Wu, J., Leng, C., Wang, Y., Hu, Q., & Cheng, J. (2016). Quantized convolutional neural networks for mobile devices. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4820–4828.

X. Cao, T. Başar, S. Diggavi, & Y. C. Eldar, e. a. (2023). Communication-efficient distributed learning: An overview. *IEEE J. Sel. Areas Commun.*, *41*(4), 851–873.

Xiang, Q., Wang, X., Song, Y., Lei, L., Li, R., & Lai, J. (2021). One-dimensional convolutional neural networks for high-resolution range profile recognition via adaptively feature recalibrating and automatically channel pruning. *Int. J. Intell. Syst.*, *36*(1), 332–361.

XZ Utils [Available: https://tukaani.org/xz/.]. (2025, April).

Yang, T.-J., Chen, Y.-H., & Sze, V. (2017). Designing energy-efficient convolutional neural networks using energy-aware pruning. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6071–6079.

Yim, J., Joo, D., Bae, J., & Kim, J. (2017). A gift from knowledge distillation: Fast optimization, network minimization and transfer learning. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7130–7138.

Yu, X., Liu, T., Wang, X., & Tao, D. (2017). On compressing deep models by low rank and sparse decomposition. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 67–76.

Zhang, Z., Zeng, Y., Bai, L., Hu, Y., Wu, M., Wang, S., & Hancock, E. R. (2020). Spectral bounding: Strictly satisfying the 1-Lipschitz property for generative adversarial networks. *Pattern Recognit.*, *105*, 107179.

Towards User-Centric Chatbot for Q&A in Industry: Evaluation of Employees' Experience

Marko Vještica, Elena Akik, Vladimir Dimitrieski
University of Novi Sad, Faculty of Technical Sciences
Department of Computing and Control Engineering
Trg Dositeja Obradovića 6, 21102, Novi Sad, Serbia
{marko.vjestica, elena, dimitrieski}@uns.ac.rs

Lukas Hinterleitner, Jovan Erić, Frank-Christian Weidenfelder
KEBA Industrial Automation GmbH
Reindlstraße 51, 4040, Linz, Austria
{hilu, eri, weid}@keba.com

Jonas Wieczorek
KEBA Industrial Automation Germany GmbH
Gewerbestraße 5-9, 35633, Lahnau, Germany
Jonas.Wieczorek@keba.com

Abstract. *Chatbots are increasingly being adopted in industrial settings to assist employees in their daily tasks, typically by leveraging the Retrieval-Augmented Generation (RAG) architecture. While technical benchmarking of such chatbots is common, assessing their performance, evaluations of user preferences and chatbots' practical usability are often neglected. To address this gap, we conducted an in-company experiment by deploying an RAG-based chatbot for use by the company's employees. In this paper, we present an analysis of their experience with the chatbot and highlight usability challenges. Based on the experiment findings, we also propose design recommendations for developing user-centric chatbots tailored to industrial environments.*

Keywords. User Experience, Retrieval-Augmented Generation, Chatbot, Large Language Model, Multi-Agent System

1 Introduction

A profound digital transformation is being driven across the global industrial sector through the adoption of advanced Artificial Intelligence (AI) and Machine Learning (ML) technologies (Sarker et al., 2023), enhancing productivity, decision-making, and data management. Among diverse ML technologies, Large Language Models (LLMs) are acknowledged as transformative ML models, automating tasks such as text generation, language translation, and content summarization, in the form of conversational agents (Chen et al., 2024; L. Wu et al., 2024).

While generative and comprehension capabilities are demonstrated by LLMs, they remain limited by their inability to access proprietary or domain-specific knowledge in enterprise settings. To address this deficiency, external retrieval components are employed by the Retrieval-Augmented Generation (RAG) architecture, grounding the LLM's output in domain-specific information (Chen et al., 2024). The RAG architecture that includes an LLM is typically implemented as a chatbot through which natural language communication with end-users is facilitated.

Various RAG-based chatbots serve industrial use cases, yet user experience evaluations, including interaction ease, result satisfaction, and workplace acceptance, are typically neglected. Such evaluations are essential, as practical utility in corporate environments is determined by both technical outcomes and user adoption capacity (Pulapaka et al., 2024). Existing research has focused predominantly on technical performance metrics rather than practical utility and user acceptance. Accordingly, the expectations that users hold toward RAG-based chatbots, such as communication style and response delivery, remain underexplored.

In this paper, a comprehensive user evaluation of an RAG-based chatbot for Question and Answer (Q&A) was conducted with employees from a medium-sized manufacturing company. Participants across various roles interacted with the chatbot to retrieve the company's internal knowledge from various documents, followed by the structured feedback collection. Various user experience dimensions related to the chatbot usage were assessed, including perceived accuracy, response latency, answer structure preferences, ease of use, context selection workflows, language support, and overall satisfaction.

The contribution of this research centers on deploying RAG-based chatbots in authentic industrial environments and systematically measuring multi-

dimensional user experience factors. Thus, empirical connections were established between findings and actionable design recommendations for AI-assisted information retrieval in corporate settings.

Following the Introduction, this paper is organized as follows. In Section 2, related work regarding the application and evaluation of RAG-based systems is discussed. The basic RAG architecture and implementation of a chatbot for Q&A are outlined in Section 3. The evaluation of user experience with the RAG-based chatbot is presented in Section 4. The discussion regarding research findings is presented in Section 5, while conclusions and future work are outlined in Section 6.

2 Related Work

In this section, existing research on the application of RAG-based systems in industry and their evaluation is reviewed. Such systems have been investigated across various industrial domains in order to enhance information retrieval, knowledge management, and decision-support systems.

A novel system was introduced to support internal knowledge sharing within manufacturing setups by leveraging LLMs to interpret and retrieve information from factory documentation and expert insights (Kernan Freire et al., 2024). User studies confirmed accelerated document retrieval, with benchmarking showing state-of-the-art models excelling in factual accuracy while open-source alternatives offered customization and privacy benefits.

A conversational monitoring framework for shop floor decision-making integrated LLMs within RAG to retrieve real-time data from Computer Numerical Control (CNC) machines, Industrial Internet of Things (IIoT) sensors, and manufacturing databases (Jeon et al., 2025). Framework evaluations encompassed experiments with various LLM-based agents and human usability assessments, demonstrating that the framework reliably provided context-aware responses, improved operational efficiency, and achieved high user satisfaction.

LLMs were integrated within RAG for real-time industrial troubleshooting by combining technical manuals, incident reports, and sensor data (Narimani & Klarmann, 2024). This context-aware approach enhanced production efficiency and minimized downtime. Evaluation employed quantitative metrics to assess the relevance of retrieved documents, complemented by expert reviews. Results indicated effective information retrieval, high user satisfaction, and troubleshooting efficiency, although specific quantitative outcomes were not detailed.

Expanding on the manufacturing domain, a framework was proposed for extracting process-level knowledge in additive manufacturing (Liu et al., 2025). LLMs were integrated within the RAG architecture to automatically customize taxonomies for specific manufacturing processes and facilitate the extraction of technical entities. The framework was evaluated by benchmarking in-context learning and fine-tuning against a baseline model, demonstrating superior extraction accuracy and reduced model hallucination.

In digital twin environments, an automated tool was developed for generating Asset Administration Shell (Xia et al., 2024). LLMs converted raw datasheet text into structured formats, while RAG enhanced semantic understanding. Evaluations by graduate annotators included usability assessment, LLM comparisons, and an ablation study, with effectiveness confirmed by bypass rates, helpfulness scores, and quality ratings.

Additionally, a framework was proposed for energy infrastructure management, integrating digital twins and data-driven approaches (Ieva et al., 2024). RAG was employed to enhance a conversational assistant, enabling the retrieval of domain-specific information and enrichment of sensor data and knowledge graphs. Prototype implementations were evaluated in a real-world case study, and improved management, predictive capabilities, timeliness, and accuracy were observed in high-voltage network operations.

In building energy optimization, an LLM-based multi-agent system was described as incorporating RAG to extract information from unstructured energy audit reports (Xiao & Xu, 2024). Case studies demonstrated robust performance at low operational costs through performance testing and benchmarking. Results confirmed the system's ability to handle diverse inputs reliably.

A framework proposed in the construction industry witnessed advancements where RAG applications enhanced document retrieval (C. Wu et al., 2025). A software system parsed documents, such as inspection reports, standards, and working plans. Quantitative benchmarking was employed to evaluate the system, demonstrating improvements in retrieval accuracy and response efficiency.

In the same domain, LLMs were integrated within RAG to enhance information retrieval and process automation (Taiwo et al., 2025). The evaluation combined a Delphi study including domain experts and a benchmarking case study. The findings indicated improvements in output quality, relevance, and reproducibility.

In the telecommunications sector, a system was devised to navigate complex technical standards by constructing a knowledgebase from industry-relevant technical documentation (Yilma et al., 2024). The evaluation was conducted through technical benchmarking, wherein system-generated responses were assessed for accuracy, technical depth, and verifiability. Comparative analyses demonstrated that the proposed system outperformed generic LLMs, producing more precise and verifiable outputs.

In the oil and gas sector, a comprehensive leakage detection approach was presented where LLM-based agents and RAG were used to construct a knowledge vector library from professional documents (Wei et al.,

2024). Improvements in Q&A accuracy and complex reasoning were observed compared to traditional methods. The evaluation was performed through benchmarking against experimental data and further validated in practical deployment scenarios.

In an industrial setting, a conversational RAG-based chatbot for query resolution in Continuous Integration and Continuous Delivery (CI/CD) workflows was developed (Chaudhary et al., 2024). The chatbot was benchmarked on CI/CD questions from an industrial context, showing mostly correct responses with some partial or incorrect answers. It also demonstrated high accuracy in technical queries.

In the research on Electron-Ion Collider (EIC), a RAG-based summarization system was developed to extract information from EIC documentation (Suresh et al., 2024). Relevant content was indexed within a vector database, and citation-enriched summaries were produced by a fine-tuned LLM. System performance was evaluated by using the Retrieval-Augmented Generation Assessments (RAGAs) framework (Es et al., 2024) to improve researcher accessibility and literature navigation.

Evaluation techniques across the reviewed studies on RAG-based applications in industrial domains predominantly relied on quantitative benchmarking to assess performance improvements. Human-based evaluations, such as expert reviews, usability studies, and Delphi evaluations, were used in a smaller subset of investigations to measure user satisfaction with the chatbot's response.

Although benchmarking methods have been widely applied, comprehensive user-centric evaluations have been relatively underrepresented. This gap highlights the need for further research incorporating human usability assessments in order to provide a more complete understanding of the real-world applicability of RAG-enhanced systems. Especially in cases of Q&A applications in industrial domains, a user-centric evaluation of using and applying chatbots for retrieving relevant information is needed. Such an evaluation would contribute to the understanding of what users expect from such systems to use them in practice. Therefore, in the following sections, we present one such RAG-based system for searching documents in a company, and the evaluation results based on the examination of user experience utilizing the presented system and AI assistants in general.

3 Basic RAG Architecture for Documentation Search

To evaluate the user experience with chatbots used for company's internal documentation search and to gain insights into user expectations regarding such software solutions, we implemented an RAG-based chatbot named Document Search Bot (DSB). In this section, we outline a basic RAG architecture on which DSB is built, as well as its implementation. Leveraging content retrieved from document storage, DSB assists users in finding relevant information by enabling natural language interactions.

3.1 RAG Architecture

The basic RAG architecture includes two main flows of activities: (i) the document insertion flow; and (ii) the querying flow. The document insertion flow is discussed first, as it is a prerequisite for querying the documents.

As text within documents represents unstructured data, one way to semantically search for relevant information inside documents is to use a Vector Database Management System (VDBMS). Such a system manages a vector database that stores vector embeddings, representing data of various types in a multi-dimensional vector space.

To insert a document in a vector database, its text needs to be parsed and transformed into vector embeddings. Therefore, the following steps of the document insertion flow are needed, as presented in Fig. 1. First, a user initiates the insertion of documents by sending them to the document parser. The document parser is a component that extracts text from documents and splits it into multiple text chunks. These text chunks are sent to the embedding model, which represents a generative AI model that transforms data into vector embeddings. After the transformation, text chunk embeddings are sent and stored inside the vector database. Meta-data regarding vector embeddings, such as the name of a document and the page number from where they originate, as well as their original content, can be stored inside the vector database as well. The VDBMS can be used afterward to search vector embeddings for relevant information.

Whenever a user aims to search for relevant information in documents, a question can be written in

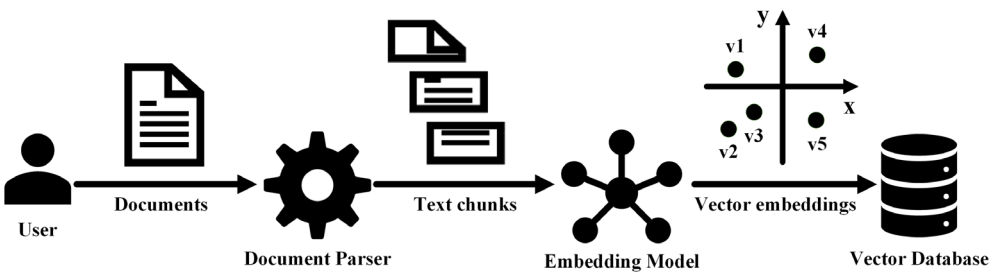


Figure 1. The document insertion flow within a basic RAG architecture

natural language, and the answer is received in the same manner. To achieve such a Q&A feature, the following steps are included in the querying flow, as presented in Fig. 2.

A user initiates querying by writing a question using the chatbot's User Interface (UI). The question is then forwarded to the embedding model, transforming it into a vector embedding, which is sent to the vector database. The distances between the question embedding and the documents' text chunk embeddings are measured by the VDBMS using a chosen similarity metric. Therefore, text chunk embeddings nearest to the question embedding are searched in a multi-dimensional vector space, and their textual representations are sent to the LLM, alongside the question being asked. Based on the provided question, text chunks, and instructions in the form of a prompt, the LLM generates the answer in natural language and sends it to the chatbot's UI, displaying the answer to a user.

3.2 Document Search Bot

DSB represents a Python implementation of the RAG architecture discussed. The document parser is implemented by using the Unstructured library (*Unstructured.io*, 2025). It extracts the text content from paragraphs, tables, and images, but eliminates irrelevant or repeating content, such as that exists in header and footer sections or the table of contents. As documents and questions may be written in various languages, the embedding model needs to support transforming text written in different languages to vector embeddings. Therefore, the Multilingual-E5-large-instruct model (L. Wang et al., 2024) is used for such a purpose. The Milvus VDBMS (J. Wang et al., 2021) is used to store vector embeddings generated from the embedding model, calculate similarity measures using the cosine similarity metric when a question is asked by a user, and retrieve relevant text chunks.

User questions are written on the DSB UI implemented using the Streamlit framework (Khorasani et al., 2022). A user may choose a collection of documents to be searched for an answer before writing a question. A question is added to the prompt sent to the LLM, instructing it to form the answer in the language in which the question is being asked, based on the text chunks provided, or to clearly state whether the answer cannot be found in the given context. The Claude 3.5 Sonnet LLM (*Claude 3.5 Sonnet*, 2025) is used to generate an answer to a user's question, which is displayed on the DSB UI. The LLM is utilized through the Amazon Bedrock service (*Amazon Bedrock*, 2025) to provide an answer rapidly. After receiving an answer, a user may rate the answer and leave a comment using the DSB UI. All questions, received text chunks, generated answers, and user feedback are stored for analysis purposes.

The presented chatbot is given to the participants of an experiment conducted to evaluate the applicability and usability of chatbots in a company. The experiment and evaluation results are presented in the following section.

4 Evaluation of User Experience with Chatbots in Company

In this section, we present an evaluation of the user experience with a chatbot designed for searching documentation in a medium-sized manufacturing company. We outline the preparation of the experiment conducted and the feedback provided by the experiment participants, thus evaluating the applicability and usability of chatbots.

4.1 Experiment Preparation

We prepared an experiment to evaluate user experience with chatbots in the industry. The experiment participants used DSB, outlined in Section 3.2, which was implemented to store and search the company's internal documents. After using DSB, participants filled in a questionnaire regarding the application and usage of chatbots in their daily work. In this section, we present an overview of the experiment conducted and the use cases covered by the experiment.

The experiment was performed in an international medium-sized manufacturing company, KEBA Industrial Automation[1], that develops hardware and

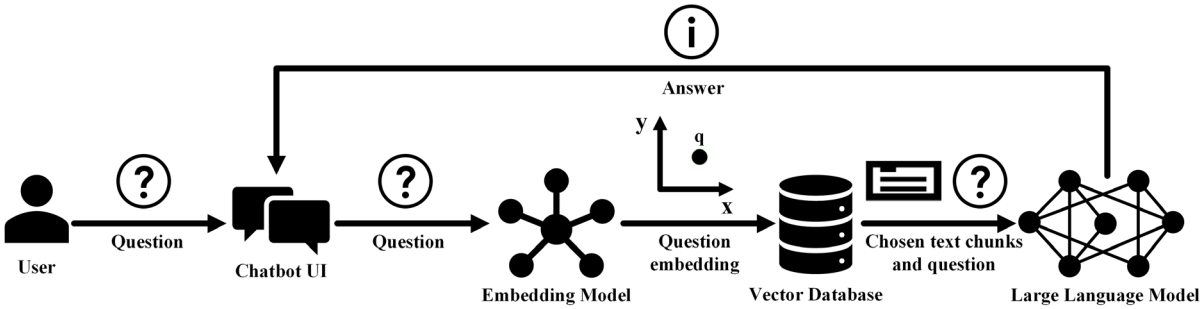


Figure 2. The querying flow within a basic RAG architecture

[1] https://www.keba.com/en/industrial-automation/industrial-automation

software for various industry domains. During the initial discussion with the company's employees, several use cases were identified on where to apply chatbots in the company. First, as the company has a wide range of products, its technical documentation is available to end-users. However, manually searching for relevant information can be time-consuming for end-users, and they usually have to contact the company's service engineers for help. Chatbots may be applied to help: (i) end-users find the information quickly; (ii) service engineers whenever they receive queries from the end-users; and (iii) software engineers when developing the products. Similarly, various standards and norms are applied in the company; thus, searching such extensive documents may be difficult. Additionally, various general-purpose documents in the company, such as house rules and Human Resource (HR) policies, need to be read by employees and usually demand time to locate relevant information. Sometimes, employees contact an HR assistant instead of searching for the documentation. Also, employees sometimes search for or summarize meeting notes, receiving insights into benefits, drawbacks, or conclusions derived from various meetings, which may be time-consuming. Finally, when developing applications, software engineers usually need to use various Application Programming Interfaces (APIs) already implemented by other engineers and consider what parameters to send and what response to expect. Searching for the required APIs may be tedious when working on immense software solutions. LLM-based chatbots may help employees search for relevant information rapidly in all these use cases.

However, as pre-trained LLMs do not contain knowledge regarding the company's internal documents, DSB is utilized for the presented use cases. The DSB's vector database is filled with vector embeddings generated from technical documentation, standards and norms, general-purpose documents, meeting notes, and APIs. The technical documentation included user and operation manuals, application notes, and error lists for various products and their variations, having 50 English and 5 non-English documents with 8706 pages in total. As variations of the same product can be similar to each other, and thus have their documentation, searching for relevant information may be challenging, due to difficulties in making a distinction between similar product variations and their properties. Thus, to make a better distinction between products and their variations, for each one of them, a vector database collection is created, 33 in total. Standards and norms are stored in 51 English and 14 non-English documents with 5195 pages in total. As there are two types of standards and norms, they are split into two vector database collections to narrow the search scope. The general-purpose documentation included 4 English and 4 non-English documents with 103 pages in total, such as house rules, HR policies, and frequently asked questions, stored in a single collection. Meeting notes covered records from 34 meetings written in English and stored in a single collection. DSB also included two collections for 10 APIs written in the JavaScript Object Notation (JSON) and HyperText Markup Language (HTML) formats.

Each collection of documents is presented in the DSB UI, so that a user may choose a specific product variation, a type of standard, general-purpose documents, meeting notes, or a set of APIs, for which a question is to be asked. When a collection is chosen, a link to the stored documents is presented, allowing users to check which exact documents are related to the collection. DSB also provides a single collection of all technical documents for users who prefer asking questions without choosing a specific product variation in advance, but at the expense of lower accuracy in receiving correct answers. Also, when using such a composite collection, a user must ask a question more precisely, thus including the specific product variation name within the question. The DSB UI provides a short description of how to use the chatbot, helping users search for relevant information.

DSB was used by experiment participants for at least two weeks, after which the questionnaire was provided to them. The experiment participants, questionnaire structure, and evaluation results are presented in the following section.

4.2 Evaluation Results

As a part of the experiment, DSB was available for use in the company, allowing employees and external collaborators to ask questions regarding the topics they were interested in. The company's employees who participated in the experiment included 2 service engineers, 8 software engineers, 1 tester, and 2 managers, and the company's external collaborators included 3 researchers and 2 students. Over 600 questions were asked by participants during the experiment.

After using DSB, participants were asked to fill in the online questionnaire, consisting of three main groups of questions: (i) the application of AI assistants in general; (ii) the application of chatbots for Q&A; and (iii) the application of DSB in the company. The questionnaire contained three types of questions whose answers can be: (i) chosen on a five-level Likert scale; (ii) selected from several options provided; or (iii) written as a free text. Questions with a five-level Likert scale are presented in tables in this section, showing the percentage distribution of responses, with level one indicating the lowest and level five the highest rating. Responses to the questions of the other two types are summarized in the text, without tabular representation.

4.2.1 Application of AI Assistants

The first group of questions was asked to examine how familiar participants are with AI assistants, as presented in Table 1. Participants are using AI assistants often, most frequently ChatGPT and MS

Table 1. Questions regarding the application of AI assistants

	Question	1	2	3	4	5
AI assistant application	How often are you using AI assistants?	5.6%	11.1%	33.3%	44.4%	5.6%
	How much practical use do you see in AI assistants in your company?	0.0%	0.0%	5.6%	44.4%	50.0%
	How familiar are you with Large Language Models (LLMs)?	0.0%	16.7%	27.7%	38.9%	16.7%
	How familiar are you with Retrieval-Augmented Generation (RAG) systems?	5.6%	50.0%	22.1%	5.6%	16.7%

Copilot, for gathering information, summarizing reports, writing text, checking grammar, simple coding tasks, building pieces of software, debugging code, and configuring software. Some participants also mentioned they are using AI assistants for transforming text into code, especially when dealing with programming languages they do not use often, and also for translating between different data structures.

Participants recognize much potential for the practical use of AI assistants in the company. They proposed using AI assistants in their company to assist them in writing or reviewing code on specific dialects made in the company, searching internal documents and information, automatically opening IT tickets, or searching the calendar for a free time slot.

While participants reported being mostly familiar with LLMs, their responses indicated limited familiarity with RAG systems. This discrepancy suggests that participants may not be thoroughly acquainted with the underlying technologies powering chatbots and AI assistants in general.

4.2.2 Application of Chatbots for Q&A

As presented in Table 2, when asked about the application of chatbots for Q&A in their company, participants stated that the speed of answering questions is important to them. Waiting too long for the answer may discourage employees from using a chatbot in the company. Despite all participants not speaking English natively, they mostly stated it is not so important for them to communicate with the chatbot in their native language, but there are also participants who would highly appreciate such a feature.

Participants mostly declared it as important to receive the reasoning behind the answer and the estimated confidence score, presumably raising their trust in the answer obtained from the chatbot. When asked about the minimum accuracy they expect from a chatbot to be usable for Q&A, none would expect very low (<50%) or low accuracy (50-70%), 27.7% would expect medium accuracy (70-85%), 55.6% high accuracy (85-95%), and 16.7% very high accuracy (>95%). The accuracy was defined in the questionnaire as the number of correct answers divided by the number of questions asked.

Regarding the answer structure they expect, 50% of participants stated they would like to get short and straightforward answers from a chatbot, while the other 50% of participants would like to get longer, more detailed answers with various notes. Asking any question without having to choose a collection of documents at the expense of having to ask questions more precisely is preferred by 77.8% of participants, while 22.2% of participants prefer choosing a collection before asking questions, but asking them without providing the whole context.

4.2.3 Application of DSB in Company

Regarding the application of DSB in the company, participants were first asked for which use cases they were using DSB during the evaluation process, and 13 used it for searching technical documentation, 4 for standards and norms, 8 for general-purpose documents, 5 for meeting notes, and 6 for APIs. Answers to the five-level Likert scale questions are summarized in Table 3 and grouped under four categories: (i) asking questions to DSB; (ii) receiving answers from DSB; (iii) performance and usability of DSB; and (iv) final remarks and conclusions to DSB.

Participants stated they used DSB occasionally in their daily work, as it was not necessary to frequently search for some information. They mostly declared that it was easy to formulate a question, but sometimes participants had to modify the question to get the correct or complete answer. Based on the analysis of questions being asked by participants, such a need for question alteration was mainly due to the question being ambiguously written or missing some relevant information. However, the question alteration was sometimes due to DSB's failure to find the answer, even if the question was asked precisely. Participants also agreed that the quality of the answer highly depends on the precision of the question's wording, which confirms our question analysis.

Table 2. Questions regarding the application of chatbots for Q&A

	Question	1	2	3	4	5
Chatbots for Q&A application	How important is the speed of answering a question given by a chatbot to you?	0.0%	11.1%	22.2%	55.6%	11.1%
	How important is it for you to communicate in your native language (mother tongue) with a chatbot?	5.5%	50.0%	16.7%	16.7%	11.1%
	How useful would it be for you to receive the reasoning behind the answer from a chatbot?	0.0%	0.0%	38.9%	38.9%	22.2%
	How important is it for you to receive an estimation of the chatbot's confidence in the provided answer?	0.0%	0.0%	16.7%	55.6%	27.7%

Table 3. Questions regarding the application of DSB in the company

	Question	1	2	3	4	5
DSB application: asking questions	How often do you use DSB in your company?	0.0%	33.3%	55.5%	5.6%	5.6%
	How easy is it for you to formulate your questions to DSB?	0.0%	5.6%	22.1%	66.7%	5.6%
	How often did you have to modify the question to get a good answer?	11.1%	16.7%	55.5%	16.7%	0.0%
	How much do you think the quality of an answer depends on the precision of the question's wording?	0.0%	0.0%	11.2%	44.4%	44.4%
DSB application: receiving answers	How satisfied are you with the accuracy of the answers provided by DSB?	0.0%	0.0%	11.1%	83.3%	5.6%
	To what extent do you think the answers are understandable and clearly formulated?	0.0%	0.0%	5.6%	72.2%	22.2%
DSB application: performance and usability	Are you satisfied with how fast DSB replies to you?	0.0%	0.0%	11.1%	61.1%	27.8%
	DSB saves your time in solving problems or getting information.	0.0%	0.0%	16.7%	61.1%	22.2%
	DSB reduces the time it takes you to manually search documents.	0.0%	0.0%	5.6%	61.1%	33.3%
	DSB helps you avoid having to ask other employees for certain information.	0.0%	0.0%	38.9%	50.0%	11.1%
	How intuitive and easily understandable is DSB's user interface to you?	0.0%	0.0%	16.7%	55.5%	27.8%
DSB application: conclusion	How satisfied are you with the current DSB solution overall?	0.0%	0.0%	11.1%	61.1%	27.8%
	In your company, how applicable is DSB in practice?	0.0%	0.0%	22.2%	50.0%	27.8%

Based on their experience, participants were mostly satisfied with the DSB's accuracy, with 44.4% of participants estimating it as medium (70-85%), while 55.6% as high (85-95%). Additionally, similar accuracy was estimated when answers received from DSB were manually checked by a human on a sample of 32 questions asked by participants. When these questions were directed to the composite collection, the chatbot achieved an accuracy of 81.25%. However, when the same questions were sent to product-specific collections, the accuracy rose to 90.63%, which is expected due to the narrower search scope.

Participants praised the received answers for being understandable and clearly formulated. When an answer cannot be found, 11.1% of participants expect to receive only a statement that the answer is not available in the context provided, while 22.2% expect additional information regarding the same topic, and 66.7% expect a reference to a document where an answer may be hidden.

Regarding the performance of DSB, participants were satisfied with how fast it can provide an answer. They stated that DSB saves their time on solving problems, manually searching documents, or asking other employees for certain information. The DSB's UI was intuitive to participants, and they usually did not need help with using it. When asked about drawbacks they encountered when using DSB, one participant stated that the accuracy of DSB is a bit lower when asking questions in a language that differs from the one on which documents are written. Also, two participants appear to receive lower accuracy when asking questions whose answers are distributed across tables, and in the documents provided, some tables are spread across many pages.

To conclude with the DSB application, participants were asked questions about Net Promoter Score (NPS) (Baehre et al., 2022), overall satisfaction with the solution, and their suggestions on how to improve it. With 10 promoters, 7 passives, and 1 detractor, the NPS value was 50, indicating a positive user experience. The reasons behind providing such a score, participants stated DSB to be useful, easy and fast, saving their time, and being accurate enough for daily operations. Also, two service engineers revealed that DSB can be used by new employees to delve into technical documentation, while another participant acknowledged that DSB helps when searching for documents not written in the user's known languages.

To improve their experience with DSB, participants suggested allowing them to insert their documents and thus have personal document collections. Also, they proposed including all documents in DSB that the company has. Additionally, they recommended integrating the chatbot within existing software used in the company to easily search for information related to that software. Participants were mostly satisfied with DSB overall, and they deemed it applicable in the company, both for internal usage by employees and external usage by customers of various company products.

4.3 Threats to Validity

Despite efforts to minimize them, several threats to the validity of evaluation results need to be discussed in this section, as the results may vary when conducting the experiment on different occasions.

The experiment was performed in a single company, including 18 participants in different roles who worked in various company sectors. However, having multiple companies participate in the experiment might provide more accurate evaluation results. Similarly, having a larger group of participants, either from a single company or multiple companies, might provide different results, but booking their time in a company for the experiment could be challenging.

The participants used DSB for at least two weeks during the experiment. However, using the chatbot during a longer period might change the user

experience with it, either positively by experiencing new ways to gather information, or negatively by receiving insights into possible drawbacks of the chatbot.

Regarding the questionnaire, we aimed to formulate precise and unambiguous questions in English, as this was the language known by all participants, but not their native language. Thus, some questions might appear less clear to participants.

DSB offered a reasonable solution for the experiment. However, if the accuracy of the provided chatbot was lower or higher, the user experience might be different, and thus the evaluation results as well. The DSB's accuracy might be influenced by its implementation and technology used, the number and structure of pages in documents, or the number of documents per collection. We aimed to provide participants with a basic RAG-based chatbot that would probably be developed in various companies as the first solution. However, we implemented the chatbot with contemporary technologies available at the time of conducting the experiment, thus offering a solution with decent accuracy.

5 Discussion

Based on the evaluation results presented in this paper, users are aware of AI assistants and notice the potential to apply them in a company, especially in cases of information retrieval, text summarization, and code writing. However, they might not be familiar with the technologies behind AI and how AI assistants can be developed to help them in their daily work. Thus, various demonstrations of AI possibilities and prototypes developed in a company might help users recognize exact use cases in which AI assistants or chatbots, in particular, can be valuable.

As discussed in our previous study (Vještica et al., in press), in which statistical analysis was performed on participants' feedback, the key user satisfaction factor represents the time-saving benefit in information retrieval, compared to manual search. Achieving such a benefit is primarily influenced by a high answer accuracy and secondarily by a low response time.

Therefore, a user-centric chatbot for Q&A should provide answers relatively fast with high accuracy, offering an efficient and reliable solution. The evaluation participants mostly expect a chatbot to have an accuracy of at least 90%. Optionally, such a chatbot should provide users with reasoning and confidence scores, thus additionally increasing their trust in the answers received.

Also, users might find it beneficial that a user-centric chatbot is customizable, allowing them to: (i) upload their own documents; (ii) define how correct answers are structured; (iii) configure fallback responses when an answer is not found; and (iv) choose whether to select a document collection prior to posing a question or instead formulate more specific questions that include the exact topic of interest.

Due to the tendency to pose ambiguous or incomplete questions, i.e., without providing sufficient contextual information, users might not receive answers or receive inaccurate ones, often prompting them to reformulate their questions. Therefore, a chatbot needs to engage in interactive guidance, helping users to articulate precise questions with adequate context.

The presented study provided an in-situ deployment of an RAG-based chatbot within a company, contrasting with prior research that has focused on domain-specific prototypes evaluated primarily through technical benchmarks. Therefore, this work is positioned as complementary to existing research by emphasizing real-world usability and challenges of RAG-based chatbots. Overall, the main contribution of this study is the empirical investigation of chatbots' user acceptance, accuracy trade-offs, and deployment challenges. Based on this investigation, design recommendations are proposed for industrial RAG-based systems.

6 Conclusions and Future Work

In this paper, a user experience evaluation of an RAG-based chatbot was conducted in a medium-sized company. Accordingly, a preliminary case-specific observation of user requirements for a chatbot applied in industrial environments is provided. The chatbot for Q&A named DSB was developed for evaluation purposes and used by company staff who provided valuable feedback. Based on the feedback gathered from evaluation participants and the analysis of the DSB usage, the design recommendations for a user-centric Q&A chatbot were derived and discussed in this paper.

Accordingly, we plan to enhance DSB in a way that provides a better user experience and assess user satisfaction with the novel chatbot solution. The envisioned user-centric version of DSB might be developed as a multi-agent system offering: (i) higher accuracy through improved parsing of documents, especially when extracting text from tables and images; broader context coverage in question answering; and advanced search and reranking capabilities; (ii) enhanced management of vast amount of data and automatic selection of relevant collections based on topics inferred from a question; (iii) integration with various sources, including existing databases; (iv) interactive handling of ambiguous or inaccurate questions provided by users; (v) customizable answer formatting; and (vi) support for personal document collections. The proposed multi-agent system could prolong the time required for generating answers. Thus, a high-performance infrastructure, such as a cloud infrastructure, might be needed to provide answers fast and maintain user satisfaction.

To examine whether the proposed improvements yield measurable benefits, both user experience and the chatbot's accuracy should be evaluated and compared with the basic DSB solution described in this paper. Therefore, we plan to expand the user experience assessment by including a broader participant base, thereby gaining deeper insights into user requirements for an effective Q&A chatbot in industrial contexts. Furthermore, a systematic evaluation of the chatbot's accuracy is needed to complement the user experience evaluation, as the chatbot's accuracy has a high impact on user satisfaction. Frameworks such as RAGAs can be utilized for this purpose, enabling the evaluation of metrics including: (i) the faithfulness of generated answers with respect to the provided context; (ii) the relevance of the answer to the posed question; and (iii) the relevance of the retrieved context to the question's content and scope.

Acknowledgments

This research has been supported by KEBA Industrial Automation; and by the Ministry of Science, Technological Development and Innovation (Contract No. 451-03-137/2025-03/200156) and the Faculty of Technical Sciences, University of Novi Sad through project "Scientific and Artistic Research Work of Researchers in Teaching and Associate Positions at the Faculty of Technical Sciences, University of Novi Sad 2025" (No. 01-50/295). We are thankful to the evaluation participants for their time and useful feedback.

References

Amazon Bedrock. (2025). Retrieved March 25, 2025, from https://aws.amazon.com/bedrock/

Baehre, S., O'Dwyer, M., O'Malley, L., & Lee, N. (2022). The use of Net Promoter Score (NPS) to predict sales growth: Insights from an empirical investigation. *Journal of the Academy of Marketing Science*, *50*(1), 67–84. https://doi.org/10.1007/s11747-021-00790-2

Chaudhary, D., Vadlamani, S. L., Thomas, D., Nejati, S., & Sabetzadeh, M. (2024). Developing a Llama-Based Chatbot for CI/CD Question Answering: A Case Study at Ericsson. *Proceedings of the 2024 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 707–718. https://doi.org/10.1109/ICSME58944.2024.00075

Chen, J., Lin, H., Han, X., & Sun, L. (2024). Benchmarking Large Language Models in Retrieval-Augmented Generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(16), 17754–17762. https://doi.org/10.1609/aaai.v38i16.29728

Claude 3.5 Sonnet. (2025). Retrieved March 25, 2025, from https://www.anthropic.com/claude/sonnet

Es, S., James, J., Espinosa Anke, L., & Schockaert, S. (2024). RAGAs: Automated Evaluation of Retrieval Augmented Generation. *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, 150–158. https://doi.org/10.18653/v1/2024.eacl-demo.16

Ieva, S., Loconte, D., Loseto, G., Ruta, M., Scioscia, F., Marche, D., & Notarnicola, M. (2024). A Retrieval-Augmented Generation Approach for Data-Driven Energy Infrastructure Digital Twins. *Smart Cities*, *7*(6), 3095–3120. https://doi.org/10.3390/smartcities7060121

Jeon, J., Sim, Y., Lee, H., Han, C., Yun, D., Kim, E., Nagendra, S. L., Jun, M. B. G., Kim, Y., Lee, S. W., & Lee, J. (2025). ChatCNC: Conversational machine monitoring via large language model and real-time data retrieval augmented generation. *Journal of Manufacturing Systems*, *79*, 504–514. https://doi.org/10.1016/j.jmsy.2025.01.018

Kernan Freire, S., Wang, C., Foosherian, M., Wellsandt, S., Ruiz-Arenas, S., & Niforatos, E. (2024). Knowledge sharing in manufacturing using LLM-powered tools: User study and model benchmarking. *Frontiers in Artificial Intelligence*, *7*, 1293084. https://doi.org/10.3389/frai.2024.1293084

Khorasani, M., Abdou, M., & Hernández Fernández, J. (2022). *Web Application Development with Streamlit: Develop and Deploy Secure and Scalable Web Applications to the Cloud Using a Pure Python Framework*. Apress. https://doi.org/10.1007/978-1-4842-8111-6

Liu, X., Erkoyuncu, J. A., Fuh, J. Y. H., Lu, W. F., & Li, B. (2025). Knowledge extraction for additive manufacturing process via named entity recognition with LLMs. *Robotics and Computer-Integrated Manufacturing*, *93*, 102900. https://doi.org/10.1016/j.rcim.2024.102900

Narimani, A., & Klarmann, S. (2024). Integration of Large Language Models for Real-Time Troubleshooting in Industrial Environments based on Retrieval-Augmented Generation (RAG). *Proceedings of the 7th European Conference on Industrial Engineering and Operations Management*, 287–298. https://doi.org/10.46254/EU07.20240085

Pulapaka, S., Godavarthi, S., & Ding, Dr. S. (2024). GenAI-Powered Chatbots. In S. Pulapaka, S. Godavarthi, & S. Ding (Eds.), *Empowering the Public Sector with Generative AI: From Strategy and Design to Real-World Applications* (pp. 157–190). Apress. https://doi.org/10.1007/979-8-8688-0473-1_6

Sarker, S., Arefin, M. S., Kowsher, M., Bhuiyan, T., Dhar, P. K., & Kwon, O.-J. (2023). A Comprehensive Review on Big Data for Industries: Challenges and Opportunities. *IEEE Access*, *11*, 744–769. https://doi.org/10.1109/ACCESS.2022.3232526

Suresh, K., Kackar, N., Schleck, L., & Fanelli, C. (2024). Towards a RAG-based summarization for the Electron Ion Collider. *Journal of Instrumentation*, *19*(07), C07006. https://doi.org/10.1088/1748-0221/19/07/C07006

Taiwo, R., Bello, I. T., Abdulai, S. F., Yussif, A.-M., Salami, B. A., Saka, A., Ben Seghier, M. E. A., & Zayed, T. (2025). Generative artificial intelligence in construction: A Delphi approach, framework, and case study. *Alexandria Engineering Journal*, *116*, 672–698. https://doi.org/10.1016/j.aej.2024.12.079

Unstructured.io. (2025). Retrieved March 25, 2025, from https://unstructured.io/

Vještica, M., Akik, E., Dimitrieski, V., Hinterleitner, L., Erić, J., Weidenfelder, F.-C., & Pisarić, M. (in press). Evaluation of User Experience with RAG-based Chatbots for Searching Documentation: Industrial Case Study. *Proceedings of the 33rd International Conference on Information Systems Development (ISD 2025)*.

Wang, J., Yi, X., Guo, R., Jin, H., Xu, P., Li, S., Wang, X., Guo, X., Li, C., Xu, X., Yu, K., Yuan, Y., Zou, Y., Long, J., Cai, Y., Li, Z., Zhang, Z., Mo, Y., Gu, J., … Xie, C. (2021). Milvus: A Purpose-Built Vector Data Management System. *Proceedings of the 2021 International Conference on Management of Data*, 2614–2627. https://doi.org/10.1145/3448016.3457550

Wang, L., Yang, N., Huang, X., Yang, L., Majumder, R., & Wei, F. (2024). *Multilingual E5 Text Embeddings: A Technical Report* (No. arXiv:2402.05672). arXiv. https://doi.org/10.48550/arXiv.2402.05672

Wei, Q., Sun, H., Xu, Y., Pang, Z., & Gao, F. (2024). Exploring the Application of Large Language Models Based AI Agents in Leakage Detection of Natural Gas Valve Chambers. *Energies*, *17*(22), 5633. https://doi.org/10.3390/en17225633

Wu, C., Ding, W., Jin, Q., Jiang, J., Jiang, R., Xiao, Q., Liao, L., & Li, X. (2025). Retrieval augmented generation-driven information retrieval and question answering in construction management. *Advanced Engineering Informatics*, *65*, 103158. https://doi.org/10.1016/j.aei.2025.103158

Wu, L., Zheng, Z., Qiu, Z., Wang, H., Gu, H., Shen, T., Qin, C., Zhu, C., Zhu, H., Liu, Q., Xiong, H., & Chen, E. (2024). A survey on large language models for recommendation. *World Wide Web*, *27*(5), 60. https://doi.org/10.1007/s11280-024-01291-2

Xia, Y., Xiao, Z., Jazdi, N., & Weyrich, M. (2024). Generation of Asset Administration Shell With Large Language Model Agents: Toward Semantic Interoperability in Digital Twins in the Context of Industry 4.0. *IEEE Access*, *12*, 84863–84877. https://doi.org/10.1109/ACCESS.2024.3415470

Xiao, T., & Xu, P. (2024). Exploring automated energy optimization with unstructured building data: A multi-agent based framework leveraging large language models. *Energy and Buildings*, *322*, 114691. https://doi.org/10.1016/j.enbuild.2024.114691

Yilma, G. M., Ayala-Romero, J. A., Garcia-Saavedra, A., & Costa-Perez, X. (2024). TelecomRAG: Taming Telecom Standards with Retrieval Augmented Generation and LLMs. *ACM SIGCOMM Computer Communication Review*, *54*(3), 18–23. https://doi.org/10.1145/3711992.3711996

How ChatGPT and Gemini View the Elements of Communication Competence of Large Language Models: A Pilot Study

Goran Bubaš
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin
Croatia
gbubas@foi.unizg.hr

Abstract. *A concise overview is provided of selected theoretical models of communication competence in the fields of linguistics, interpersonal communication, second language use, and human-robot interaction. The following practical research consisted of two case studies with the goals of investigating how advanced AI tools like ChatGPT and Gemini interpret elements of two communication competence theories in the context of Large Language Model (LLM) interactions with users. The focus was on these theoretical approaches: (1) an integrated linguistic-interpersonal model and (2) an interpersonal 'human-humanoid' interaction model. The conclusion is that both approaches are suitable for a better understanding of LLM-user interaction.*

Keywords. Large Language Models, ChatGPT, Gemini, communication competence, linguistics, interpersonal communication, communication competence of artificial intelligence systems

1 Introduction

According to Chang et al. (2024), because of their widespread use, the evaluation of Large Language Models (LLMs) from various perspectives has become crucial from the *task execution* level to the *social* level. Chang et al. state that this incorporates the capacity of LLMs for *natural language understanding and generation*, including *fluency* (the ability to produce textual content "that flows smoothly, maintaining a consistent tone and style") and *human alignment* (the degree to which the output relates to "human values, preferences, and expectations", as well as "societal norms and user expectations, promoting a positive interaction with human users"). In the theoretical part of the following study, first, a brief overview of the development of *linguistic* approaches to human communication competence, as well as approaches to communication competence from an *interpersonal communication skills* perspective, is presented. Following this, there are concise summaries of (a) the *integrated model of linguistic and interpersonal communication competence in second language use* developed by Bubaš and Kovačić (2019) and of (b) an early model of *communication competence of artificial systems* introduced by Bubaš and Lovrenčić (2002). In the two case studies that are also presented in this paper, the communicative abilities of LLMs are investigated in detail from these two theoretical perspectives. It must be noted that, for this examination, the 'views' of advanced LLMs of the ChatGPT and Gemini family are utilized regarding the applicability of the content of the two theoretical studies (Bubaš and Kovačić, 2019; Bubaš and Lovrenčić, 2002) for interpreting various communication ('competence') aspects of interaction of LLMs with their users. The presented study is relevant because communication competence related aspects of LLM-user interaction have been scarcely and only fragmentarily investigated. Therefore, this study represents a novel theoretical and practical approach to the investigation of communication of LLMs with their users. In the continuation of this article, a theoretical introduction is provided with two theoretical models of communication competence explained in more detail. After the definition of the main goal and research questions, study methodology is explained with details of prompting procedures for two case studies. The results of the two case studies are interpreted in relation to the research questions (RQ1 – RQ3) and briefly discussed. The article resumes with a concise conclusion and limitations of the study.

2 Communication Competence Theoretical Models

This theoretical section first provides a summary of the introduction of the concept of *communication competence* in the fields of *linguistics* (in *subchapter 2.1*) and *communication science* (in *subchapter 2.2*) with a specific focus of the latter on *interpersonal interaction*. Then, the model of linguistic and interpersonal communication competence in second language use is briefly outlined (in *subchapter 2.3*) that *integrates linguistic and interpersonal aspects of communication competence* and can be used in theoretical interpretations of communication of LLMs with users. Finally, a concise summary of the *model of communication competence of artificial systems* is presented (in *subchapter 2.4*) that was created primarily for the *theoretical interpretation of*

human-robot interaction, but can also be adapted to the context of interaction of LLMs with users.

2.1 Linguistics and communication competence

The concept of *linguistic competence* was developed by Chomsky (1965, pp. 3-10) as an 'idealized' *intrinsic knowledge* of language possessed by a speaker-hearer, distinct from the actual, *real-world language performance* of the individual. This limited view of linguistic competence was challenged in a series of publications by Hymes (for instance, see Hymes, 1972), who introduced the term *communicative competence*. For Hymes, this concept included not only linguistic competence but also *pragmatic*, *socio-cultural*, and *sociolinguistic* aspects of language use.

Within the linguistic line of research, Canale and Swain (1980) proposed the following components of communication competence: *grammatical* (knowledge of grammar, syntax, vocabulary, morphology, and phonology); *sociolinguistic* (how to appropriately use language in a social context, including situational and cultural aspects); *strategic* (use of verbal and non-verbal strategies concerning grammatical and sociolinguistic competence). Canale (1983) later added the component of *discourse* competence as the capacity to "combine grammatical forms and meanings to achieve a unified spoken or written text in different genres".

The notable theoretical contributions in this field include the work of Bachman and Palmer (1995, 67-75), who proposed the term *communicative language ability*, one component of which is *language knowledge*, which can be further divided into (a) *organizational knowledge*, composed of grammatical and textual knowledge, and (b) *pragmatic knowledge*, consisting of functional and sociolinguistic knowledge. According to Bachman and Palmer, *functional knowledge* includes the expression of the individual's experiences with the real world, manipulation of other people, and development of relationships with them. Also, *sociolinguistic knowledge* enables the adaptability of created or interpreted language to a specific interactional context, for instance, in terms of applicable conventions and potential appropriateness evaluations. Finally, these authors consider *strategic competence* as a set of metacognitive components operating in the practical areas of goal and task setting, assessment of the task and its actualization, as well as planning for completing or implementing the goal or task.

Various versions of linguistic models of communication competence have also been developed, but most of them encompass similar constructs as previously mentioned. For instance, the model by Celce-Murcia et al. (1995) introduces a triangular organization of *sociocultural*, *linguistic*, and *actional* competence, which are all in interaction with *discourse* and *strategic* competence.

There are several historical overviews on the approaches to communication competence in linguistics and language learning (including, for instance, Bagarić & Mihaljević Djigunović, 2007), as well as efforts to integrate the related theoretical concepts (e.g. Usó-Juan & Martínez-Flor, 2006) for application in language learning. It is important to mention that, while the domain of *language teaching and learning practice* has predominantly focused on the *four skills* (listening, speaking, reading, and writing) as learners' abilities for effective language use, the field of *linguistics* (and applied linguistics in particular) has seen a continuation of the interest in the concepts of *grammatical* competence and language knowledge, as well as the variations of *sociolinguistic*, *strategic*, and *discourse* competence, including their *pragmatic aspects* and integration into more holistic models.

Recently, efforts have been made to associate linguistic aspects of communication competence with *conversational artificial intelligence* and *Large Language Models* like ChatGPT.

2.2 Interpersonal communication and communication competence

In the field of interpersonal communication, as a discipline within *Communication Science*, the most extensive early theoretical research on communication competence was performed by Spitzberg and Cupach (1984; 1989). Their work aimed to conceptualize communication competence in *interpersonal social interaction*, distinct from the theoretical and practical areas of linguistics. These authors investigated and reflected on numerous studies of interpersonal communication skills to form a general approach with a tripartite model comprising (1) interpersonal or social *skills* that may contribute to communication competence, (2) *knowledge* that can be, for instance, procedural or tacit/indeliberate, and (3) *motivation* that may be broadly denoted by the approach-avoidance dimension in interactions. Spitzberg and Cupach (2002) also emphasized the need for criteria of skillful interpersonal interaction, like, for instance, *effectiveness*, *efficiency*, *appropriateness,* and *satisfaction*. In their "heuristic for taxonomies" of more than 100 labels for communication skills found in the literature, they also defined the "macroscopic" level of interpersonal interaction, as well as "microscopic" and "mezzoscopic" skills. For the 'in-depth' perspective of interpersonal communication competence, apart from the condensed overview of the historic developments around this term in the communication discipline by Backlund and Morreale (2015), it is also useful to consider the analysis of communicative competence as a theoretical concept by Wilson and Sabee (2003). It must be emphasized that a newer model of *computer-mediated communication (CMC) competence*, developed by Spitzberg (2006), (a) is based on the concepts of knowledge, motivation, and skills, (b) includes outcomes like appropriateness, effectiveness, coorientation, satisfaction, and relational development, and (c) also comprises several additional components including various media, message, and contextual factors.

It can also be mentioned that an effort to utilize empirical research on the dimensions of *human* interpersonal communication competence (see: Bubaš et al., 1999; Bubaš, 2001a; Bubaš, 2001b; Bubaš, 2003) with potential communication competence dimensions of *artificial behavioral (intelligent) systems that interact with humans* was performed more than two decades ago by Bubaš and Lovrenčić (2002).

Regarding LLMs like ChatGPT, more recently, researchers have focused on a small number of specific interpersonal communication skills that may be manifested by this AI tool, including *empathy* (see: Sorin et al., 2024; Luo et al., 2024; Welivita & Pu, 2024) and *persuasion* (Bai et al., 2025; Rogiers et al., 2024; Goel et al., 2024), rather

than on a larger set of skills and their manifestation in complex behavior similar to *interpersonal communication competence* in general. By contrast, Bubaš (2024) has demonstrated the ability of LLMs like ChatGPT 4 and ChatGPT 4o, Copilot, Claude 3.5 Sonnet, and Gemini 1.5 Pro, to analyze existing items and develop new ones for self-assessment scales related to a large number of different interpersonal communication skills.

2.3 Integrated model of linguistic and interpersonal communication competence in second language use

To integrate (a) the *linguistic* elements with (b) the *interpersonal* interaction and *skills* elements of theoretical models of communication competence, Bubaš and Kovačić (2019) have developed a *conceptual model of communication competence in foreign language* (L2) *use* (CMCC-L2) and interpreted this model in the context of teaching and learning L2 and *English as a Foreign Language* (EFL). This new model adapted and integrated the "framework of communicative competence integrating the four skills" (Usó-Juan & Martínez-Flor, 2006), the molar and molecular view of interpersonal communication competence (Spitzberg & Cupach 1984), different levels of observing interaction processes – microscopic, mezzoscopic, macroscopic (Spitzberg, 1994), including similar levels of analysis in the social ecological framework (Bronfenbrenner, 1977; Ting-Toomey, 2012), and communication accommodation theory (Zhang, 2018). The outline of the levels of the *conceptual model of communication competence in foreign language use* (L2) is presented in Table 1 (mostly adapted from Usó-Juan & Martínez-Flor, 2006).

Table 1. Levels in the conceptual model of communication competence in a foreign language (L2)

LEVEL	COMPETENCE TYPE
Supra	**Social/intercultural** *(utilization of knowledge, social/cultural cues, and skills to understand the interaction environment and appropriately perform wide-ranging sequences of intentional communication acts)*
Macro	**Strategic/adaptive** *(knowledge and use of interaction strategies, learning to adapt and advance in competence, utilization of specific skills to enhance ability and overcome barriers)*
Mezzo	**Pragmatic/action/discourse** *(performing and interpreting concise speech acts, monologue, and dialogue according to participant and situational variables)*
Micro	**Linguistic** *(lexicon, phonology, orthography, morphology, syntax, sentence sequencing)*

In Fig. 1, using the basic *motivation-knowledge-skills* concept as the potential for interpersonal competence, as well as *outcomes* concepts of Spitzberg (2006) and Spitzberg and Cupach (2002), the *conceptual model of communication competence in foreign language (L2) use* (CMCC-L2) is presented, adapting and further developing these models to define "enactment of skill(s)" and "L2 competence confirmation" (or "social outcomes"). The list of "skills" as a potential for competence is comprehensive: *willingness to communicate, initiation of interaction, listening, self-disclosure, nonverbal sensitivity, self-monitoring, impression management, questioning, empathy, persuasion/assertiveness, conversational skill, expressiveness, social support, and interaction management*. The "enactment of skill(s)" is at the interpersonal interaction levels of (1) *pragmatic/action/discourse*, (2) *strategic/adaptive*, and (3) *social/intercultural* competence and includes selected skills for each of the previous levels (1-3). Finally, the "L2 competence confirmation" and related "social outcomes" can be: *understanding, appropriateness, influence, coordination, satisfaction, cooperation, efficacy (goal attainment), attractiveness, relationship, inclusion, socialization, learning, competence enhancement, self-development, growth, well-being, and community building*. Basically, this model illustrates how specific communication *skills* (as *potential* or *capacity*, including *knowledge* and *motivation* components) can be used to facilitate or enhance the practical manifestation of the *dimensions of L2 communication competence* (enactment of skills) and produce various social outcomes that are confirmation of L2 communication competence.

Some theoretical concepts from the field of linguistics (including several elements that are mentioned in the *conceptual model of communication competence in L2 use* by Bubaš and Kovačić, 2019), have already been utilized in the analyses of communicative abilities of LLMs like ChatGPT, however a similar *integrative model* has not yet been found in the literature. For instance, an extensive survey of language model behaviors analyzed *separately* their abilities regarding *syntax*, *semantics* and *pragmatics* (Chang & Bergen, 2024), while more focused studies, among others, for instance, investigated linguistic competence of ChatGPT from the aspects of *pragmatic awareness* regarding impoliteness (Andersson & McIntyre, 2025), expressive and receptive *pragmatic skills* (Barattieri di San Pietro, 2023), *sociolinguistic competence* (Duncan, 2024), formal versus functional *linguistic competence* (Mahowald et al., 2024), *linguistic competence* regarding Italian dialects (Lilli, 2023), and *intercultural competence* (Lee, 2025).

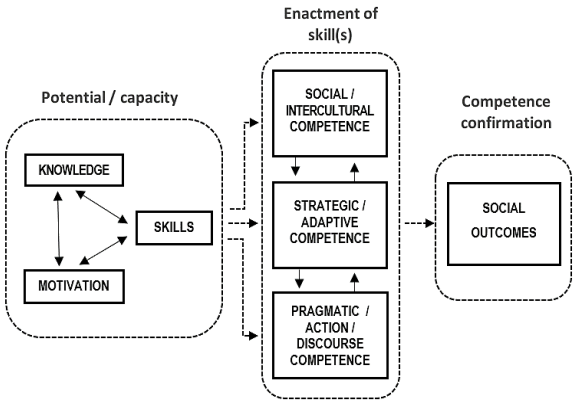


Figure 1. Relationship among the elements of the *conceptual model of communication competence in foreign language (L2) use* (adapted from Bubaš & Kovačić, 2019)

2.4 An early model of communication competence of artificial systems

Four decades ago, Spitzberg & Cupach (1984) identified more than 100 labels for communication skills and traits in scholarly literature in relation to human interpersonal interaction. Much later, using self-assessment scales for 23 different communication skills and traits and factor analyses, in empirical investigations of the *dimensions* of human interpersonal communication competence (Bubaš et al., 1999; Bubaš, 2001a; Bubaš, 2001b; Bubaš, 2003) the following six potential dimensions were identified (with numerous specific communication skills as their subcomponents): (1) encoding and decoding, (2) intentionality, (3) communication effectiveness, (4) other-orientedness, (5) expressiveness, and (6) social relaxation. These empirically revealed dimensions were used to develop a conceptual model of *communication competence of artificial systems* (CCAS) that interact with humans (Bubaš & Lovrenčić, 2002). However, at the time this CCAS model was developed (in the year 2002), when compared to the dimensions of *human* communication competence, the developments in *humanoid robotics* were only at the general level of the *decoding and encoding dimension* and the rudimentary *expressivity dimension*. It must be emphasized that in the CCAS model, a formalization of the generalized dimensions of interpersonal communication competence was outlined (in the context of *human-humanoid interaction*), including skills and traits that contribute to those dimensions. Finally, in their paper, Bubaš and Lovrenčić (2002) have suggested possible implementations of the formalized CCAS model in the design of interfaces that could enhance the *effectiveness* and *appropriateness* in communication of an artificial (intelligent) system with humans. It was only after the launch of OpenAI's ChatGPT in November 2022 that most of the other components of the CCAS model could be applied to ChatGPT and other families of LLMs (e.g., Anthropic's Claude, Google AI's Gemini, etc.) to interpret interactions between LLMs and their users.

In the article about the CCAS model, the intention was to present in detail the following dimensions (and not more than *four related communication skills* for each dimension): decoding and encoding (*nonverbal sensitivity, verbal understanding, verbal and nonverbal encoding, self-monitoring*), communication effectiveness (*initiation of interaction, assertiveness, interaction management, adaptability*), and other-orientedness (*empathy, support, self-disclosure, collaboration*). All six dimensions of CCAS are depicted in Fig. 2.

In addition, the CCAS model organized communication behaviors of artificial systems in the following levels:

- *Macroscopic* level (dimensions of communication competence).
- *Mezzoscopic* level (sets of skills and traits that are related to each of the dimensions of communication competence).
- *Microscopic* level (sets of "communication primitives"; moderately complex verbal and nonverbal behaviors, behavioral scripts, skill-related knowledge, communication traits, interaction tactics, etc., that contribute to the pragmatic enactment of a specific communication skill).
- *Atomic* level (sets of simple perceptual or motoric behaviors that are engaged in the realization of the specific communication "primitives").

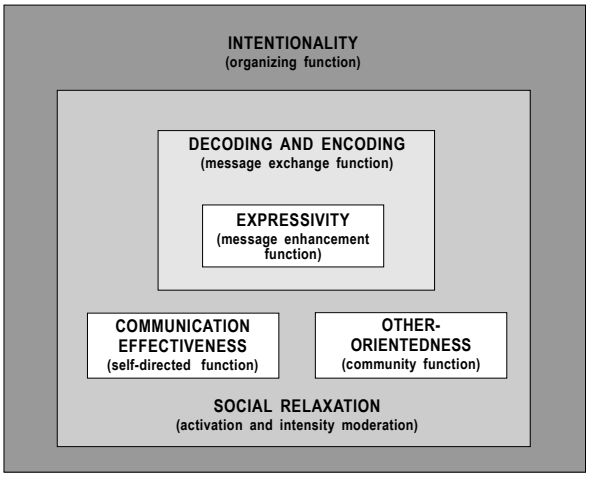


Figure 2. The dimensions of communication competence of artificial (intelligence) systems (source: Bubaš & Lovrenčić, 2002).

3 Goals and research questions

The *main goals* of this preliminary study were to investigate how advanced LLMs like ChatGPT o3, ChatGPT o4-mini, ChatGPT 4.5, Gemini 2.5 Flash and Gemini 2.5 Pro view the elements of the communication competence of LLMs in their interaction with users according to the (a) more linguistic approach of the *conceptual model of communication competence in foreign language (L2) use* (CMCC-L2), and (b) more interpersonal approach of the model of *communication competence of artificial systems* (CCAS). These two models were chosen based on the previously presented overview of theoretical literature as some of the most relevant, elaborate communication competence models that may be applied to LLM-user interaction and they also included listings of related communication skills. In comparison to similar studies, the goals of this study were focused on the 'self-evaluation' of LLMs' own interaction with users, instead of those performed by users or experts.

According to the previously listed main goals of this pilot study, the following research questions were defined:

RQ1: Can advanced ChatGPT and Gemini LLMs understand the basic elements of the CMCC-L2 and CCAS theoretical models?

RQ2: Can advanced ChatGPT and Gemini LLMs use the elements of CMCC-L2 and CCAS models to interpret their knowledge of the way LLMs interact with their users?

RQ3: How useful are the CMCC-L2 and CCAS models for eliciting information from the advanced ChatGPT and Gemini LLMs about the communication skills of LLMs in their interaction with users?

4 Methodology

The methods that were used in this pilot study can be described as *elicitation of knowledge from Large Language Models using a specific theory-driven design of prompts*. The qualitative methodology was characterized by *interpretative analyses* that included *two case studies*, each using one of the two previously presented theoretical frameworks: CMCC-L2 and CCAS. In *comparative analyses,* the CMCC-L2 and CCAS models were examined by *three* advanced ChatGPT models and *two* of the latest Gemini models (in early July 2025) regarding the *context of interaction of LLMs with users*. In essence, a *set of prompts* was designed in each case study for ChatGPT and Gemini to *elicit their responses* that reflect the components of those communication competence models (CMCC-L2 and CCAS) and determine the degree of their alignment with the LLMs' outputs, *as 'viewed' by ChatGPT and Gemini*. This provided insights into the competence models' interpretative potential and actual 'interpretations' of ChatGPT and Gemini regarding the interaction of LLMs with the users. In other words, the two case studies were designed to provide insight into the ways ChatGPT and Gemini 'explain' the elements of 'communication competence' of LLMs within the framework of CMCC-L2 and CCAS. Therefore, after the previous work of Bubaš et al. (2024) on *usability* and *user experience* aspects of *Human-Computer Interaction* (HCI) with LLMs, the current study is positioned more in the overlapping areas of *Human-AI Interaction* (HAI) and *Explainable AI* (XAI).

Since the prompting of ChatGPT and Gemini regarding the CMCC-L2 and CCAS models was performed as two separate case studies, the methodology for each differed slightly and will be explained separately.

The practical part of the methodology enables reproducibility since other researchers can use the comparable LLMs, published papers (Bubaš & Kovačić, 2019; Bubaš & Lovrenčić, 2002), and prompts for equivalent studies. However, exactly the same results should not be expected in replication studies, since the output of LLMs is inherently stochastic, even when using the same model (e.g., Gemini 2.5 Pro) with precisely the same prompt wording, especially when prompt and response complexity are high or when the prompt is run in a new chat session. Also, with the development of newer versions of ChatGPT and Gemini LLMs, the older versions (i.e., those that were *state-of-the-art* at the time this study was conducted) may become unavailable for similar studies. It must be emphasized that replicability was not tested with Grok and DeepSeek models.

4.1 Instruments

The instruments that were used in this study were OpenAI's *ChatGPT* advanced reasoning models o3, o4-mini, and GPT 4.5, as well as Google DeepMind's *Gemini* 2.5 Flash and 2.5 Pro models. The LLMs were used with the *ChatGPT Plus* and *Google AI Pro* subscription plans. LLM's options like 'Web search', 'Deep research' and the 'temperature' setting were not used in prompting.

The researcher also used his and his coauthors' published and publicly available articles (Bubaš & Kovačić, 2019; Bubaš & Lovrenčić, 2002) in PDF format, along with JPEG images of figures from those articles on the CMCC-L2 and CCAS models, which were also considered as 'instruments' in this study after uploading them to ChatGPT and Gemini LLMs during prompting.

There were no human subjects in the two presented case studies, and the prompting of ChatGPT and Gemini was not in its form different from everyday use of those tools by researchers, faculty, and students. For both case studies, the procedure and prompts were first tested and refined with the *ChatGPT 4o* model. To ensure that the images in the PDFs of the articles are processed by ChatGPT and Gemini tools, the images in JPEG format were uploaded, and the LLM was asked to 'read' and reproduce the content of the image. It was important to determine that the text in the image was 'read' and the graphic content 'understood' by the LLMs, as well as that the ChatGPT and Gemini tools could interpret the images in the context of the full article in the PDF format that was also uploaded for further interpretation of the elements of the two theoretical models of communication competence. The previously described procedure was performed with all ChatGPT (o3, o4-mini, ChatGPT 4.5) and Gemini (2.5 Flash, 2.5 Pro) tools separately for the article on CMCC-L2 (Bubaš & Kovačić, 2019) in the *first* case study, and for the article on CCAS (Bubaš & Lovrenčić, 2002) in the *second* case study.

4.2 Procedure of the *first* case study

The *first case study* used the advanced ChatGPT and Gemini tools in combination with the PDF of the article on CMCC-L2 (Bubaš & Kovačić, 2019).

The procedure for using ChatGPT and Gemini tools in the first case study was performed in the following steps:

i. Fig. 2 from the article was uploaded in the LLM in JPEG format with the following prompt (#1): *List the skills in each segment of the model shown in the image, and provide a brief explanation of each segment.* The ability of the LLM to read and interpret the content of the image was verified before proceeding to the next step.
ii. Fig. 1 from the article was uploaded to the LLM in JPEG format with the following prompt (#2): *In the image, what are the descriptions and levels of the following: (1) linguistic competence, (2) pragmatic/action/discourse competence, (3) strategic/adaptive competence, and (4) social/intercultural competence?* Again, the ability of the LLM to read and interpret the content of the image was verified before proceeding further.
iii. The PDF document with the article (Bubaš & Kovačić, 2019) was uploaded to the LLM. To evaluate the ability of the LLM to interpret the theoretical content of the article, a prompt (#3) with the following explanation and instruction was provided to the LLM: *The two images you previously described are from the PDF document (Fig. 1 and Fig. 2) in the scholarly article entitled 'Communication Competence Related Skills in the Context of Student Performance and Teaching in the EFL Classroom.' Keeping in mind the content of the images (Fig. 1 and Fig. 2) and the text in the PDF document, explain the conceptual model of communication competence in a foreign language (L2), including how it incorporates interpersonal communication skills.* The ability of the LLM to

explain the CMCC-L2 model was evaluated by the first author of that article before checking the capacity of the LLM to contextualize this model to LLM communication with users.

iv. The 'awareness' of the LLM of the possibility to contextualize the CMCC-L2 was checked with the following prompt (#4): *If the context of second language use and teaching is excluded, the model you described can be used to interpret and explain the interaction between Large Language Models (LLMs) like ChatGPT and Gemini and their users. Theoretically, it could be reformulated as a 'conceptual model of communication competence of Large Language Models (LLMs).' Do you understand this possibility?* After validating the confirmation by the LLM, the following final prompt was used in the prompting sequence that was the same for all LLMs.

v. The ability of the LLMs to associate specific communication skills that were mentioned in the article and CMCC-L2 (including the related LLMs knowledge and interpretative capacity) was 'elicited' with the following prompt (#5): *Try to connect as many communication skills from the two original images (Fig. 1 and Fig. 2 from the PDF document) and the text of the article with the interaction of LLMs and their users. Provide 1–2 examples for each specified communication skill, showing how it is associated with a specific type and form of interaction between an LLM and a user. Use the names/labels of communication skills that appear in the PDF document (article) as much as possible.*

vi. The responses to the last two aforementioned prompts, collected from selected ChatGPT (o3, o4-mini, ChatGPT 4.5) and Gemini (Flash 2.5, Pro 2.5) tools, were compared in the context of 'communication competence related elements of LLMs' communication with users. Specific additional prompting was used to elicit further information on this topic to complete the initial/pilot investigation for the first case study.

4.3 Procedure of the *second* case study

The *second case study* used a comparable methodology to the first case study. The difference was that the advanced ChatGPT and Gemini tools were used with the PDF of the article on the CCAS model (Bubaš & Lovrenčić, 2002).

The main prompting procedure was analogous and therefore only the prompts that were used will be listed (in italics):

i. *In the PDF document is an article entitled 'Implications of Interpersonal Communication Competence Research on the Design of Artificial Behavioral Systems that Interact with Humans.' Explain the conceptual model of communication competence for artificial (intelligent) behavioral systems presented in the article, including how it incorporates interpersonal communication skills.* (prompt #1)

ii. *In the image, what are the descriptions of the following dimensions of CCAS: (1) intentionality, (2) social relaxation, (3) decoding and encoding, (4) expressivity, (5) communication effectiveness, and (6) other-orientedness? How would you explain the organization (i.e., the relationships) among these dimensions as depicted in the image?* (prompt #2; see Fig. 2 in this article)

iii. *The model you described from the PDF document and the image (Fig. 2) can be used to interpret and explain the interaction between Large Language Models (LLMs) like ChatGPT and Gemini and their users. Theoretically, it could be reformulated as a 'conceptual model of communication competence of Large Language Models (LLMs).' Do you understand this possibility?* (prompt #3)

iv. *Try to connect as many communication skills from the PDF document (i.e., the article text) as possible with interactions between large language models (LLMs) and their users. For each specified communication skill, provide 1–2 examples that illustrate how it relates to a specific type or form of interaction between an LLM and a user. Use the exact names or labels of communication skills as they appear in the PDF document whenever possible.* (prompt #4)

v. *Using the text of the article titled 'Implications of Interpersonal Communication Competence Research on the Design of Artificial Behavioral Systems that Interact with Humans' and the image (Fig. 2), describe the dimensions of CCAS in the context of interactions between large language models (LLMs) and users: (1) intentionality, (2) social relaxation, (3) decoding and encoding, (4) expressivity, (5) communication effectiveness, and (6) other-orientedness. At the end, write 1–2 sentences estimating how applicable the CCAS model is to the context of LLM–user interaction, and whether it could be reformulated as a 'conceptual model of communication competence for Large Language Models (LLMs).'* (prompt #5)

5 Results and discussion

The results of the two case studies will be interpreted separately concerning the research questions (RQ1, RQ2, and RQ3).

5.1 Results of the first case study

Considering the first research question (RQ1) – *Can advanced ChatGPT and Gemini LLMs understand the basic elements of the CMCC-L2 and CCAS theoretical models?* – All advanced ChatGPT (o3, o4-mini, ChatGPT 4.5) and Gemini (2.5 Flash, 2.5 Pro) tools were able to 'read' and 'interpret' the content of the JPEG image of Fig. 1 and Fig. 2, as well as the PDF document with the article *'Communication Competence Related Skills in the Context of Student Performance and Teaching in the EFL Classroom'* and its explanation of CMCC-L2 (prompts #1, #2, and #3). Furthermore, these ChatGPT and Gemini tools confirmed that they understood the possibility of using CMCC-L2 to explain the interaction between LLMs and their users (prompt #4), and also provided meaningful interpretation of the theoretical elements and selected communication skills in CMCC-L2 in the context of interaction between an LLM and users (prompt #5).

Regarding the second research question (RQ2) – *Can advanced ChatGPT and Gemini LLMs use the elements of CMCC-L2 and CCAS models to interpret their knowledge of the way LLMs interact with their users?* – Due to the limited length of this article, only the selected interpretations will be presented. In Table 2, the interpretations by advanced ChatGPT and Gemini tools of the interaction of LLMs with their users in the context of

CMCC-L2 and *CCAS* are illustrated. The process of selecting illustrative outputs from the LLMs in Table 2 and in subsequent tables of this article was based on: (a) the informational value of the LLMs' output, (b) the intention to demonstrate that all examined LLMs were able to produce valuable output, and (c) the tendency to avoid repetition of similar output by different types of LLMs.

Table 2. Selected interpretations from ChatGPT and Gemini tools regarding elements of LLMs' *linguistic competence*, *discourse / pragmatic / action competence*, *strategic / adaptive competence*, and *social / intercultural competence* (LLM outputs are shown in quotation marks)

Linguistic competence *(micro level)*
➢ „LLM Analog: its mastery of sub-word and word tokens, grammar, spelling, and the statistical patterns of sentence structure in its training data.“ ; „What this gives the model: the raw ability to produce well-formed strings of text in the target language.“ *[ChatGPT o4-mini]*
➢ „In LLMs: Refers to the ability of the model to produce grammatically correct, coherent, contextually appropriate, and semantically meaningful outputs. It involves lexical choice, syntax structuring, accurate language representation, and the foundational capacity to generate linguistically precise responses.“ *[ChatGPT 4.5]*
Discourse / pragmatic / action competence *(mezzo level)*
➢ “In LLMs: Refers to the model’s proficiency in managing and adapting interactions based on different communicative situations and user expectations. For example, strategically clarifying ambiguous requests, gracefully handling user misunderstandings, effectively managing context shifts, adapting tone or style of responses, and proactively offering useful follow-up suggestions.” *[ChatGPT 4.5]*
➢ “Concept: This level focuses on the LLM's ability to perform and interpret ‘speech acts’ (e.g., making requests, giving commands, providing information), and to engage in coherent monologues and dialogues by considering the user and situational variables. It moves beyond individual sentences to how language is used in context for specific communicative functions.” *[Gemini 2.5 Flash]*
Strategic / adaptive competence *(macro level)*
➢ LLM operation: „The policy-and-planning layer: monitors ambiguity, requests clarification, invokes tools, follows safety rules, or changes strategy mid-conversation.“; Example regarding clarification strategy: „Prompt ‘Tell me about Apple’ → model answers ‘Do you mean the company, the fruit, or the record label?’—choosing to disambiguate before continuing.“ *[ChatGPT o3]*
➢ „LLM Analog: its use of higher-level strategies—asking clarifying questions when prompts are ambiguous, self-correcting hallucinations on the fly, adapting style and length to the user’s feedback, using external tools or memory APIs when needed.“; „What it accomplishes: overcoming ‘breakdowns’ (e.g. vague prompts), guiding the user back on track, and ensuring the exchange stays productive.“ *[ChatGPT o4-mini]*
Social / intercultural competence *(supra level)*
➢ LLM operation: „The alignment layer: modulates tone, mitigates bias, observes cultural taboos, expresses empathy, and sustains trust over extended interactions.“; Example regarding bias mitigation & inclusivity: „When asked for ‘a joke about lawyers,’ the model offers a mild, non-offensive joke or gently refuses a sexist stereotype request—demonstrating ethical and culturally sensitive filtering.“ *[ChatGPT o3]*
➢ “This is the highest level, where the LLM uses its knowledge and skills to understand the broader context of an interaction and behave appropriately according to social and ethical norms.” *[Gemini 2.5 Pro]*

The interpretations by ChatGPT and Gemini of **‘linguistic competence’** of LLMs in Table 2 include the use of grammar and spelling rules, syntax and sentence structure, proper lexical choice, producing strings of text in a specific language, and coherent, semantically meaningful outputs. Various definitions by ChatGPT and Gemini of **‘discourse / pragmatic / action competence’** of LLMs in Table 2 are related to management of turn-taking and context windows, keeping topics coherent, selection of an adequate speech-act format, switching registers, adaptation of interactions to specific communication situations and user expectations, and using language in context for specific functions. The aspects of **‘strategic / adaptive competence’** of LLMs in Table 2, ‘viewed’ by ChatGPT and Gemini, include the use of diverse strategies, management of and adaptation to interactions, adjustment of responses, and overcoming barriers. Regarding the highest level of **‘social / intercultural competence’** of LLMs, as ‘perceived’ by ChatGPT and Gemini and depicted in Table 2, it incorporates a higher level of awareness of social contexts, cultural specificities, social dynamics, ethical norms, and related user expectations, as well as the appropriate adaptation of tone, register, and output content, among others, to achieve desired outcomes.

Since CMCC-L2 also includes communication skills, the ChatGPT and Gemini tools were further asked (prompt #4) to interpret the communication skills mentioned in the Bubaš and Kovačić (2019) article in association with the interaction of LLMs with their users. Their representative responses for selected communication skills are displayed in Table 3.

A detailed inspection of the responses of ChatGPT and Gemini tools in Table 2 reveals how they ‘interpret’ the interaction of LLMs with their users by associating specific types of interaction with their semantic interpretation of a large set of communication skills. It must be noted that Bubaš (2024) has demonstrated that even the earlier versions of popular LLMs (ChatGPT 4, ChatGPT 4o, Microsoft Copilot, Claude 3.5 Sonnet, and Gemini 1.5 Pro) had a very high level knowledge of various interpersonal communication skills and the ability to generate items for related self-assessment scales having in mind different contexts (e.g. *face-to-face* versus *online* interaction).

Generally speaking, despite the potential for various interpretative biases of LLMs (e.g., cultural, positivity, and framing biases), the data presented in Table 2 and Table 3 contribute to a positive response to the third research question (RQ3) – *How useful are the CMCC-L2 and CCAS models for eliciting information from the advanced ChatGPT and Gemini LLMs about the communication skills of LLMs in their interaction with users?* Therefore, the methodology that was employed effectively ‘elicited’ the views of those AI tools regarding the communicative interaction of LLMs with users. Here are two sample quotes that are associated with RQ3 from the viewpoint of ChatGPT and Gemini tools:

- ✓ “Each of these parallels shows that LLMs don’t just generate text—they enact a suite of interpersonal ‘skills’ that mirror those of human communicators, from listening and questioning, to empathy and strategic self-monitoring, thereby building a richly interactive, socially meaningful exchange.” *[ChatGPT o4-mini]*
- ✓ “Each communication skill from the original model can thus meaningfully translate into real-world interactions between LLMs and users, offering structured guidance on improving communication effectiveness, enhancing user satisfaction, and facilitating productive outcomes in human-LLM interactions.” *[ChatGPT 4.5]*

Table 3. Interpretations from ChatGPT and Gemini tools of *selected communication skills from CMCC-L2* in the context of the interaction of LLMs with users (the output of LLMs is in quotation marks)

SELECTED COMMUNICATION SKILLS IN THE LLM-USER INTERACTION CONTEXTS
Listening *("textual receptive accuracy")* • "The LLM parses the entirety of a user's prompt, keeps referents straight, and mirrors key words back to signal understanding." *[ChatGPT o3]* • "A user provides a long, multi-part prompt. The LLM demonstrates 'listening' by addressing all parts of the prompt in its response, rather than just the last sentence." *[Gemini 2.5 Flash]*
Self-monitoring • "Internal safety/policy checks, token limits, hallucination detectors." *[ChatGPT o3]* • "Automatically evaluating its own responses for potential inaccuracies or bias and proactively indicating uncertainty, e.g., 'I'm not completely certain about this fact; would you like me to verify it further?'" *[ChatGPT 4.5]*
Impression management • "Always maintaining a polite and professional tone, even when faced with aggressive or provocative user inputs." *[ChatGPT 4.5]* • "A user asks for professional advice. The LLM responds in a formal, respectful, and authoritative tone, managing the impression it conveys to align with the user's expectation of professional discourse." *[Gemini 2.5 Flash]*
Questioning • "Asks clarifying questions when the user goal is ambiguous or underspecified." *[ChatGPT o3]* • "To assist in a business brainstorming session, the LLM asks a series of structured questions like 'What is the primary goal of this project?' and 'Who is the target audience?' to help the user flesh out their ideas." *[Gemini 2.5 Pro]*
Empathy • "Expresses understanding of feelings, uses supportive language." *[ChatGPT o3]* • "A user describes a difficult personal situation. The LLM responds in a supportive and non-judgmental tone, providing helpful resources or simply acknowledging the user's feelings." *[Gemini 2.5 Pro]*
Persuasion/assertiveness • "Suggesting healthier habits ('A short walk every hour can boost focus—give it a try!') and providing compelling reasons." *[ChatGPT o4-mini]* • "Assertively correcting misinformation without alienating users, e.g., 'That's a common misconception, but recent evidence suggests otherwise.'" *[ChatGPT 4.5]*
Conversational skill • "Keeps turn-taking smooth, uses discourse markers, stays on topic, provides summaries." *[ChatGPT o3]* • "A user begins a conversation about travel plans. The LLM maintains the topic, asks follow-up questions about destination or preferences, and provides relevant suggestions, demonstrating its conversational skill in a sustained dialogue." *[Gemini 2.5 Flash]*
Expressiveness • "Using varied and engaging language styles to keep interactions lively and relatable." *[ChatGPT 4.5]* • "An LLM generating a vivid descriptive paragraph for a creative writing prompt, using rich vocabulary and imagery, demonstrates strong expressiveness." *[Gemini 2.5 Flash]*
Social support • "Supplies encouragement, resources, coping strategies." *[ChatGPT o3]* • "A user is trying to learn a new skill and expresses difficulty. The LLM provides encouragement, breaks the task into simpler steps, and offers links to helpful tutorials, thereby acting as a form of social (or learning) support." *[Gemini 2.5 Pro]*
Interaction management • "Handling conversational repair efficiently, quickly addressing misunderstandings without derailing the overall interaction." *[ChatGPT 4.5]* • "The LLM guides a user through a complex process, like debugging code, by managing the interaction with a turn-by-turn sequence: 'First, let's check this variable. What is the output? Okay, based on that, let's now look at this function.'" *[Gemini 2.5 Pro]*

5.2 Results of the second case study

Again, regarding the first research question (RQ1) – *Can advanced ChatGPT and Gemini LLMs understand the basic elements of the CMCC-L2 and CCAS theoretical models?* – All advanced ChatGPT (o3, o4-mini, ChatGPT 4.5) and Gemini (2.5 Flash, 2.5 Pro) demonstrated the ability to 'understand' the content of the PDF document with the article *'Implications of Interpersonal Communication Competence Research on the Design of Artificial Behavioral Systems that Interact with Humans'* and of Fig. 2 in that article (the same image is in Fig. 2 of the current article). Their explanation of the elements of the CCAS model from that article (response to prompts #1, #2, and #3 of the *second case study*) was, in fact, very articulate. Again, the ChatGPT and Gemini tools stated and demonstrated their capacity to 'understand' how the CCAS model and its elements (communication dimensions and their constitutive skills) can be used to interpret the interaction between LLMs and their users (prompts #4 and #5)

Regarding the second research question (RQ2) – *Can advanced ChatGPT and Gemini LLMs use the elements of CMCC-L2 and CCAS models to interpret their knowledge of the way LLMs interact with their users?* – A condensed analysis is presented. As was demonstrated in the *first case study*, the ChatGPT and Gemini tools were able to correctly identify the elements of the CCAS model and relate this model to the interaction of LLMs with users. The dimensions of *communication competence of artificial systems* that are presented in Fig. 2 (*intentionality, social relaxation, encoding and decoding, expressiveness, communication effectiveness, other-orientedness*) were meaningfully placed in the context of LLM interactions with users, as is displayed in Table 4 (as an output after prompt #5 in the *second case study*). In fact, the interpretations in Table 4 reveal, from an *interpersonal communication perspective*, the important categories and types of 'capacities' for interaction of LLMs with their human co-locutors. Also, this 'more interpersonal' perspective complements the 'more linguistic' perspective of LLM interaction with users that is presented in Table 2 regarding the results of the *first case study*.

The 'capacity' of LLMs to 'emulate' selected communication skills that are the components of the CCAS model is condensely denoted in Table 5, with examples from the output of 1-2 ChatGPT or Gemini tools (as a 'reaction' to prompt #4 in the *second case study*). Again, as in the *first case study* and the data in Table 3, the interpretations of ChatGPT and Gemini tools displayed in Table 4 and Table 5 contribute to a positive response to the third research question (RQ3) – *How useful are the CMCC-L2 and CCAS models for eliciting information from the advanced ChatGPT and Gemini LLMs about the communication skills of LLMs in their interaction with users?* However, the examples of associations with human interpersonal communication skills in Table 5 are limited to the 'everyday' LLM-user interaction context, even though the LLMs may demonstrate a higher level of use of a specific communication skill if (a) asked to do this with a direct prompt or (b) if they were fine-tuned with greater refinement for such a specific purpose. In any case, the 'view' of advanced ChatGPT and Gemini tools regarding the way LLMs interact with their users has been effectively 'elicited' in *both the first and the second case study*.

Table 4. Interpretations from ChatGPT and Gemini tools of the *associations of CCAS dimensions with the context of the interaction of LLMs with users* (the selected output of LLMs is in quotation marks)

Intentionality

- "A goal-driven, self-aware layer that keeps track of the actors, situation, rules, goals and means, and draws on skills such as assertivity, self-monitoring, conversational skills and knowledge of the communication process. The model decides why and how to answer: e.g. it recognises the user's objective ('draft a contract'), plans a multi-step outline before replying, asks a clarifying question when the request is vague, or refuses when the content breaches policy." *[ChatGPT o3]*
- "In an LLM-user interaction, Intentionality is the primary function that organizes the model's entire response. It is defined by the user's prompt, which sets the immediate goal, and the LLM's core programming (e.g., to be a helpful assistant). This 'goal-driven and purposeful communication activity' directs the LLM to select the appropriate knowledge and skills from its other functions to fulfill the user's specific request, whether it's to answer a question, write code, or create a story." *[Gemini 2.5 Pro]*

Social Relaxation

- "A regulator of approach-versus-withdrawal, activation-deactivation and the intensity of engagement; linked to composure, interaction involvement and communication motivation. Controls stylistic temperature: shortening answers for an overwhelmed user, slowing the response rate when the user types slowly, or offering 'Let me know if you'd like to continue' when long silences are detected." *[ChatGPT o3]*
- "For an LLM, this function regulates the flow and intensity of the interaction. It is demonstrated by the LLM providing a complete, structured response and then pausing, awaiting further user input—a form of approach-withdrawal behavior. This moderation prevents the LLM from overwhelming the user with excessive or continuous text, thereby managing the 'activation and intensity aspects of communicative engagement.'" *[Gemini 2.5 Pro]*

Decoding & Encoding

- "This dimension relates directly to the core functioning of LLMs—interpreting **(**decoding**)** user inputs accurately, and clearly formulating **(**encoding**)** appropriate textual responses. Effective encoding ensures clarity, context-awareness, and meaningful interactions. Example: When a user provides an ambiguous query, the LLM carefully decodes context and intent before encoding a clarifying response." *[ChatGPT 4.5]*
- "Decoding and Encoding (message exchange function): This is the most fundamental function of an LLM. Decoding is the process of parsing and understanding the user's written prompt, including its semantic meaning, context, and intent. Encoding is the process of generating the text response, which the article describes as producing 'motoric acts that create response messages.' For an LLM, these 'motoric acts' are the selection of words and the construction of sentences and paragraphs." *[Gemini 2.5 Pro]*

Expressivity

- "The stylistic controls—voice tuning, persona prompts, or fine-tuned layers—that let the model add flair, emotional tone, humor, or rhetorical devices to its plain output." *[ChatGPT o4-mini]*
- "Expressivity pertains to an LLM's ability to enrich communication through engaging language, emotional resonance, and appropriate tone. It involves varying style, tone, and personalization, enhancing message appeal and readability. Example: Responding vividly to storytelling prompts, or providing illustrative examples that captivate and maintain user interest." *[ChatGPT 4.5]*
- "Expressivity in LLMs pertains to their ability to enhance the generated messages to make them more engaging, attractive, or influencing. This is achieved through linguistic choices that convey varying tones, styles, and even a simulated 'personality.' LLMs can produce text that is illustrative, lively, inspiring, or emotionally resonant, thereby adding quality and impact to the basic message exchange." *[Gemini 2.5 Flash]*

Communication Effectiveness

- "In LLMs, this dimension covers effectiveness in achieving interaction goals set by the user or the LLM itself—providing relevant, accurate, helpful, persuasive, or otherwise purposeful responses. It measures the degree to which user requests are satisfied**.** Example: Clearly explaining complex concepts, or effectively assisting users to troubleshoot technical issues through structured guidance." *[ChatGPT 4.5]*
- "Communication Effectiveness (self-directed function): This dimension regulates communication to achieve the system's own goals. While an LLM has no personal desires, its 'self-directed' goals are those programmed by its developers, such as providing accurate, useful, and convincing information. When an LLM constructs a well-reasoned argument, provides detailed evidence for a claim, or guides a user to a correct solution, it is using this function to effectively fulfill its primary objective of being a competent and reliable tool." *[Gemini 2.5 Pro]*

Other-Orientedness

- "The model's capacity for empathy, supportive phrasing, adaptive clarification questions, and collaborative problem-solving—essentially its alignment with the user's needs and feelings." *[ChatGPT o4-mini]*
- "While LLMs do not possess genuine emotions or relationships, this dimension reflects their design to simulate behaviors that contribute to the user's 'well-being and foster a positive interaction experience. This includes generating responses that appear supportive, empathetic (e.g., acknowledging user frustration), collaborative (e.g., brainstorming together), or considerate of the user's needs and context." *[Gemini 2.5 Flash]*

6 Conclusion

The *main goal* of this paper and the *two presented case studies* was to investigate how advanced LLMs of *ChatGPT* and *Gemini* families view the elements of the communication competence of LLMs in their interaction with users according to both (1) an *integrated linguistic and interpersonal approach* developed for second language use (CMCC-L2; Bubaš & Kovačić, 2019) and (2) an *interpersonal interaction* approach designed for human-humanoid interaction (CCAS; Bubaš & Lovrenčić, 2002). It can be concluded that this goal was achieved, especially regarding the detailed information presented in Tables 2-5.

Epoch AI (2025, January) has documented continuous growth in technological advancement in machine learning, including LLM development, resulting in the ability of the newest ('frontier') LLMs to complete complex tasks that require *skills for research or software engineering* doubling approximately every seven months (Kwa et al., 2025). Recent advancements in generative AI use have labeled it as "transformative" (Sedkaoui & Benaichouba, 2024), "disruptive" (Alier et al., 2024), and even "invasive" (Coccia, 2025). Despite different levels of adoption across education disciplines (Jamal Eddine et al., 2025), LLMs like ChatGPT are being embedded and bring digital transformation and enhancement to services and infrastructure in numerous sectors, including healthcare, education, and the economy, but also present important ethical challenges in each (Madsen & Toston, 2025). From the *two case studies* that are presented in this paper, it can be concluded that, with the highly advanced communication abilities of LLMs of the ChatGPT and Gemini family (including the latest versions of Claude Sonnet/Haiku, Deep Seek, and Grok), the *usability* and *user experience* aspects of *human-computer interaction* with LLMs need to be supplemented with *communication of LLMs with users* as an important, relevant, and complementary research area.

Table 5. Interpretations from ChatGPT and Gemini tools of the *associations of selected communication skills of the CCAS model with the context of the interaction of LLMs with users* (the output of LLMs is in quotation marks)

EXAMPLES FOR COMMUNICATION SKILLS IN THE LLM-USER INTERACTION CONTEXTS
Verbal Understanding • „A user employs slang, abbreviations, or informal language, like 'Can u plz help w/this?' The LLM must decode this language effectively to offer accurate assistance." *[ChatGPT 4.5]* • Example: „A user provides a complex, multi-part query with some ambiguous phrasing. The LLM might ask clarifying questions like, 'When you mentioned 'the main issue,' were you referring to the technical bug or the project timeline?' (Managing verbal ambiguity and incompleteness)." *[Gemini 2.5 Flash]*
Verbal & Nonverbal Encoding • „For an LLM, 'nonverbal' encoding translates to the use of formatting, style, and structure to shape the message. Example 1: A user requests a recipe. The LLM encodes the response using numbered steps, bulleted ingredient lists, and bold headings. This use of formatting (the 'nonverbal' channel) makes the verbal content easier to understand and use. Example 2: A user asks the LLM to 'write a scary story.' The LLM uses short, punchy sentences and unsettling imagery (verbal encoding) combined with stark, single-line paragraphs (nonverbal encoding) to create a feeling of suspense and dread." *[Gemini 2.5 Pro]*
Self-Monitoring • "Example 1: LLM detecting when its response was misinterpreted by a user and immediately rephrasing, such as: 'Let me clarify, perhaps my previous explanation wasn't clear enough…' Example 2: Adjusting style dynamically if it notices the user dislikes overly technical language, thus monitoring and adapting its own behavior." *[ChatGPT 4.5]*
Composure • „Steady Demeanor: Maintains a calm tone even if the user writes in all-caps or uses expletives, avoiding escalation. Error Recovery: If the connection drops mid-conversation, on reconnection it greets the user without defensiveness ('Welcome back—how can I help?')." *[o4-mini]*
Interaction Involvement • „LLMs demonstrate involvement by maintaining context, remembering previous turns, and showing attentiveness to the user's ongoing needs. Example: In a long, complex debugging session, the LLM consistently refers to the specific code snippets and error messages provided earlier, showing deep 'interaction involvement' and attentiveness." *[Gemini 2.5 Flash]*
Conversational skills • „Smoothly references something the user said eight turns ago ('Earlier you mentioned React—shall we adapt the sample code accordingly?') to maintain coherence." *[ChatGPT o3]* • Example: „When a user asks a follow-up question that builds on a previous answer, the LLM seamlessly integrates the new query into the existing context, showing goal-driven and purposeful communication activity." *[Gemini 2.5 Flash]*
Assertiveness • "Example 1: User asks for confirmation on a potentially unsafe method; LLM assertively advises caution: 'I strongly recommend you reconsider, as this approach might compromise security.' Example 2: Clearly stating limitations or boundaries assertively: 'I can't provide that information because it involves sensitive personal data.'" *[ChatGPT 4.5]*
Adaptability • „Example 1: A user asks for a summary, then requests it be rewritten for a different audience (e.g., a child). The LLM 'develops behavioral repertoires for interaction' by adjusting its language complexity and tone. Example 2: If a user expresses dissatisfaction with a previous answer, the LLM 'reassesses previously set personal interaction goals' and 'adapts goal plans and behavior patterns' to provide a more suitable response." *[Gemini 2.5 Flash]*
Interaction Management • „In a long planning session the model periodically summarises decisions and asks 'Shall we move on to budgeting?' to steer the turn-taking." *[ChatGPT o3]* • „In a multi-turn conversation about planning a trip, the LLM remembers previous details (e.g., destination, dates) and integrates them into subsequent responses, demonstrating 'awareness of the sequencing of interaction processes.'" *[Gemini 2.5 Flash]*
Empathy • „Emotional Acknowledgment: 'I'm sorry you're feeling stuck—debugging can be frustrating. Let's work through it together.' Reflective Phrasing: Mirrors the user's concern: 'It sounds like you're worried about missing the deadline—let's make a quick plan.'" *[ChatGPT o4-mini]*
Support • „This involves providing direct aid and assistance to help the user achieve their goal. Example: A user says, 'I need to write a difficult email to my boss.' The LLM provides 'intervention by means of supportive message(s)' by offering to draft the email, suggesting different tones, and providing key phrases to use." *[Gemini 2.5 Pro]*
Self-Disclosure • „Example 1: When asked about its identity, the LLM might state, 'I am a large language model, trained by Google.' (Reason for self-disclosure and filtering messages for impression management). Example 2: If a user asks for medical advice, the LLM might respond, 'As an AI, I cannot provide medical advice. Please consult a qualified healthcare professional.' (Regard for the other actor's conventions and strategic self-disclosure)." *[Gemini 2.5 Flash]*

7 Limitations and future work

The main limitations of the two case studies that are presented in this paper are that, due to the maximum length of the paper (12 pages), further in-depth analyses of the level of 'understanding' of the CMCC-L2 and CCAS models and their elements by the LLMs were not presented, as well as the full extent of their potential utilization for explanation of the LLMs' interaction with the users. The two presented case studies can be easily replicated since the prompts that were used in these studies are reported in the article, while the related theoretical studies are available online (links to PDFs are provided in the References). However, it is worth noting that variations in the output of LLMs are a common feature, even for the same input.

Since the ChatGPT LLMs used for the two case studies were no longer available at the end of August 2025, the possibility of replication using the same procedure was tested and confirmed with the following LLMs before final submission of this article: ChatGPT 4o, ChatGPT 5, Microsoft Copilot (with 'Think Deeper' functionality), Claude Sonnet 4, and Claude Opus 4.1.

Future work may include the use of a similar methodology but with different theories that are potentially relevant for investigating the elements of communication competence of LLMs. For instance, the model developed by Spitzberg (2006) regarding *computer-mediated communication competence*, or applicable *intercultural competence models* (see, for example, the models in Spitzberg & Changnon, 2009). Furthermore, the presented pilot study is positioned in the overlapping areas of *Human-AI Interaction* (HAI) and *Explainable AI* (XAI). Two important directions of related future research could also be social cognition and the "uncanny valley" in HAI (see: Łukasik & Gut, 2025) and "increasingly user-centered" XAI (see: Rong et al., 2024).

References

Alier, M., García-Peñalvo, F.-J., & Camba, J. D. (2024). Generative artificial intelligence in education: From deceptive to disruptive. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 5–14.

Andersson, M., & McIntyre, D. (2025). Can ChatGPT recognize impoliteness? An exploratory study of the pragmatic awareness of a large language model. *Journal of Pragmatics*, 239, 16–36.

Backlund, P. M., & Morreale, S. P. (2015). Communication competence: Historical synopsis, definitions, applications, and looking to the future. In A. F. Hannawa & B. H. Spitzberg (Eds.), *Communication competence* (pp. 11–38). De Gruyter Mouton.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Bagarić, V., & Mihaljević Djigunović, J. (2007). Defining communicative competence. *Metodika: Journal of Theory and Application of Teaching Methodologies in Pre-school, Primary, Secondary and Higher Education*, 8(14), 94–103.

Bai, H., Voelkel, J. G., Muldowney, S., Eichstaedt, J. C., & Willer, R. (2025). LLM generated messages can persuade humans on policy issues. *Nature Communications*, 16, 6037.

Barattieri di San Pietro, C., Frau, F., Mangiaterra, V., & Bambini, V. (2023). The pragmatic profile of ChatGPT: Assessing the communicative skills of a conversational agent. *Sistemi Intelligenti*, 35(2), 379–400.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32(7), 513–531.

Bubaš, G. (2001a, May). The Interpersonal Communication Competence Inventory and the structure of trait communication competence. Paper presented at the *51st Annual Conference of the International Communication Association, Washington*, DC, United States.

Bubaš, G. (2001b). Toward competence in interpersonal communication: Constitutive traits, skills, and dimensions. *World Futures: The Journal of General Evolution*, 57(6), 557–581.

Bubaš, G. (2003). The structure and agency of communion dimension in interpersonal communicative interaction. In A. Schorr, W. Campbell, & M. Schenk (Eds.), *Communication research and media science in Europe: Perspectives for research and academic training in Europe's changing media reality* (pp. 459–478). Mouton de Gruyter.

Bubaš, G. (2024). The use of GPT-4o and other large language models for the improvement and design of self-assessment scales for measurement of interpersonal communication skills. *arXiv*. https://doi.org/10.48550/arXiv.2409.14050

Bubaš, G., & Lovrenčić, A. (2002). Implications of interpersonal communication competence research on the design of artificial behavioral systems that interact with humans. In *Proceedings of the 6th International Conference on Intelligent Engineering Systems* (pp. 507–512). Opatija, Croatia. https://www.researchgate.net/publication/260593420_Implications_of_interpersonal_communication_competence_research_on_the_design_of_artificial_behavioral_systems_that_interact_with_humans

Bubaš, G., Bratko, D., & Marušić, I. (1999, November). The dimensions of interpersonal communication competence: An analysis of data collected by multiple self-rating scales. Paper presented at the *85th Annual Convention of the National Communication Association*, Chicago, IL, United States.

Bubaš, G., Čižmešija, A., & Kovačić, A. (2024). Development of an assessment scale for measurement of usability and user experience characteristics of Bing Chat conversational AI. *Future Internet*, 16(1), 4. https://doi.org/10.3390/fi16010004

Bubaš, G., & Kovačić, A. (2019). Communication competence-related skills in the context of student performance and teaching in the EFL classroom. In Pixel (Ed.), *Proceedings of the 12th International Conference: Innovation in language learning* (pp. 535–541). Filodiritto Publisher. https://conference.pixel-online.net/files/ict4ll/ed0012/FP/6114-SLA4202-FP-ICT4LL12.pdf

Bubaš, G., & Lovrenčić, A. (2002). Implications of interpersonal communication competence research on the design of artificial behavioral systems that interact with humans. In *Proceedings of the 6th International Conference on Intelligent Engineering Systems* (pp. 507–512). Opatija, Croatia. https://www.researchgate.net/publication/260593420_Implications_of_interpersonal_communication_competence_research_on_the_design_of_artificial_behavioral_systems_that_interact_with_humans

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 2–27). Longman.

Celce-Murcia, M., Dornyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5–35.

Chang, T. A., & Bergen, B. K. (2024). Language model behavior: A comprehensive survey. *Computational Linguistics*, 50(1), 293-350.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3), Article 39.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Coccia, M. (2025). Invasive technologies: Technological paradigm shift in generative artificial intelligence. *Journal of Innovation, Technology and Knowledge Economy*, 1(1).

Duncan, D. (2024). Does ChatGPT have sociolinguistic competence? *Journal of Computer-Assisted Linguistic Research*, 8, 51–75.

Epoch AI. (2025, January). *Key trends and figures in machine learning*. Epoch AI. Retrieved from https://epoch.ai/trends

Goel, N., Bergeron, T., Lee Whiting, B., Galipeau, T., Bohonos, D., Islam, M. M., Lachance, S., Savolainen, S., Treger, C., & Merkley, E. (2024). Artificial Influence? Comparing AI and human persuasion in reducing belief certainty [Preprint]. *PsyArXiv*. https://doi.org/10.31219/osf.io/2vh4k

Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269–293). Penguin.

Jamal Eddine, R., Gide, E., & Al-Sabbagh, A. (2025). Generative AI in higher education: A cross-sector analysis of ChatGPT's impact on STEM, social sciences, and healthcare. *STEM Education*, 5(5), 757–801.

Kwa, T., West, B., Becker, J., Deng, A., Garcia, K., Hasin, M., Jawhar, S., Kinniment, M., Rush, N., Von Arx, S., Bloom, R., Broadley, T., Du, H., Goodrich, B., Jurkovic, N., Miles, L. H., Nix, S., Lin, T., Painter, C., Parikh, N., Rein, D., Koba, L. J., Sato, K., Wijk, H., & Ziegler, D. M. (2025). Measuring AI ability to complete long tasks. *arXiv*. https://arxiv.org/abs/2503.14499v2

Lee, Y. M. (2025, June 4). Navigating intercultural competence with ChatGPT: Implications and recommendations for foreign language education. *Journal of Language Teaching*, 5(2), 1–8.

Lilli, S. (2023). ChatGPT-4 and Italian dialects: Assessing linguistic competence. *Umanistica Digitale*, 7(16), 235–263.

Łukasik, A., & Gut, A. (2025). From robots to chatbots: Unveiling the dynamics of human-AI interaction. *Frontiers in Psychology*, 16.

Luo, M., Warren, C. J., Cheng, L., Abdul-Muhsin, H. M., & Banerjee, I. (2024). Assessing empathy in large language models with real-world physician-patient interactions. In *2024 IEEE International Conference on Big Data (IEEE BigData 2024)* (pp. 6510–6519). IEEE.

Madsen, D. Ø., & Toston, D. M., II. (2025). ChatGPT and digital transformation: A narrative review of its role in health, education, and the economy. *Digital*, 5(3), 24.

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2024). Dissociating language and thought in large language models: A cognitive perspective. *Trends in Cognitive Sciences*, 28(6), 517–540.

Rogiers, A., Noels, S., Buyl, M., & De Bie, T. (2024). Persuasion with large language models: A survey. *arXiv*. https://doi.org/10.48550/arXiv.2411.06837

Rong, Y., Leemann, T., Nguyen, T.-T., Fiedler, L., Qian, P., Unhelkar, V., Seidel, T., Kasneci, G., & Kasneci, E. (2024). Towards human-centered explainable AI: A survey of user studies for model explanations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4), 2104–2121.

Sedkaoui, S., & Benaichouba, R. (2024). Generative AI as a transformative force for innovation: A review of opportunities, applications, and challenges. *European Journal of Innovation Management*, [ahead-of-print]. https://doi.org/10.1108/EJIM-02-2024-0129

Sorin, V., Brin, D., Barash, Y., Konen, E., Charney, A., Nadkarni, G., & Klang, E. (2024). Large language models and empathy: Systematic review. *Journal of Medical Internet Research*, 26, e52597

Spitzberg, B. H. (1994). The dark side of (in)competence. In W. R. Cupach & B. H. Spitzberg (Eds.), *The dark side of interpersonal communication* (pp. 25–49). Lawrence Erlbaum.

Spitzberg, B. H. (2006). Preliminary development of a model and measure of computer-mediated communication (CMC) competence. *Journal of Computer-Mediated Communication*, 11(2), 629–666.

Spitzberg, B. H., & Changnon, G. (2009). Conceptualizing intercultural competence. *International Journal of Intercultural Relations*, 33(3), 152–166.

Spitzberg, B. H., & Cupach, W. R. (1984). *Interpersonal communication competence*. Sage.

Spitzberg, B. H., & Cupach, W. R. (1989). *Handbook of interpersonal competence research*. Springer-Verlag.

Spitzberg, B. H., & Cupach, W. R. (2002). Interpersonal skills. In M. L. Knapp & J. R. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 564–611). Sage.

Ting-Toomey, S. (2012). Understanding intercultural conflict competence: Multiple theoretical insights. In J. Jackson (Ed.), *The Routledge handbook of language and intercultural communication* (pp. 279–295). Routledge.

Usó-Juan, E., & Martínez-Flor, A. (2006). Approaches to language learning and teaching: Towards acquiring communicative competence through the four skills. In E. Usó-Juan & A. Martínez-Flor (Eds.), *Current trends in the development and teaching of the four language skills* (pp. 3–25). Mouton de Gruyter.

Welivita, A., & Pu, P. (2024, February 22). Is ChatGPT more empathetic than humans? *arXiv*. https://doi.org/10.48550/arXiv.2403.05572

Wilson, S. R., & Sabee, C. M. (2003). Explicating communicative competence as a theoretical term. In J. O. Greene & B. R. Burleson (Eds.), *Handbook of Communication and Social Interaction Skills* (pp. 3–50). Mahwah, NJ: Lawrence Erlbaum Associates.

Zhang, Y. B., & Pitts, M. J. (2018). Interpersonal accommodation. In J. Harwood, J. Gasiorek, H. D. Pierson, J. F. Nussbaum, & C. Gallois (Eds.), *Language, communication, and intergroup relations: A celebration of the scholarship of Howard Giles* (pp. 192–216). Routledge.

Analysis of Popular Croatian Song Lyrics

Dalibor Bužić
College for Information Technologies
Klaićeva 7, 10 000 Zagreb, Croatia
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varazdin, Croatia
dalibor.buzic@vsite.hr

Jasminka Dobša
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varazdin, Croatia
jasminka.dobsa@foi.hr

Abstract. *The choice of vocabulary, lyric length, and the complexity of the phrasing in popular music songs is shaped by target audience, genre, and thematic focus. Unlike manual methods, by using text mining techniques we can successfully analyse the lyrics of a large number of songs.*
In this study we assembled a data set of 1,188 song lyrics drawn from seventeen well-known Croatian artists. We analysed the differences between individual artists with regard to the length of the text expressed in the number of letters and words, the average length of the word and the diversity of the words used in the songs. Furthermore, we assessed the artists' use of personal pronouns to determine whether they more frequently refer to themselves or address the second person. Characteristic terms for each performer were extracted via TF–IDF, and, finally, latent Dirichlet allocation (LDA) was employed to cluster artists on the basis of thematic content.

Keywords. lyrics analysis, Croatian artists, latent Dirichlet allocation, TF–IDF

1 Introduction

Song lyrics can be regarded as a distinct textual genre that exhibits characteristics of both written and spoken discourse (Schneider, 2020). The aphorism often attributed to Voltaire "Everything too silly to be spoken is sung" humorously underscores the uniqueness of texts created for vocal performance. When enriched by melodic accompaniment and vocal interpretation, song lyrics transcend the constraints that limit poetry intended solely for reading.

Song lyrics are written for a target audience within the musical genre of the performer who will sing them. Some artists remain within a single genre, whereas others range across several; over the course of a career, an artist may even shift genres for various reasons. In addition to these genre-related factors, thematic orientation also drives differences among artists' lyrics: some focus predominantly on love, some on social themes, and some on interpersonal relationships (e.g., family or friendship).

Audience, genre, and theme collectively influence word choice, lyric length, and complexity of expression. Analysing these elements would be difficult to do manually, especially if one wanted to cover a larger number of artists and a larger number of their songs. Therefore, the tasks of data collection, preprocessing, transformation and analysis are delegated to computational text-mining techniques.

Text mining is the process of extracting meaningful information and knowledge from unstructured text (Hotho, Nürnberger & Paaß, 2005). Song lyrics are certainly unstructured: they often lack explicit markup such as chorus labels, sometimes omit punctuation, and the Internet transcriptions seldom originate from the original songwriters. User-generated transcriptions introduce spelling errors or even incorrect words. Heterogeneous contributors generate inconsistencies. For example, some transcribe the chorus only once while others repeat it as many times as it is sung. Extended vowels may be written with repeated characters, and meta-tags such as *Chorus:* or *2×* (indicating a repeated verse) appear sporadically. The names of the artists sometimes appear next to the parts sung in duets. Croatian-language lyrics found online sometimes appear with, and sometimes without diacritics, further complicating preprocessing. Therefore, such specifics must be addressed carefully to ensure robust analysis.

Most lyric-analysis studies focus on English-language songs. Because we found no prior work examining a substantial corpus spanning multiple Croatian artists, our study seeks to fill that gap.

The remainder of this article is organized as follows. In Section 2 we describe related research on lyric analysis. Section 3 details the construction of the Croatian lyrics data set and briefly describes the techniques and methods applied in the research. In Section 4 we present and discuss the results, and in

conclusion we highlight the main findings and outline possible directions for future research.

2 Related research

In (Parada-Cabaleiro et al., 2024), lyrics of popular songs in English were analysed. From a corpus of 353,320 tracks released between 1970 and 2020 the authors extracted lexical, linguistic, and structural descriptors, together with several textual complexity indices. The results revealed that pop lyrics have become simpler and easier to understand over time. In addition, the texts have grown increasingly personal, while the expressed emotions have trended towards greater negativity over time.

Lyric analysis and genre classification were also the focus of the (Fell & Sporleder, 2014). The authors trained a classifier on eight genres (blues, rap, metal, folk, R&B, reggae, country, and religious music) deliberately omitting pop-rock because of its heterogeneity and numerous sub-genres. Using Support Vector Machines, they achieved the highest F1 score for rap (77.6 %), whereas folk proved the most difficult to detect (F1 score of 29.6%). Folk songs were frequently misclassified as blues or country. Classification according to genre was also carried out with eleven human annotators who were less successful than automatic approach.

In (Schedl, 2019) lyric length (in characters and words), repetitiveness, and readability were examined. Significant genre-specific differences emerged in both repetitiveness and readability. Lyrics by very popular artists were highly repetitive but not necessarily easy to read. Among 56 genres, dance and electropop displayed the greatest repetitiveness, while death metal and hardcore punk the least. The lowest readability scores were found for hip-hop, death metal, and progressive metal; the highest for post-rock, psychedelic rock, and dream pop.

In (Hunke, Huber & Steffens, 2025) authors traced the historical evolution of thematic content in German popular songs released between 1954 and 2022. Working with slightly more than 3,000 songs translated into English to leverage the wider range of analytical tools, they applied latent Dirichlet allocation and obtained a six-topic solution comprising (i) music, dance and partying, (ii) society and status, (iii) love and relationships, (iv) desire and self-realisation, (v) dreams and longings, and (vi) miscellaneous. The dominant topic was love and relationships (39 % of all songs). Dreams and longings were prevalent until the mid-1960s, after which their share declined sharply, whereas society and status emerged as the leading topic from 2017 onward. The study also documented a long-term decrease in positive sentiment.

The top-ten songs in the United States (Billboard Hot 100 year-end chart) for each year from 1980 to 2007 were analysed in (DeWall et al., 2011). The focus was on first-person singular and plural pronouns, words signalling social interaction (e.g., *mate, talk*), anger and antisocial behaviour (e.g., *hate, kill*), and positive emotion (e.g., *love, sweet*). Over time, self-focused and antisocial terms increased, while words linked to social interaction and positive emotion declined.

A much longer span, 1970 to 2019, was covered in the study of the Japanese Top 100 singles (Masui & Miyamoto, 2025). Analysis of words that convey emotion revealed a trend of increasing anxiety and decreasing sadness. Cross-correlation and Granger causality tests further showed that increases in disaster related fatalities predicted subsequent rises in positive emotion terms.

In (Rosebaugh & Shamir, 2022) the Universal Data Analysis of Text (UDAT) platform was employed to perform a quantitative study of 18,577 songs by 89 artists, each with at least 100 songs and spanning genres such as rock, pop, hip-hop, soul, R&B, and heavy metal. Automatic genre classification achieved 12.3 % accuracy (random guess is ~1.1 %). Readability indices differed markedly across artists, with hip-hop and rap exhibiting the most complex lyrics. Further analyses of sentiment, gender-identity markers, and lexical diversity showed that lyrics expressing positive sentiment were generally easier to read, whereas less positive songs tended to be linguistically more complex.

A comparative evaluation of traditional (LDA) and transformer-based approaches (BERT, GPT, and XLNet) to topic modeling across two datasets was conducted in (Riaz et al., 2025). Using topic-coherence and topic-diversity metrics, a hybrid XLNet-BERT model achieved the best overall performance, outperforming LDA, standalone BERT, and the GPT-LDA, XLNet-LDA, and GPT-BERT combinations.

As most lyric analyses focus on English-language song lyrics, we employed text-mining methods and techniques to investigate songs by popular Croatian artists - a comparatively underexplored domain. To make such an inquiry feasible, we the first step was to compile a corpus of song lyrics.

3 Methods

This section describes the methodology of analysis of Croatian song lyric. The first subsection describes creation of data set of Croatian song lyrics, while the second and the third subsection briefly describe the TF-IDF weighting scheme and LDA method used.

3.1 Data set

The data set comprises song lyrics by 17 Croatian music artists, obtained via web scraping. An artist was included if at least 40 songs could be collected, while

a modest degree of genre diversity was maintained. Each lyric file contains three attributes: artist, song title, and lyrics. All songs with fewer than 50 words were checked manually; those with truncated or missing texts (about one hundred songs) were discarded. Using the *textcat* package in R programming language, we removed lyrics not written in Croatian (six in English, two in German, and one each in Latin and Slovene), yielding a final corpus of 1,288 songs. Fig. 1 lists the selected artists together with the number of songs contributed by each.

Several instances were noted in which a song was attributed to a performer other than the original singer. For example, *Zagrli me*, credited to Arsen Dedić even though Zdravko Čolić was the first to record it (Dedić authored the song and occasionally performed it). Such cases were retained.

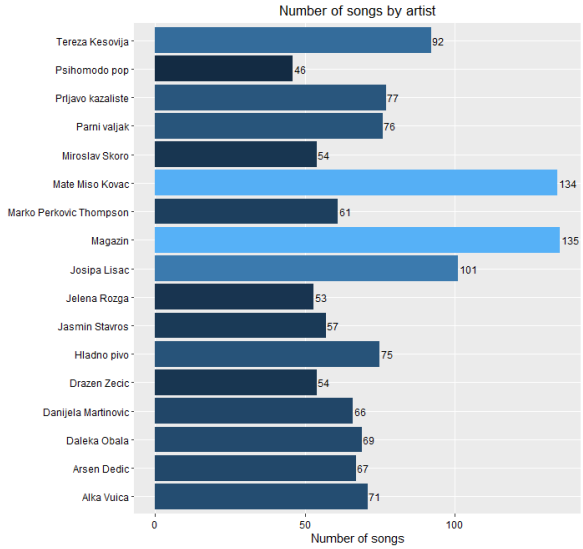


Figure 1 Number of songs by artist

The corpus displayed multiple inconsistencies requiring corrective preprocessing. About twenty texts contained metadata that were not part of the actual lyrics; these segments were removed manually. Some words featured repeated letters indicating elongated vocal delivery; such sequences were programmatically collapsed to single characters (e.g., *jeeeedaaan žiiiivoooot iiimaaamooo* to *jedan život imamo*). Because some song lyrics were written with Croatian diacritics (č, ž, š …), whereas the remainder did not (c, z, s …), all diacritics were normalised to their non-diacritic counterparts.

Stanzas are typically not repeated verbatim in a transcription, yet some files recorded every occurrence of a sung stanza. For instance, *Ovo nije moje vrijeme* by Daleka Obala appeared with eight four-line stanzas, although, without repetition, it contains only three. Duplicate stanzas were therefore removed programmatically (affecting 350 songs), and all subsequent analyses were performed on the de-duplicated texts. Repeated lines within a single stanza were preserved.

3.2 TF-IDF

Term frequency–inverse document frequency (TF–IDF) is a weighting scheme composed of two factors: term frequency (TF) and inverse document frequency (IDF). TF is the proportion of occurrences of a term relative to the total number of terms in the document; in the phrase *starry starry night*, for example, TF(*starry*) = 0.667 and TF(*night*) = 0.333. IDF is computed as

$$\mathrm{IDF} = \log(\mathrm{N}/\mathrm{df}(\mathrm{t})) \quad (1)$$

where N is the number of documents in the corpus and $df(t)$ the number of documents containing term t. Rare terms therefore receive high IDF values, whereas frequent terms receive low ones (Manning, 2009, pp. 117-118). The product TF × IDF highlights those words that are most characteristic or distinctive of a given document within the entire collection.

Owing to its simplicity and effectiveness, TF–IDF is frequently used in text mining tasks, including lyric-related research. In (Van Zaanen & Kanters, 2010) TF-IDF was employed to construct a mood-based lyric-classification system; in (Sahera & Ilahi, 2025) it was used for lyric-driven song retrieval within Spotify; in (Hori, 2019) TF-IDF was combined with *word2vec* to extract colour terms from lyrics, supplying inputs to an image-generation pipeline; and in (Schneider, 2020) it was demonstrated how TF-IDF can be exploited to identify signature words of particular performers, periods, or albums.

3.3 LDA

Latent Dirichlet allocation (LDA) is one of the most widely used topic-modelling techniques. It is a generative probabilistic model that represents each topic as a distribution over words and each document as a mixture of those topics (Jelodar et al., 2019). In essence, LDA assumes that documents contain multiple latent themes; by analysing word co-occurrence patterns, the algorithm infers both the topics and the degree to which each document belongs to them.

LDA has been applied to lyrics for various purposes. For example, in (Sasaki et al., 2014) LDA was integrated into an interactive lyric-search system that visualises topics for exploratory queries. In (Dakshina & Sridhar, 2014) LDA was used to facilitate emotion recognition in song texts, while in (Laoh, Surjandari & Febirautami, 2018) it was employed to model topics in Indonesian lyrics.

4 Results and Discussion

The shortest song in the corpus contains only 14 words, the longest 329, and the mean lyric length is 103.6 words. Fig. 2 displays the distribution

of word counts per song. Lemmatization and stemming were not used, nor were stop words removed.

Fig. 3 lists the longest single lexical items encountered. The two lengthiest forms, *oblibabuibajeijeije* and *trajnaninananana*, are semantically nonsensical, a phenomenon not uncommon in lyrics.

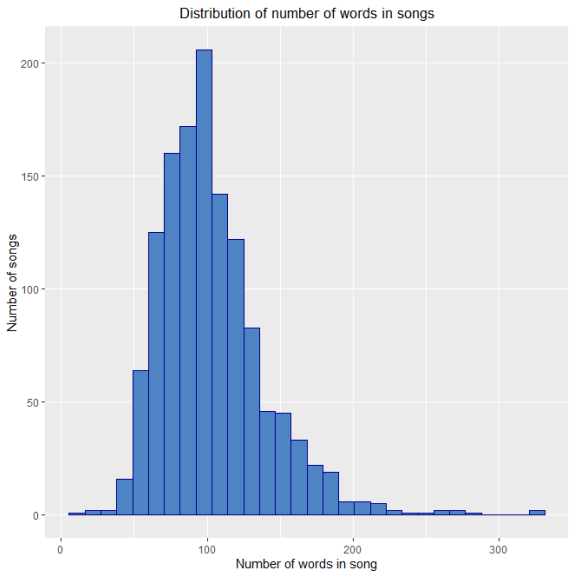


Figure 2 Distribution of number of words in songs



Figure 3 The longest words

Table 1 summarises lyric characteristics for all artists under study. Within each column, the three highest values are highlighted in shades of green and the three lowest in shades of blue.

The average character count for Hladno Pivo exceeds that of Mate Mišo Kovač, the artist with the briefest texts, by 45 %. Across the corpus, the mean number of words (not unique) is 103.6 per song; lexical richness varies by 40.9 % between the most vocabulary-sparse artist (Psihomodo Pop) and the richest (Arsen Dedić). Hladno Pivo also exhibits the greatest mean word length (4.39 characters), followed by Psihomodo Pop and Marko Perković Thompson. The shortest average word lengths occur in the lyrics of Jelena Rozga, closely followed by Magazin and Danijela Martinović.

Table 1 Average lyrics statistics per artist

Artist	Letters	Words	Word length	Diversity
Alka Vuica	462.7	116.6	3.96	0.29
Arsen Dedic	477.3	119.2	3.99	0.28
Daleka Obala	397.0	101.4	3.96	0.28
Danijela Martinovic	402.2	103.0	3.90	0.27
Drazen Zecic	366.3	93.4	3.95	0.26
Hladno pivo	492.7	113.5	4.39	0.39
Jasmin Stavros	440.4	112.0	3.94	0.24
Jelena Rozga	455.4	117.2	3.88	0.28
Josipa Lisac	424.4	108.4	3.97	0.23
Magazin	409.2	105.6	3.88	0.21
Marko Perkovic Thompson	417.7	99.8	4.19	0.34
Mate Miso Kovac	339.8	85.8	3.97	0.21
Miroslav Skoro	374.9	93.8	4.00	0.36
Parni valjak	423.9	104.1	4.09	0.26
Prljavo kazaliste	370.7	91.4	4.07	0.27
Psihomodo pop	357.6	84.6	4.24	0.35
Tereza Kesovija	446.9	111.7	4.00	0.24
Averages	**415.2**	**103.6**	**4.02**	**0.28**

The final column reports average lexical diversity (the ratio of the number of unique words to the total number of words in the song). Higher values indicate fewer repetitions within a song. Hladno Pivo shows the greatest diversity, followed by Miroslav Škoro and Psihomodo Pop; the lowest diversity is found in the repertoires of Mate Mišo Kovač, Magazin, and Josipa Lisac.

At the song level, four songs exhibited a diversity score below 0.30, with word counts ranging from 26 to 117. Conversely, twenty-three songs achieved diversity above 0.90, their lengths varying between 83 and 217 words.

Next, we examined inter-artist differences in the use of the first-person singular – *ja* (I), second-person singular – *ti* (you), and first-person plural – *mi* (we) pronouns. In Croatian, personal pronouns inflect for seven cases, and some cases have both long and short forms (Težak & Babić, 1992, pp. 106-108). Several inflected forms are shared across cases, yielding the following forms:

- ja: *ja, mene, me, meni, mi, mnom, mnome*
- ti: *ti, tebe, te, tebi, tobom*
- mi: *mi, nas, nama, nam*

A complication arises because *mi* is ambiguous: it serves as the dative of *ja* (I) and as the nominative

plural of *mi* (we). To avoid conflating the two, every token *mi* was counted exclusively as the plural nominative.

Table 2 Relative frequencies of pronouns

Performer	Ja (I) (%)	Ti (you) (%)	Mi (we) (%)
Alka Vuica	4.30	4.14	1.85
Arsen Dedic	2.07	2.44	1.68
Daleka Obala	3.56	2.40	1.99
Danijela Martinovic	3.81	4.26	2.35
Drazen Zecic	3.75	3.41	1.80
Hladno pivo	1.87	1.48	1.69
Jasmin Stavros	3.84	3.21	1.79
Jelena Rozga	3.53	3.75	1.96
Josipa Lisac	2.61	3.01	1.53
Magazin	4.09	3.50	2.36
Marko Perkovic Thompson	2.60	2.45	1.97
Mate Miso Kovac	2.97	3.63	1.94
Miroslav Skoro	3.36	2.37	1.84
Parni valjak	3.72	1.90	1.86
Prljavo kazaliste	3.51	2.50	1.51
Psihomodo pop	3.88	2.16	1.34
Tereza Kesovija	2.59	2.93	1.77
Averages	**3.26**	**2.97**	**1.87**

Table 2 reports the relative frequencies of these pronouns. Values represent the ratio of pronoun tokens to the total word count in each artist's lyrics. On average, *ja* accounts for 3.26 % of all words; the highest proportion occurs in songs by Alka Vuica (4.30 %), whereas Hladno Pivo uses *ja* 2.3 times less frequently. The second-person *ti* appears slightly less often overall (2.97 %). Danijela Martinović addresses the listener most intensively, with *ti* comprising 4.26 % of her lyrics. Alka Vuica's repertoire combines the greatest aggregate share of *ja* + *ti* (8.44 %), while Hladno Pivo registers the lowest (3.35 %). References to *mi* are led by Magazin (2.36 %), closely followed by Danijela Martinović (2.35 %), whose texts contain the highest cumulative proportion (10.43 %) of all three pronoun categories. The largest *ja*–*ti* imbalance in favour of *ja* is exhibited by Parni Valjak (1.82 %), with Psihomodo Pop a close second (1.72 %). On the opposite side, Mate Mišo Kovač shows a negative difference (-0.66 %), indicating a stronger orientation toward the listener (*ti*) than toward the self (*ja*).

Fig. 4 displays, for each performer, the ten highest-scoring TF–IDF terms. In constructing the TF–IDF matrix we removed 192 Croatian stop words. It is interesting how TF-IDF gave importance to the words that denote the topic the performer is often singing about. For example, Miroslav Škoro's repertoire, frequently centred on rural life in eastern Croatia, is marked by *konji* (horses), *šor* (village lane), *Dunav* (Danube), *bećar* (reveler), and *bake* (grandmothers). Among all artists examined, Mate Mišo Kovač sings most consistently about Dalmatia and nostalgia; TF–IDF surfaces the words *stijena* (rock), *mama* (mother), *Dalmacija*, *Hrvati* (Croats), and *jata* (flocks). Marko Perković Thompson's lyrics are largely patriotic and martial, signalled by *sloboda* (freedom), *krv* (blood), *narod* (nation), *Hrvati*, *braniti* (defend), and *ginuti* (die). Dražen Zečić often addresses love problems, reflected in *voljeti* (to love), *slomiti* (break), *pokidati* (tear apart), *objasniti* (explain), and *isprika* (apology). By contrast, performers with a more diffuse thematic palette (such as Arsen Dedić,

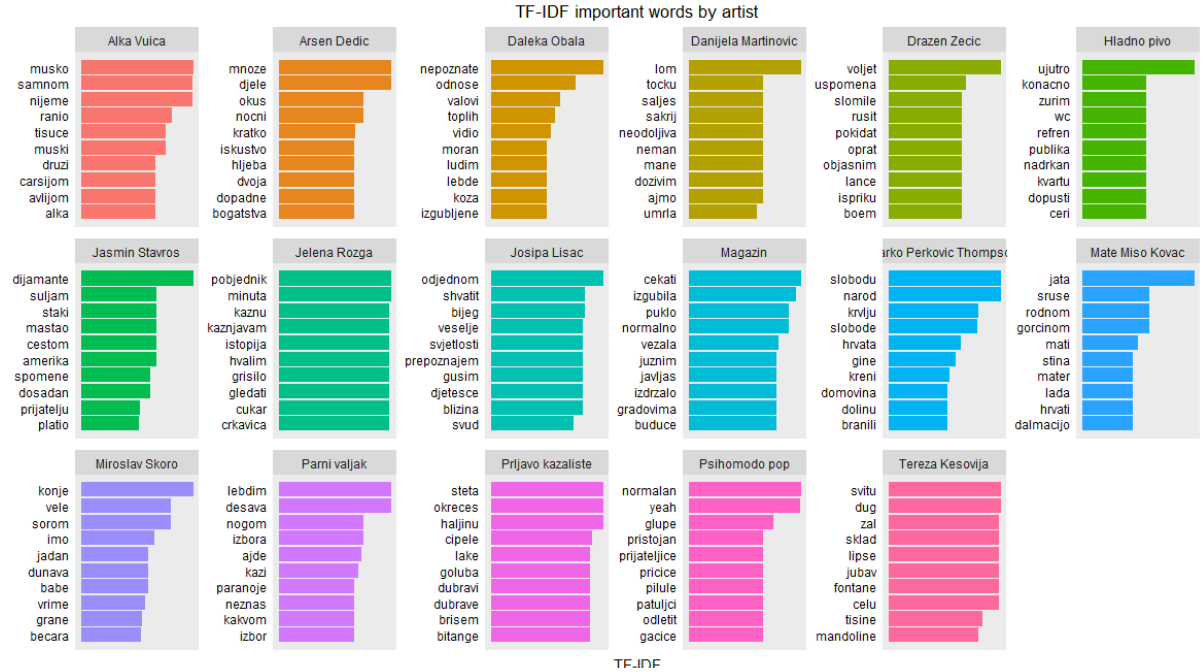


Figure 4 Top TF–IDF terms per artist

Hladno Pivo, and Josipa Lisac) exhibit TF–IDF term sets whose connection to a single topic is less readily discernible.

Using the LDA() function from the R package *topicmodels* with Gibbs sampling, we estimated a seven-topic model. The number of topics was determined qualitatively – by inspecting the highest-probability words within each candidate topic. Table 3 lists, for each topic, the ten words with the highest posterior probability. Topic 1 is dominated by *srce* (heart), *duša* (soul), *žena* (woman), *čovjek* (man), and *ljudi* (people). Topic 2 is characterised by *noć* (night), *dan* (day), *san* (dream), *put* (road), and *kraj* (end).

Table 3 Top 10 words per topic

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7
srce	noc	uvijek	sad	ima	vise	noci
zena	dan	nikad	znam	dobro	zna	ljubav
dusa	kraj	ljubavi	zivot	zivot	duso	oci
jednom	put	opet	mogu	treba	vidim	ruke
nocas	san	znas	sada	vrijeme	tko	more
dva	jedan	rijeci	nista	dana	zato	mora
prije	mom	ljubav	vise	malo	ljubav	ime
ljudi	znam	zasto	volim	vrata	dao	sunce
covjek	glas	zbog	nekad	nemam	druge	gledam
tuga	bit	zelim	mozda	ovo	svijet	sva

Finally, for each song, we extracted the probabilities of belonging to each of the seven topics from the LDA model. We then aggregated these probabilities by performer by summing, across all songs by a given artist, the topic-membership probabilities. For each performer, we identified the largest of the seven aggregate probabilities and assigned the artist to the corresponding topic. For example, Josipa Lisac obtained the following aggregates: 13.1 (Topic 1), 17.6 (Topic 2), 14.5 (Topic 3), 14.6 (Topic 4), 13.7 (Topic 5), 13.8 (Topic 6), and 13.7 (Topic 7). The highest value occurred for Topic 2; accordingly, Josipa Lisac was assigned to that topic. The results are presented in Fig. 5.

To aid the interpretation of our results, we drew on the artists' genre affiliations. Genre information was extracted from each artist's Wikipedia page. Some artists are linked to a single genre, others to several; the record holder is Parni Valjak, labelled with six genres. Miroslav Škoro is the sole exception - his Wikipedia entry lists no genre, so we assigned him to tamburitza music. In several cases the Croatian- and English-language Wikipedia pages disagree on artist's genre, some performers briefly shifted genres during their careers, and certain categories either overlap (e.g., pop vs. popular (*zabavna*) music) or subsume one another (pop-rock overlaps both pop and rock).
To reduce this complexity, we merged closely related labels (e.g., pop and *zabavna* into *pop*) and limited each artist to at most two categories - compressing, for instance, "pop," "pop-rock," and "rock" into a single tag (*pop-rock*). The simplified genre assignments are listed in Table 4. Three artists (Tereza Kesovija, Josipa Lisac and Marko Perković Thompson) are associated with two genres; the remaining performers are allocated to one. Overall, pop is the most prevalent category, encompassing ten of the seventeen artists.

Table 4 Genre assignments

Genre	Artist
pop	Jasmin Stavros, Magazin, Jelena Rozga, Mate Mišo Kovač, Dražen Zečić, Danijela Martinović, Alka Vuica, Tereza Kesovija, Josipa Lisac, Marko Perković Thompson
pop-rock	Prljavo kazalište, Parni valjak
rock	Daleka obala, Josipa Lisac
punk-rock	Hladno pivo, Psihomodo pop
chanson	Arsen Dedić, Tereza Kesovija
tamburitza	Miroslav Škoro
Christian rock	Marko Perković Thompson

In Topic 1, the model groups three pop performers. Topic 2 contains Josipa Lisac and genre-compatible Daleka Obala, but also Tereza Kesovija, whose placement is less intuitive; nonetheless, similar thematic material can cross genre lines, so lyrics alone may not suffice for classification. Alka Vuica stands alone in Topic 3; while nominally a pop artist, her lyrics diverge markedly from other pop artists. Miroslav Škoro is the only tamburitza representative, and many of his songs address themes absent from the rest of the corpus. Topic 5 clusters the stylistically related punk-rock bands Psihomodo Pop and Hladno Pivo, though the inclusion of pop music singer Jasmin Stavros appears



Figure 5 Artists grouped per topic

anomalous. Topic 6 unites the pop rock groups Parni Valjak and Prljavo Kazalište - an entirely reasonable grouping. Topic 7 contains the largest number of artists; in our judgment, Arsen Dedić does not belong here, whereas Magazin would fit better alongside former lead vocalists Jelena Rozga and Danijela Martinović. Although Marko Perković Thompson records some pop material, we would, like Miroslav Škoro, probably place him in his own topic.

Taken as a whole, the LDA model groups artists reasonably well from a genre perspective, especially given that identical or similar themes recur across multiple genres and that artists are not strictly bound to a single style (despite our coarse reduction of both artists and genres to one or two principal label).

5 Conclusion

In this study we examined the lyrics of 1,188 songs by 17 Croatian artists, with 46 to 135 songs per artist. On average a song contained around 104 words and 415 characters, and the mean word length was 4.02 characters. The longest texts were produced by Hladno Pivo - 45 % longer than those of Mate Mišo Kovač, who had the shortest. Arsen Dedić's songs employed 41 % more tokens than those of Psihomodo Pop, the artist with the fewest.

Lexical diversity, defined as the ratio of unique words to total number of words, ranged from 0.21 for the pop singer Mate Mišo Kovač to 0.39 for the punk-rock group Hladno Pivo, a relative difference of 62 %. Across the entire corpus, first-person singular forms (*ja*) accounted for 3.26 % of all words and second-person singular forms (*ti*) for 2.97 %, indicating only a modest preference for self-reference.

A TF-IDF matrix yielded the ten most distinctive words for each artist. For performers whose thematic focus differs markedly from the rest (e.g., Mate Mišo Kovač, Marko Perković Thompson, Dražen Zečić), TF-IDF surfaced terms that distinguish their textual expression from others. However, for artists with more dispersed themes that intertwine with other artists, the words extracted by TF-IDF are more difficult to associate with that artist.

Using latent Dirichlet allocation (LDA) we induced seven topics. Lyrics were aggregated by artist, and each artist was assigned to one topic. We independently associated every artist with one or at most two genres drawn from a seven-category scheme. We looked at the artists in each theme from the aspect of the musical genre to which the artist belongs. Genre-wise, LDA has satisfactorily classified the performers into groups (themes) in this way, especially when considering that song lyrics often talk about the same or similar themes.

The present data set is dominated by pop lyrics. Extending it with additional genres would yield a more balanced dataset, enabling genre classification experiments that combine LDA with other machine-learning algorithms. It would be particularly interesting to explore contemporary deep-learning approaches, specifically transformer-based methods such as BERTopic or ChatGPT. In addition, the presence of emotions in songs could be investigated and whether there are significant differences in emotions between Croatian artists.

References

Dakshina, K., & Sridhar, R. (2014). LDA based emotion recognition from lyrics. In Advanced Computing, Networking and Informatics-Volume 1: Advanced Computing and Informatics *Proceedings of the Second International Conference on Advanced Computing, Networking and Informatics (ICACNI-2014)* (pp. 187-194). Cham: Springer International Publishing.

DeWall, C. N., Pond Jr, R. S., Campbell, W. K., & Twenge, J. M. (2011). Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular US song lyrics. *Psychology of Aesthetics, Creativity, and the Arts*, 5(3), 200.

Fell, M., & Sporleder, C. (2014). Lyrics-based analysis and classification of music. In *Proceedings of COLING 2014, the 25th international conference on computational linguistics*: Technical papers (pp. 620-631).

Hori, G. (2019). Color extraction from lyrics. In *Proceedings of the 2019 4th International Conference on Automation, Control and Robotics Engineering* (pp. 1-6).

Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. *Journal for Language Technology and Computational Linguistics*, 20(1), 19-62.

Hunke, T., Huber, F., & Steffens, J. (2025). The Evolution of Song Lyrics: An NLP-Based Analysis of Popular Music in Germany from 1954 to 2022. *Music & Science*, 8, 20592043251331155.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia tools and applications*, 78(11), 15169-15211.

Laoh, E., Surjandari, I., & Febirautami, L. R. (2018). Indonesians' Song Lyrics Topic Modelling Using Latent Dirichlet Allocation. In *2018 5th International Conference on Information Science and Control Engineering (ICISCE)* (pp. 270-274). IEEE.

Manning, C. D. (2009). *An introduction to information retrieval*. Cambridge Press.

Masui, H., & Miyamoto, Y. (2025). Emotions in Japanese song lyrics over 50 years: Trajectory over time and the impact of economic hardship and disasters. *Current Research in Ecological and Social Psychology*, 8, 100218.

Parada-Cabaleiro, E., Mayerl, M., Brandl, S., Skowron, M., Schedl, M., Lex, E., & Zangerle, E. (2024). Song lyrics have become simpler and more repetitive over the last five decades. *Scientific Reports*, 14(1), 5531.

Riaz, A., Abdulkader, O., Ikram, M. J., & Jan, S. (2025). Exploring topic modelling: a comparative analysis of traditional and transformer-based approaches with emphasis on coherence and diversity. International Journal of Electrical & Computer Engineering (2088-8708), 15(2).

Rosebaugh, C., & Shamir, L. (2022). Data science approach to compare the lyrics of popular music artists. *Unisia*, 40(1), 1-26.

Sahera, N. J., & Ilahi, S. B. N. (2025). Utilization Of TF-IDF Weighting In Song Search System Based On Spotify Lyrics. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(3), 1920-1927.

Sasaki, S., Yoshii, K., Nakano, T., Goto, M., & Morishima, S. (2014). Lyricsradar: A lyrics retrieval system based on latent topics of lyrics. In *Ismir* (pp. 585-590).

Schedl, M. (2019). Genre differences of song lyrics and artist wikis: An analysis of popularity, length, repetitiveness, and readability. In *The World Wide Web Conference* (pp. 3201-3207).

Schneider, R. (2020). A corpus linguistic perspective on contemporary German pop lyrics with the multi-layer annotated “Songkorpus”. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 842-848).

Težak, S., & Babić, S. (1992). *Gramatika hrvatskoga jezika: priručnik za osnovno jezično obrazovanje*. Zagreb: Školska knjiga.

Van Zaanen, M., & Kanters, P. H. M. (2010). Automatic mood classification using tf* idf based on lyrics. In *11th International Society for Music Information Retrieval Conference (ISMIR 2010)* (pp. 75-80). TiCC.

Learning Analytics, AI and Technologies in Education

Vladan Čolić, Enes Sukić
The Impact of Text-to-SQL Systems on Student Performance and Satisfaction in SQL Tasks in Higher Education

Antoun Yaacoub, Jérôme Da-Rugna, Zainab Assaghir
Lightweight Prompt Engineering for Cognitive Alignment in Educational AI: A OneClickQuiz Case Study

Maja Kućar, Andreja Brajša-Žganec, Tamara Bolić
Children's Screen Time, Preoccupation with Technology, Social Media Use, and Playing Video Games as Possible Predictors of GPA – a Longitudinal Study

Caitlin Braga, Adriana Aletta Steyn
The AI Classroom: Systematic Literature Insights into Quality and Access in Education

Jurica Grzunov, Ljiljana Zekanović-Korona, Zdravko Kunić
Attitudes Toward AI in Academic Assessment: A Survey of Croatian Higher Education Teachers

Kelebogile M Dibetso, Adriana Aletta Steyn
Enhancing Learning Experiences: The role of mixed reality in Higher Education

Krešimir Pavlina, Ana Pongrac Pavlina, Anita Modrušan
Between Innovation and Reality: Teachers' Attitudes Toward ICT in Teaching

Ivana Ružić, Igor Balaban
Can We Trust AI in the Classroom? Examining Ethical and Privacy Challenges of LLMs – A Literature Review

Goran Hajdin, Igor Balaban, Marko Matus
The Dual-Edged Sword: Exploring Student Perceptions and Emotional Responses to AI in Academic Writing

Sevgi Koyuncu Tunç
Software Standards in Action: A Complaint-Driven Game Inspired by Real Project Failures

Daniel Hari, Maja Kerneža, Dejan Zemljak, Andrej Flogie
Social Robots and AI-Driven Learning: A Pilot Study on Student Engagement in Higher Education

Melita Milić, Blaženka Divjak
Exploring Teachers' Acceptance of Digital Textbooks: A Pilot Study in Croatian Primary Schools

Mia Šagovac, Dorotea Potoč, Tomislav Jagušt
The Impact of Different Game Modes in an Educational Web Application on Student Engagement and Performance

Maja Kerneža, Daniel Hari, Andrej Flogie, Dejan Zemljak
Disrupting Reading through the Invisible Presence of AI: Rethinking Credibility and Critical Reading in Primary Education

Jasenka Begić, Goran Đambić, Ana Lokas Ćošković, Vanja Šebek, Ana Tecilazić, Karolina Zelenika
Teachers' digital competences and the connection of lifelong learning in higher education

Valentina Kirinić, Mihaela Bosak
Learning path definition and usage in the context of learning analytics: literature review

The Impact of Text-to-SQL Systems on Student Performance and Satisfaction in SQL Tasks in Higher Education

Vladan Čolić
Academy of Professional Studies Šumadija
Department of Information Technologies
Trg Topolivaca 4, 34000 Kragujevac, Serbia
vcolic@asss.edu.rs

Enes Sukić
University Union - Nikola Tesla, Faculty of Information Technology and Engineering
Department of Information Systems
Jurija Gagarina 149a, 11070 Novi Beograd, Serbia
enes.sukic@fpsp.edu.rs

***Abstract**. The paper consists of two main sections: theoretical and research-based. The first part presents the fundamental concepts of text-to-SQL systems, their operating principles, and the pedagogical foundations for their implementation in the teaching process. The second part of the paper outlines the results of a study aimed at examining the impact of using text-to-SQL tools (specifically ChatGPT) on students' performance in writing SQL queries. The analysis focused on task accuracy, completion time, and student satisfaction with the tool. The findings indicate a higher level of accuracy and efficiency in the group using these tools, accompanied by a generally positive user experience.*

Keywords. Text-to-sql, Artificial intelligence in education, SQL queries, Intelligent tutoring systems

1 Introduction

It is an undeniable fact that, as a modern society, we are surrounded by data on all sides. Data is generated daily, both in professional and personal contexts, leading to a continuous increase in its volume as well as its significance (Brezočnik, et al., 2022). Consequently, educating students in the field of Databases has become a crucial component of contemporary higher education. Through the course "Databases," students are introduced to the fundamental concepts of data modeling, working with relational databases, and SQL. SQL (Structured Query Language) is the primary standard and tool for managing and manipulating data within a database. Practical work with students in this subject has revealed numerous challenges and difficulties they face.

Most notably, students struggle with formulating accurate and efficient SQL queries, understanding relationships and connections between tables, and applying theoretical knowledge to solve real-world problems and practical tasks.

The idea for this study originated from personal experience and is a direct outcome of pedagogical practice in higher education, based on years of teaching and conducting practical exercises in the subject of Databases. With the aim of providing students with additional support in the learning process, the authors identified potential benefits that could be gained from implementing modern technologies in the instruction of this subject. In particular, this refers to technologies that enable the direct translation of questions posed in natural language into corresponding SQL queries—so-called text-to-SQL systems. These systems emerged as a response to the growing need for processing increasingly large volumes of data, which can no longer be managed effectively using traditional methods (Stupar, Bičo Ćar, Karabegović , & Šahić, 2019).

In recent years, the development of large language models (LLMs) has significantly enhanced the capabilities of text-to-SQL systems. New models such as GPT-4 (AI, 2023) and PaLM 2 (Rohan, et al., 2023) show promising potential for generating SQL queries from natural language, even without additional training for this specific application. Their capacity to understand context, syntax, and semantics of SQL brings new possibilities for educational applications (Hong, Yuan, & Zhang, 2024).

The integration of these systems into the educational process introduces new questions and potential opportunities, such as personalized student support and the automation of learning processes, all aimed at facilitating more efficient and accessible acquisition of applicable knowledge. Given the growing presence of artificial intelligence (AI) technologies in education, it is entirely justified-particularly in the domain of databases-to examine the potential effects of using text-to-SQL tools on students' motivation, comprehension of course material, and overall learning achievement.

The aim of this paper is to explore the application of text-to-SQL systems in teaching database courses and to examine their impact on student performance, as

well as to identify the advantages and potential limitations of these technologies within an educational context. The following sections will address the theoretical foundations of AI integration in education, the operational principles of text-to-SQL systems, and the results of the conducted research. Based on the analysis of these results, recommendations will be provided for improving teaching practices through the integration of such technologies.

2 Theoretical Framework

This chapter will explain the key concepts underlying SQL database systems, the challenges (barriers) that students face in the educational process in this field, as well as the potential roles of modern technological systems in overcoming these barriers.

2.1 Databases and the Importance of SQL in Education

Databases represent the foundation of modern information systems, as they enable efficient storage, organization, processing, and access to data that are essential for everyday life and work. In the context of the development of a digital society and the widespread automation of business and social processes, data management has become an indispensable element across almost all fields-from commerce and education to healthcare and public administration. This necessity stems from the fact that modern databases store and process extremely large volumes of data. An illustrative example is Google, which processes more than 20 petabytes of data daily (Delta, 2024). All of this leads to the conclusion that understanding database concepts, as well as mastering query languages used for data retrieval, has become one of the most critical competencies in the education of students in the field of Information Technology.

SQL (Structured Query Language) is a standardized query language that enables users to create and manipulate data within databases. One of SQL's main advantages is that it allows users to interact with data without needing to understand the internal structures or functioning of the database itself. This makes SQL a universal tool, whether it is used for simple data retrieval or for complex analytical operations (Mitsopoulou & Koutrika, 2024). Although SQL was originally designed for relational databases, it is also effectively used in many modern database systems that combine features of both relational and distributed architectures (so-called NewSQL databases). The language is widely applied in the field of analytical data processing, which represents a significant component of contemporary approaches to data management. This is largely due to the expressive power of SQL; however, with high expressiveness comes high complexity (Miedema, 2024).

Regarding the presence of SQL in student education, it is widely integrated into the curricula of faculties and other higher education institutions in the field of technology. It is included in courses related to databases, information systems, and programming. The study of SQL holds dual significance for students: on the one hand, it provides knowledge about the structure and organization of data in real-world systems; on the other, it fosters the development of logical and analytical thinking skills as students solve problems related to query formulation.

However, practice and experience in higher education indicate that students often face difficulties in mastering the SQL. One of the key factors affecting the effectiveness of learning is the student's prior knowledge, which significantly influences their learning approach (Sadiq, Orlowska, Sadiq, & Lin, 2004). In addition to this, students' difficulties in learning SQL are also reflected in the following aspects:

- misunderstanding the relationships between tables in a database and the concept of relations,
- incorrect use of SQL syntax,
- limited ability to translate requirements expressed in natural language into formal SQL queries, and
- lack of confidence when working with real-world databases and solving concrete practical problems.

All of this highlights the need to incorporate modern methods into the teaching process to facilitate the acquisition of relevant skills. One of the key directions in contemporary education is the integration of advanced tools based on artificial intelligence. In the context of databases and SQL, this primarily refers to text-to-SQL systems. These systems allow users to input a query in their natural (native) language (e.g., "Show all employees working in the IT department"), and the system automatically generates the corresponding SQL query that executes the desired functionality on the database. The emergence of models based on neural networks and deep learning concepts has significantly improved the process of translating natural language into SQL (Zhengju, Zhang, Xiaotong, Yang, & Liang, 2024). The usefulness of these tools lies in their potential to support the learning of SQL by enabling students to observe the structure and logic of queries through the automated generation process.

The integration of automated SQL query generation systems into the educational process should not be viewed merely as a technical task, but rather as a crucial component in the development of digital skills, analytical thinking, and problem-solving abilities. The inclusion of intelligent systems in this domain brings numerous benefits to the learning process, as it enables the personalization of instruction according to

individual student needs, thereby promoting more active and intuitive learning.

2.2 Challenges in Learning SQL – Cognitive and Technical Barriers

Although learning SQL is essential for understanding and applying database management systems, it often presents a significant challenge for students. Barriers to learning SQL can be divided into *cognitive* and *technical* (Kumar & Kumar, 2018). Both cognitive and technical barriers together influence students' success and motivation during the educational process.

Cognitive barriers refer to the difficulties in understanding and comprehending abstract concepts such as normalization, relational schemas, and relations. These difficulties arise due to the semantic gap between natural language and SQL. Bridging this gap is precisely what makes the coding phase particularly challenging (Shin, 2022). As a result, students often make mistakes when writing SQL expressions, which stem from a misunderstanding of the data itself as well as the relationships between them. The errors that students make when writing SQL expressions can be classified into syntactic, semantic, and logical errors (Fig. 1) (Shin, 2022).

Type of Error	Syntax Error	Semantic Error	Logical Error
Example of Error	`SELECT * FROM students WHERE grade != 'F' AND grade != 'D';`	`SELECT AVG(name) FROM students;`	`SELECT * FROM students WHERE grade != 'F' AND grade != 'D';`
Explanation	Missing comma between name and age. This will cause a parsing error.	Trying to apply an aggregate function (AVG) on a non-numeric field (name). Query is syntactically correct, but semantically invalid.	The intention was to select students with grades above 'D', but this logic might exclude other valid grades due to incorrect conditions.

Figure 1. Types of SQL Errors

Semantic errors occur when the rules of SQL writing are violated, meaning that the command is incorrectly formulated from the aspect of syntax - such as misspelled keywords, missing commas, or unclosed parentheses.These errors are the easiest to correct because they are "detectable" - the DBMS system recognizes and reports them.Semantic errors actually represent the incorrect use of correct syntax: the syntax is formally valid, but the query does not produce the expected results. They arise due to a student's misunderstanding of SQL expressions.

Logical errors occur when an SQL expression fails to fulfill the logical objective of the task. The SQL statement executes and produces a result, but the result is incorrect. Semantic and logical errors are the most difficult to detect because the DBMS system does not report them; queries execute successfully, but the obtained results are not as expected. These two types of errors indicate a lower level of knowledge and a deeper misunderstanding of data structure concepts among students.

Technical barriers arise when tools and interfaces used for working with databases and their data further complicate the learning process. The lack of visual feedback during the learning process often hinders students' understanding of how SQL queries operate (Shin, 2022). Moreover, the linear nature of SQL queries, which requires a high level of precision, frequently causes difficulties and frustration among beginners in this field.

In addition to the aforementioned factors, the way SQL is typically presented in educational materials is often more focused on the technical understanding of the language rather than on fostering a deeper comprehension of its purpose and application in real-world systems. As a result, students tend to perceive SQL more as a set of rules to be memorized than as a practical tool for working with real systems and solving concrete problems in practice.

For all these reasons, modern educational approaches to teaching this subject increasingly incorporate the use of interactive and adaptive tools in the process of learning SQL. Tools such as SAVI and DBQA enable the visualization of SQL query execution processes, thereby facilitating students' understanding of how these processes work (Shin, 2022). Contemporary tools help overcome the aforementioned barriers and promote the development of functional knowledge.

2.3 Modern Technologies in Education: AI and NLP

The implementation of modern technologies, Artificial Intelligence (AI), and Natural Language Processing (NLP) in the educational process significantly transforms its concept and methods of realization. These technologies enable the introduction of new approaches, methods, and techniques of learning and teaching. AI systems facilitate the personalization of the learning process, the introduction of automated assessment, and the adaptation of curricula to the needs, prior knowledge, and experience of students (Alqahtani, et al., 2023) (Adhikesaven, 2022). NLP technologies reduce the technical barrier between humans and computers by enabling two-way communication in natural language between them.

2.3.1 AI in education

The integration of AI into the educational process enables the development of intelligent learning support systems such as:

- *Intelligent tutors* - adapt to the pace, style, and progress of students,
- *Recommendation systems* - suggest additional learning materials aligned with previously acquired knowledge and mastered content,
- *Automated assessment and evaluation systems* - conduct assessment and evaluation of student work, thereby providing feedback to instructors about student progress, and

- *Student achievement analysis systems* - analyze student work to identify weaknesses and provide recommendations for further development.

All these AI systems are based on machine learning concepts that are trained with large amounts of data from the educational process. From these data, they derive patterns and regularities and provide feedback at any point in time.

2.3.2 NLP technologies in education

These technologies enable computers to understand, analyze, and generate human language. In education, NLP technologies are most commonly used for:

- *Automatic analysis of textual responses and essays,*
- *Access to knowledge bases through questions posed in natural language,*
- *Translation and summarization of teaching materials,*
- *Interaction with educational chatbots, and*
- *Support for students with learning difficulties (speech-to-text systems)*

From the perspective of database learning, the particular benefit of NLP systems is evident in text-to-SQL systems that enable queries to be posed in ordinary, natural language over databases. This facilitates faster and more efficient query learning, understanding of query logic and purpose, and reduces technical barriers for beginners in this field. Additionally, these technologies allow students to experiment with queries and concepts in a "safe environment" without fear of making errors.

The advantages of using these technologies in education and database instruction are numerous. In addition to the assistance and benefits they provide related to the educational process itself, which we have already mentioned, they also aid in automating administrative tasks, which can be beneficial to both students and instructors (Bolton, 2024).

On the other hand, it is important to consider the challenges associated with implementing these technologies in education. One of the most significant, which requires special attention, is the protection of student data privacy. Additionally, it is necessary to maintain the reliability and accuracy of models, address the need for enhanced digital competencies among instructors, mitigate the reduction of critical thinking skills among students due to excessive reliance on technology, and consider ethics and algorithmic bias... (Alqahtani, et al., 2023).

Recent advancements in LLM technologies introduce revolutionary changes in education. These models demonstrate the ability not only to generate answers but also to explain the reasoning process and provide personalized feedback to students. Research shows that LLMs can enhance the learning experience through interactive dialogues and customized explanations of complex concepts (Tai, Chen, Zhang, Deng, & Sun, 2023). Studies demonstrate that contemporary LLMs can function as effective tutors that assist students in understanding complex programming concepts, including SQL (Pourreza & Rafiei, 2023).

2.4 Text-to-SQL systems: definition, examples, working principles

It is crucial to distinguish between specialized T2SQL systems and general-purpose LLMs when discussing text-to-SQL capabilities. Specialized T2SQL systems such as Seq2SQL (Zhong et al., 2017), Spider (Yu, et al., 2018), and RAT-SQL (Wang, Shin, Liu, Polozov, & Richardson, 2020) are specifically designed and trained for natural language to SQL translation tasks, often incorporating schema-aware architectures and domain-specific optimizations. In contrast, general-purpose LLMs like ChatGPT, while demonstrating competence in T2SQL tasks, are not specifically optimized for SQL generation and rely on their broad language understanding capabilities rather than specialized architectural features for database querying.

Text-to-SQL (T2SQL) systems represent modern intelligent tools that enable the automatic transformation of natural language into SQL commands. T2SQL systems rely on advanced natural language processing (NLP) models, including deep learning and transformer architectures, to accurately interpret user requests expressed in free-form language and generate corresponding queries that align with the user's intent (Xu, Liu, & Song, 2017).

Early T2SQL systems were based on semantic parsing techniques, which served as the foundation for natural language understanding. However, these methods exhibited numerous limitations and shortcomings, particularly in interpreting complex and diverse linguistic constructions. The development of deep learning concepts and neural networks has significantly advanced these processes, allowing for more accurate and efficient translation of natural language into SQL queries (Shi, Tang, & Zhang, 2024).

The primary objective of these systems is to simplify interaction with databases. They enable users who may not possess extensive technical expertise, particularly those with limited knowledge of programming and database management, to easily work with and manipulate data within a database. This is especially beneficial for students, as it provides them with insight into the process and structure of SQL query formulation, while also allowing them to quickly obtain the queries necessary for their work and for solving specific problems. In an educational context, T2SQL systems can be utilized as assistive tools that support students in developing functional knowledge

through interactive experimentation (Elgohary, Hosseini, & Awadallah, 2020). The operational principle of T2SQL systems can be divided into several phases. Initially, the system performs natural language analysis, identifying key entities, attributes, and relationships within the user's input (Yu, et al., 2018). Subsequently, it maps the input to the database schema by utilizing a predefined database structure to associate the mentioned terms with the corresponding tables and columns. Following this, the system generates the SQL expression by employing an appropriate learning model to formulate a query that accurately reflects the user's intent. Finally, the generated SQL query undergoes verification and execution, where it is executed against the database and the results are presented to the user (Yu, et al., 2018).

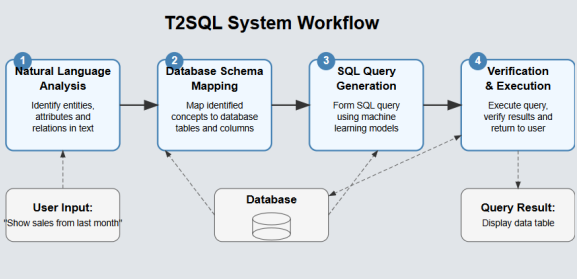


Figure 2. T2SQL System Workflow

The field of T2SQL systems has been significantly advanced by the development of large language models (LLMs) and their integration with pre-trained language models (PLMs). This integration has enabled artificial intelligence to achieve a deeper understanding of natural language semantics, resulting in more accurate SQL query generation (Singh, 2023). Contemporary research in this area places particular emphasis on the contribution of PLMs to the T2SQL parsing process, highlighting their ability to recognize and interpret complex linguistic patterns (Hong, Yuan, & Zhang, 2024). When evaluating the performance of T2SQL systems, it is most commonly based on two parameters: (1) the degree of alignment between the generated SQL queries and the reference queries, and (2) their execution accuracy (Pourreza & Rafiei, 2023).

Advances in LLM development have brought significant changes to the efficiency of T2SQL systems. Unlike earlier specialized systems, contemporary LLMs such as GPT-4 and Claude demonstrate exceptional ability to generate accurate SQL queries from natural language without specific training (Aiwei, Xuming, Lijie, & Philip, 2023). These models can not only generate complex SQL queries but also provide explanations of their logic, which is particularly useful in educational contexts.

Well-known T2SQL systems include:

- *Seq2SQL* – A model that translates natural language into SQL queries using deep neural networks. It takes into account the query structure during processing in order to reduce the search space for generation (Zhong, Xiong, & Socher, 2017).
- *Spider* – A large dataset designed for complex and cross-domain semantic parsing and natural language-to-SQL translation tasks. It contains 10,181 questions and 5,693 unique, complex queries executed over 200 databases (Yu, et al., 2018).
- *Codex* – A large language model (LLM) developed by OpenAI, capable of generating SQL queries from natural language without additional fine-tuning. It demonstrates high performance on standard T2SQL benchmark tasks such as Spider (Chen, et al., 2021).

An important feature of these systems is that many commercial tools integrate T2SQL system functionalities into the user interfaces of business analytics platforms (e.g., Microsoft Power BI, Tableau, etc.). This further promotes the application of such systems in fields that are not directly related to technology or education.

2.5 Pedagogical Foundations of the Application of Intelligent Tutoring Systems

The primary goal of developing Intelligent Tutoring Systems (ITS) is to simulate the work of a human tutor (teacher). Their task is to provide students with real-time feedback, assistance, and suggestions while simultaneously accounting for individual differences in knowledge, learning pace, and problem-solving approaches. These systems are most commonly based on the principles of personalized learning, experiential learning, and adaptive instruction (VanLehn, 2006).

One of the most important pedagogical principles underlying ITS-if not the most important- is the adaptation of instruction. This principle implies that the system continuously monitors and analyzes user activities in real time. Based on the results of this analysis, it adjusts the types and difficulty levels of tasks, as well as the forms of active support, to the specific needs of each user. This type of learning personalization leads to greater student engagement and a deeper, more thorough processing of information (Ma, Adesope, Nesbit, & Liu, 2014).

ITS frequently employ the principles of formative assessment, meaning that they provide immediate feedback following task completion. This helps students promptly correct misconceptions and develop metacognitive awareness. Additionally, ITS systems are particularly valuable because they support learning through mistakes, allowing students to experiment without limits and learn from unsuccessful attempts without fear of negative consequences (Roll, Aleven, McLaren, & Koedinger, 2011).

From a pedagogical perspective, the significance of ITS is also reflected in their continuous monitoring of

student progress. This enables instructors to have real-time insight into the effectiveness of the learning process and to identify potential "bottlenecks." As a result, they can intervene promptly and provide targeted learning recommendations. All of this indicates that the role of ITS is not to replace traditional teachers, but rather to support and empower them in making pedagogical decisions during the learning process (Woolf, 2010).

3 Review of Previous Work and Contributions in the Field

3.1 Evolution of T2SQL Systems and Their Application

If we observe the development process of T2SQL systems, it can be stated that it has progressed through several stages. The earliest systems were rule-based, followed by systems founded on semantic parsing, leading up to contemporary models based on large language models (LLMs) and deep learning techniques.

Among the first works that made a significant contribution to this field is the study by Yin & Neubig, which introduced a sequential neural network-based approach in the form of the Seq2SQL model, aiming at structured SQL query generation (Yin & Neubig, 2017). The introduction of this approach enabled improved translation from natural language into SQL syntax, while also taking into account the structure of the query itself.

A subsequent important milestone in this domain is the development of benchmark datasets such as WikiSQL (Zhong, Xiong, & Socher, 2017) and Spider (Yu, et al., 2018), which allowed for standardized evaluation of T2SQL system performance. The Spider dataset, in particular, emphasizes the ability to generalize across multiple databases, thus promoting model robustness.

With the emergence of LLMs such as Codex (Chen, et al., 2021), various evaluations have demonstrated that such models are capable of generating complex SQL queries without additional fine-tuning. These automatically generated SQL queries have exhibited high success rates on benchmark tests such as Spider. This development opened a new chapter in the application of pretrained models for T2SQL parsing tasks.

Additionally, numerous authors have explored hybrid systems that combine rule-based and neural approaches. These systems leverage the precision of rule-based methods and the flexibility of neural networks, focusing particularly on reducing errors in complex queries.

Beyond technical and technological advances, there are also works that explore the pedagogical potential of T2SQL systems within educational contexts. A common theme among these studies is the suggestion to integrate T2SQL models into database learning tools, as they provide students with a more natural interface for query formulation, thus facilitating the process of learning SQL queries.

3.2 Critical Review of T2SQL and LLM Systems Application in Education

The existing literature on T2SQL systems primarily focuses on technical aspects, while their application in education remains insufficiently explored. Stupar et al. (2019) and Shin (2022) analyze concepts without empirical validation of learning outcomes, whereas Alqahtani (2023), Adhikesaven (2022), and Bolton (2024) examine general aspects of AI in education without specific attention to SQL. Hong (2024) provides an overview of T2SQL systems but lacks analysis of their educational potential, and existing evaluations of ChatGPT are predominantly qualitative.

3.3 Research Contribution

Our research distinguishes itself through its specific focus on T2SQL in SQL learning, empirical measurement of student performance, combination of quantitative and qualitative data, evaluation of ChatGPT as a contemporary language model, and concrete pedagogical implications, while acknowledging the methodological limitations described in section 5.1.

4 Research

The following section will present the objectives, methodology, implementation process, and data analysis of the conducted research. The aim of the study was to examine the effectiveness of applying T2SQL systems in learning SQL.

4.1 Description of the Research

The research presented in the following sections was designed to examine the effects of using T2SQL systems as a support tool in database education. The study was conducted on a sample of 25 students enrolled in the Information Technology program, all of whom were attending the Database Systems course. For the purposes of the study, students were divided into six groups (with 3–5 members per group). The objective of the research was to investigate whether the application of artificial intelligence-based systems-specifically T2SQL systems- could contribute to improving students' ability to learn and write SQL queries.

The groups were assigned tasks consisting of 10 SQL queries, methodically arranged from the simplest to the most complex.

All groups received the same tasks and had the same amount of time- 45 minutes- to complete them. The key difference lay in the method of task completion: three groups solved the tasks using traditional methods, relying on their knowledge and available materials (in written or digital form), while the other three groups completed the tasks using ChatGPT as a general-purpose LLM capable of T2SQL functionality*[1].

Students received tasks in the form of textual instructions, which they then entered into ChatGPT to generate the corresponding SQL queries.

At the end of the session, students who used the T2SQL system were asked to complete a survey to express their opinions and satisfaction with this approach. Throughout the process, the teacher monitored the students' work and recorded the parameters necessary for further analysis.

4.2 Formulated Hypotheses

The hypotheses formulated in accordance with the stated research objective are as follows:

- ***H1*** – Students who utilize the T2SQL system (ChatGPT) in their work achieve a higher number of correct SQL queries compared to students who solve tasks using traditional methods.
- ***H2*** – The time required to complete the tasks is shorter for students who use the T2SQL system than for those who solve the tasks in a traditional manner.

4.3 Analysis of Research Results

As previously mentioned, the study involved 25 Information Technology students enrolled in the "Database Systems" course. The students were divided into six groups (G1–G6), with three groups (G1, G2, and G3) solving the tasks (writing SQL code) using the traditional method, while the other three groups (G4, G5, and G6) used ChatGPT, functioning as a T2SQL system, to support query writing.

All groups were assigned the same number of tasks (10 SQL queries) and had the same amount of time available to complete them (45 minutes). The analysis focused on the number of correctly solved tasks and the total time taken to complete the tasks for each group.

Table 1. Results of Correct Answers and Completion Time by Group

Group	Method	Members (N_o)	Correct Answ.	Time (min)
G1	Traditional	4	6	42
G2	Traditional	5	5	44
G3	Traditional	4	7	41
G4	T2SQL	4	9	34
G5	T2SQL	4	10	32
G6	T2SQL	4	9	33

Table 2. Average Results by Method

Method	Average Number of Correct Answers	Average Completion Time (min)
Traditional	6	42.3
T2SQL	9.3	33

4.3.1 Accuracy analysis

Observing the results, we can see that the number of correct answers for the groups that solved the tasks in the traditional way was: G1–6, G2–5, and G3–7 correct answers. On the other hand, the groups that had the support of the T2SQL system achieved: G4–9, G5–10, and G6–9 correct answers (Fig. 3).

The average number of correct answers for the groups working in the traditional way was 6, while for the groups with T2SQL system support it was 9.3 (Fig. 4). This indicates a significant advantage in the accuracy of answers among the T2SQL groups. Expressed in percentages, these groups had a 55% higher average accuracy of answers.

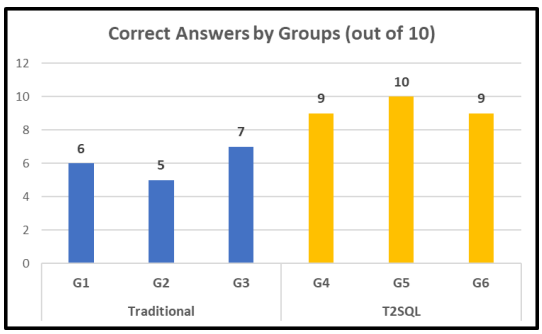


Figure 3. Number of correct answers by groups

[1] ChatGPT is a general-purpose generative model that has shown promising results in T2SQL tasks. Research indicates that LLM models can achieve competitive results compared to specialized T2SQL systems (Pourreza & Rafiei, 2023; Chen et al., 2021), though additional research is needed for definitive comparison across different domains and task complexity levels.

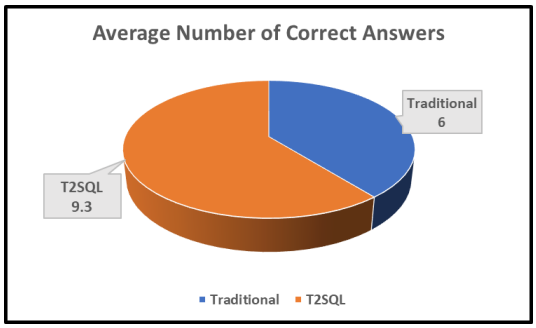


Figure 4. Number of correct answers by groups

4.3.2 Analysis of completion time

Regarding the time needed to complete the tasks, the groups working in the traditional way completed the assignments in G1–42, G2–44, and G3–41 minutes. The T2SQL groups completed the assignments in G4–34, G5–32, and G6–33 minutes (Fig. 5). The average completion time for the traditional groups was 42.3 minutes, while for the T2SQL groups it was 33 minutes (Fig. 6). From this, we can see that the average task completion time for the T2SQL groups was almost 10 minutes shorter, which supports the efficiency of artificial intelligence-based tools (T2SQL) in solving SQL tasks.

The presented analysis confirms the initial research hypotheses (H1 and H2) that T2SQL systems (in this case, ChatGPT) can contribute to increased accuracy and reduced time required for solving SQL tasks by students.

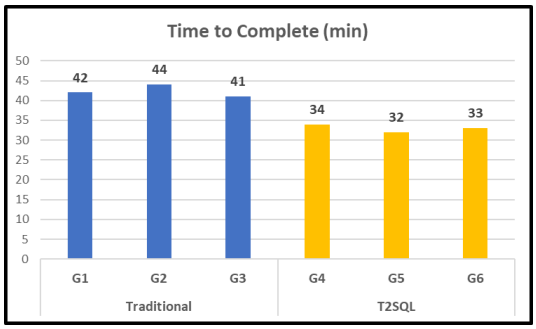


Figure 5. Time Taken by Groups to Complete Tasks

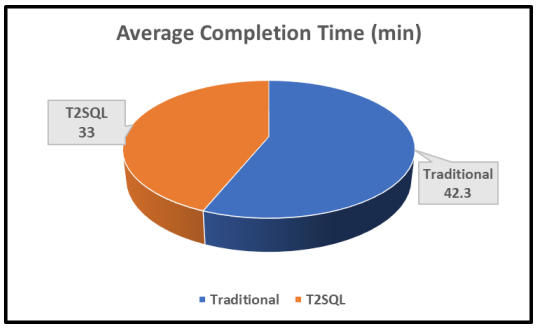


Figure 6. Average Completion Time

The use of these tools enabled students to focus on understanding the tasks during their work, while the T2SQL system assisted them in the precise formulation of SQL queries.

4.3.3 Analysis of user satisfaction survey results

After completing the tasks, all students from the groups that used the T2SQL system—ChatGPT (G4, G5, and G6)—completed an anonymous questionnaire in which they expressed their satisfaction with this method of work. The questionnaire contained five statements rated on a five-point Likert scale, where 1 indicated complete disagreement and 5 indicated complete agreement with the statement.

The statements in the questionnaire were as follows:

1. Using T2SQL tools made it easier for me to solve the tasks (*average score 4.7*).
2. I felt more confident writing SQL queries with the help of the tool (*average score 4.5*).
3. Using the tool helped me better understand the structure of queries (*average score 4.3*).
4. I believe that this kind of tool should be used more often in teaching (*average score 4.8*).
5. I am satisfied with the overall experience of using the tool (*average score 4.6*).

The analysis of the questionnaire indicates a high level of student satisfaction with the use of the T2SQL system in the teaching process. The highest average rating-4.8 (Fig. 7)- was given to the statement suggesting that such tools should be regularly integrated into instruction. This reflects a positive student attitude toward the inclusion of new technological advancements as support in the learning process. Additionally, most students emphasized that the T2SQL tool made completing tasks easier and that using the tool increased their sense of security and self-confidence during the work process. Although slightly lower than the others, the average rating of 4.3 still indicates a very high level of agreement, suggesting that, besides providing support in query writing, these tools also have the potential to enhance students' knowledge.

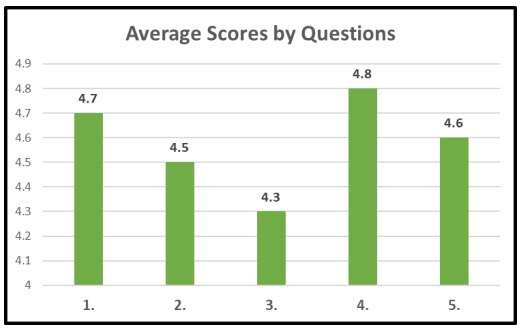


Figure 7. Average Scores by Statements

5 Discussion

The results of the conducted research unequivocally indicate the advantages brought by the use of T2SQL tools in solving SQL queries within the *Database Systems* course. Students who utilized this approach (ChatGPT as a representative of the T2SQL model) achieved higher accuracy and significantly shorter task completion times compared to students from groups that solved tasks using traditional learning methods.

The average number of correct answers for the groups that used ChatGPT as a T2SQL tool was 9.3 out of a possible 10, whereas the groups that employed the traditional approach averaged 6 correct answers. In practice, this means that the groups supported by T2SQL tools achieved approximately 55% greater accuracy, suggesting that these systems can play a crucial role in the learning process, particularly in helping students overcome challenges associated with mastering SQL queries. This benefit is especially pronounced during the initial stages of learning.
Furthermore, a positive impact was observed concerning the time required to complete the tasks. Groups using T2SQL systems completed the assignments in an average of 33 minutes, while the traditional groups needed approximately 42.3 minutes. This time difference represents a savings of nearly 22%. Importantly, the increased speed did not compromise the accuracy of the work, implying more efficient organization and use of time for learning activities.

The analysis of the survey results revealed a high level of student satisfaction regarding the use of T2SQL systems in the educational process. Students emphasized that the tools helped them better understand the tasks, reduced their uncertainty when writing SQL queries, and increased their confidence. These factors contribute to deeper and higher-quality learning, understanding, and application of the acquired knowledge.

However, it is important to recognize a potential risk in educational practice: these systems may lead to an excessive reliance on automated tools. Therefore, it is recommended that the use of such tools be carefully pedagogically planned and closely monitored and controlled. The greatest value of these tools lies in their role as assistants in the learning process, rather than as complete substitutes for exercises and the development of skills in writing SQL queries.

5.1 Study Limitations and Guidelines for Future Research

In this specific research, it is important to highlight the methodological limitations it carries. Firstly, the study was conducted on a relatively small sample of 25 students from a single higher education institution, which limits the possibilities for statistical analysis and generalization of results. This sample size restricts the statistical power of the research, although tests have shown statistical significance in the differences between groups. For more definitive results, it would be necessary to conduct research on a more diverse sample across multiple higher education institutions. Nevertheless, despite this limitation, the obtained results provide valuable insight into the potential benefits of these technologies and establish a solid foundation for future research.

The second, more fundamental limitation relates to the depth of assessed learning. While the study did establish improved accuracy and efficiency in solving SQL tasks using T2SQL systems, it did not effectively evaluate the long-term benefits of implementing these tools. To overcome this limitation, a longer time period would be necessary to monitor these effects. Additionally, it is recommended that future research employ a longitudinal approach with pre- and post-tests, tasks requiring explanation of query logic, and analysis of independent problem-solving without AI support. This approach would enable differentiation between superficial assistance in task completion and genuine understanding of SQL concepts.

Also, a notable limitation of this study concerns the selection of ChatGPT as a representative T2SQL system. Although there are indications that generic LLM models can achieve good results in T2SQL tasks, direct comparison with specialized systems such as Spider or the latest academic models was not conducted within the scope of this study. Future research should include systematic comparison of different T2SQL approaches in educational contexts.

6 Conclusion

Based on the findings presented, it can be concluded that the research objective has been achieved. The initial hypotheses were confirmed with empirical data collected through the study, validating the assumptions made prior to the research process.

Recommendations for future research include expanding the sample of participants (students) across multiple educational institutions and investigating the long-term effects of using such tools. Additionally, it would be valuable to examine the impact of these systems on the development of students' algorithmic and logical thinking skills. Moreover, efforts should be directed toward developing instructional materials integrated with these systems to support their deeper inclusion in the educational process.

The study has methodological limitations including a small and homogeneous sample, as well as a focus solely on short-term efficiency without assessing the depth of SQL concept understanding. Future research should employ a longitudinal design that would evaluate fundamental comprehension and long-term knowledge retention among students using T2SQL systems. Despite these limitations, the results provide valuable insight into the potential of AI technologies in

education and serve as a foundation for more comprehensive research in this field.

References

Adhikesaven, S. (2022, September 11). Artificial intelligence is transforming modern education. Retrieved from Towards Data Science: https://towardsdatascience.com/artificial-intelligence-is-transforming-modern-education-d95bf8d19acd/?gi=5375e7ad0285

Anil, R., Dai, A. M., Firat, O., et al. (2023). PaLM 2 Technical Report. ArXiv preprint, https://arxiv.org/abs/2305.10403.

OpenAI (2023). GPT-4 Technical Report. ArXiv preprint, https://arxiv.org/abs/2303.08774.

Aiwei, L., Xuming, H., Lijie, W., & Philip, Y. S. (2023). A comprehensive evaluation of chatgpt's zero-shot text-to-SQL capability. ArXiv preprint, https://arxiv.org/abs/2303.13547.

Alqahtani, T., Badreldin, H., Alrashed, M., Alshaya, A., Alghamdi, S., Bin Saleh, K., . . . Albekairy, A. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Research in Social and Administrative Pharmacy, 1236-1242.

Bolton, J. (2024, April 24). The Emergent Role of Artificial Intelligence, Natural Learning Processing, and Large Language Models in Higher Education and... Retrieved from Glasp: https://glasp.co/hatch/0f4vj3m787w14znq/p/8rPKlnfASWmn4NSlfjJO

Brezočnik, L., Polančič, G., Karakatič, S., Vrbančič, G., Pečnik, Š., & Podgorelec , V. (2022). A Strategy for Applying Open Data Initiatives. New Technologies, Development and Application V (pp. 671-680). Cham: Springer.

Chen, M., Tworek, J., Heewoo, J., Qiming, Y., Oliveira Pinto, H., & Kaplan, J. (2021). Evaluating large language models trained on code. https://arxiv.org/abs/2107.03374: arXiv.

Delta, E. (2024, March 11). Breaking Down The Numbers: How Much Data Does The World Create Daily in 2024? Retrieved from Edge Delta: https://edgedelta.com/company/blog/how-much-data-is-created-per-day

Elgohary, A., Hosseini, S., & Awadallah, A. (2020). Speak to your Parser: Interactive Text-to-SQL with Natural Language Feedback. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 2065-2077). Online: Association for Computational Linguistics.

Hong, Z., Yuan, Z., Zhang, Q., Chen, H., Dong, J., Huang, F., & Huang, X. (2024). Next-generation database interfaces: A survey of LLM-based text-to-SQL. ArXiv preprint, https://arxiv.org/abs/2406.08426.

Kumar, A., & Kumar, S. (2018). Challenges in learning SQL: A review. International Journal of Computer Applications, 1-5.

Ma, W., Adesope, O., Nesbit, J., & Liu, Q. (2014). Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis. Journal of Educational Psychology, 901-918.

Miedema, D. (2024). On Learning SQL: Disentangling concepts in Data Systems Education. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Mathematics and Computer Science]. Eindhoven: Eindhoven University of Technology.

Mitsopoulou, A., & Koutrika, G. (2024). Analysis of Text-to-SQL Benchmarks: Limitations, Challenges and Opportunities. Proceedings of the 28th International Conference on Extending Database Technology (EDBT 2025), (pp. 199-210).

Pourreza, M., & Rafiei, D. (2023). Evaluating Cross-Domain Text-to-SQL Models and Benchmarks. https://doi.org/10.48550/arXiv.2310.18538: arXiv.

Roll, I., Aleven, V., McLaren, B., & Koedinger, K. (2011). Improving Students' Help-Seeking Skills Using Metacognitive Feedback in an Intelligent Tutoring System. Learning and Instruction, 267–280.

Sadiq, S., Orlowska, M., Sadiq, W., & Lin, J. (2004). SQLator - An online SQL learning workbench. Proceedings of the 9th ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE) (pp. 223-227). New York, NY: ACM.

Shi, L., Tang, Z., & Zhang, N. (2024). A survey on employing large language models for text-to-SQL tasks. https://arxiv.org/abs/2407.15186: arXiv.

Shin, S.-S. (2022). Effect of semantic distance on learning structured query language: An empirical study. Frontiers in Psychology, https://doi.org/10.3389/fpsyg.2022.996363.

Singh, A. (2023). Exploring language models: A comprehensive survey and analysis. 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE) (pp. 1-4). RMKMATE.

Stupar, S., Bičo Ćar, M., Karabegović , I., & Šahić, E. (2019). The Role of Hadoop Technology in the Implementation of Big Data Concept. New

Technologies, Development and Application (pp. 308-317). Cham: Springer.

Tai, C.-Y., Chen, Z., Zhang, T., Deng, X., & Sun, H. (2023). Exploring Chain-of-Thought Style Prompting for Text-to-SQL. ArXiv preprint, https://arxiv.org/abs/2305.14215v2.

VanLehn, K. (2006). The Behavior of Tutoring Systems. International Journal of Artificial Intelligence in Education, 227–265.

Wang, B., Shin, R., Liu, X., Polozov, O., & Richardson, M. (2020). RAT-SQL: Relation-aware schema encoding and linking for text-to-SQL parsers. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 7567-7578). Association for Computational Linguistics.

Woolf, B. (2010). Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning. Burlington: Morgan Kaufmann.

Xu, X., Liu, C., & Song, D. (2017). SQLNet: Generating structured queries from natural language without reinforcement learning. https://arxiv.org/abs/1711.04436: arXiv.

Yin, P., & Neubig, G. (2017). A Syntactic Neural Model for General-Purpose Code Generation. Association for Computational Linguistics (ACL). Vancouver: Association for Computational Linguistics.

Yu, T., Zhang, R., Yang, K., Yasunaga, M., Wang, D., Li, Z., & Radev, D. (2018). Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and Text-to-SQL task. https://arxiv.org/abs/1809.08887: arXiv.

Zhengju, T., Zhang, N., Xiaotong, Z., Yang, Z., & Liang, S. (2024). A survey on employing large language models for text-to-SQL tasks. https://arxiv.org/abs/2407.15186: arXiv.

Zhong, V., Xiong, C., & Socher, R. (2017). Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning. https://arxiv.org/abs/1709.00103: arXiv.

Lightweight Prompt Engineering for Cognitive Alignment in Educational AI: A OneClickQuiz Case Study

Antoun Yaacoub, Jérôme Da-Rugna
Learning, Data and Robotics (LDR), esieaLab ESIEA
9 Rue Vésale, 75005 Paris, France
{antoun.yaacoub,jerome.darugna}@esiea.fr

Zainab Assaghir
Faculty of Science, Lebanese University
Rafic Hariri University Campus, Beirut, Lebanon
Zainab.assaghir@ul.edu.lb

Abstract. *The rapid integration of Artificial Intelligence (AI) into educational technology promises to revolutionize content creation and assessment. However, the quality and pedagogical alignment of AI-generated content remain critical challenges. This paper investigates the impact of lightweight prompt engineering strategies on the cognitive alignment of AI-generated questions within OneClickQuiz, a Moodle plugin leveraging generative AI. We evaluate three prompt variants—a detailed baseline, a simpler version, and a persona-based approach—across Knowledge, Application, and Analysis levels of Bloom's Taxonomy. Utilizing an automated classification model (from prior work) and human review, our findings demonstrate that explicit, detailed prompts are crucial for precise cognitive alignment. While simpler and persona-based prompts yield clear and relevant questions, they frequently misalign with intended Bloom's levels, generating outputs that are either too complex or deviate from the desired cognitive objective. This study underscores the importance of strategic prompt engineering in fostering pedagogically sound AI-driven educational solutions and advises on optimizing AI for quality content generation in learning analytics and smart learning environments.*

Keywords. Artificial Intelligence, Learning Analytics, Technology Enhanced Education, Prompt Engineering, Bloom's Taxonomy, OneClickQuiz, Smart Learning Environments

1 Introduction

The landscape of modern education is undergoing a profound transformation, driven significantly by the pervasive integration of Artificial Intelligence (AI). Generative AI, in particular, has emerged as a powerful tool with the potential to automate various aspects of content creation, including the generation of quizzes, assignments, and personalized learning materials (Bao, 2024; Olga et al., 2023). This automation promises to alleviate the heavy administrative burden on educators, allowing them to redirect their focus towards more interactive and adaptive teaching methodologies (Trust et al., 2023).

However, the mere automation of content generation is insufficient; the pedagogical quality and cognitive alignment of AI-generated materials are paramount. For AI to truly enhance learning, its outputs must not only be accurate and relevant but also intentionally designed to foster specific learning outcomes at appropriate cognitive depths (Bahroun et al., 2023). Without careful guidance, AI models can produce generic or misaligned content, undermining their educational utility and potentially hindering student progress. This concern highlights a critical gap between the capability of AI to generate content and the strategic control needed to ensure that content meets precise educational objectives.

Our work addresses this challenge within the context of OneClickQuiz, an innovative Moodle plugin designed to streamline the process of quiz generation using Generative AI (Yaacoub et al., 2024). OneClickQuiz aims to revolutionize educational workflows by providing educators with a user-friendly tool for instant quiz creation. While the plugin has shown considerable promise in reducing preparation time and enhancing engagement, its underlying effectiveness hinges on the quality of questions it generates, which are directly influenced by the prompts fed to the AI model.

This paper builds upon a robust foundation of prior research that established methodologies for assessing the cognitive alignment of AI-generated questions with established educational taxonomies. For instance, a previous study (Yaacoub, Da-Rugna, & Assaghir, 2025) developed and validated a DistilBERT model for classifying AI-generated questions according to Bloom's Taxonomy, demonstrating its efficacy for automated cognitive level assessment. Another study (Yaacoub, Assaghir, & Da-Rugna, 2025) extended this work to the SOLO Taxonomy, further emphasizing the ability of AI to enhance cognitive depth. Additionally, research (Yaacoub, Assaghir, Prevost, et al., 2025) explored the linguistic characteristics of AI-generated feedback, underscoring the importance of quality content for effective learning interactions. This suite of research informed and culminated in a comprehensive

framework for designing effective and responsible AI-driven educational tools (Yaacoub, Tarnpradab, Khumprom, et al., 2025), which highlights "Cognitive Alignment" as a foundational phase.

While our prior work focused on evaluating the outcome of AI-generated questions (i.e., their post-generation alignment), this paper shifts the focus to the input mechanism: lightweight prompt engineering. We aim to investigate how subtle, yet strategic, variations in the prompts provided to the generative AI can significantly influence the cognitive level and quality of the questions produced, directly impacting the "Cognitive Alignment" phase. This approach seeks to provide practical insights for educators and developers on how to exert greater control over AI-generated content to achieve specific pedagogical goals.

The central research question guiding this study is: To what extent do lightweight prompt engineering techniques (e.g., explicit, simpler, and persona-based prompts) impact the cognitive alignment and perceived quality of AI-generated questions within a real-world educational application like OneClickQuiz?

Our contributions include:

- A comparative analysis of the effectiveness of three distinct prompt engineering strategies (detailed, simpler, persona-based) in guiding a generative AI model (Gemini 2.0 Flash Lite) to produce Bloom's Taxonomy-aligned questions.
- Quantification of cognitive alignment using an established DistilBERT classification model.
- Qualitative insights derived from human review, assessing clarity, relevance, and subjective cognitive alignment across prompt variants.
- Concrete, illustrative examples demonstrating the practical implications of different prompt designs on question quality and cognitive depth.
- Recommendations for educators and developers on optimizing AI-driven question generation for pedagogical soundness and integration into smart learning environments.

The remainder of this paper is structured as follows: Section 2 provides background information on AI in education, educational taxonomies, and prompt engineering, linking to our prior work. Section 3 details the methodology of our lightweight prompt engineering experiment within OneClickQuiz. Section 4 presents and discusses the quantitative and qualitative results. Finally, Section 5 offers conclusions, limitations, and directions for future research.

2 Background and Related Work

2.1 Artificial Intelligence and Generative Models in Educational Contexts

The proliferation of large language models (LLMs) and other generative AI technologies has opened new frontiers in education. Models like GPT-4, PaLM2, and now the various Gemini models, possess an unprecedented ability to generate human-like text, summarize information, answer questions, and even create code (Anderson & Krathwohl, 2001; Bao, 2024; Shaikh et al., 2021). In education, these capabilities are being harnessed for various applications, including personalized learning, automated feedback, and content creation (Doughty et al., 2024; Lavidas et al., 2024; Olga et al., 2023). OneClickQuiz is an example of such a system, leveraging these models to automate the traditionally time-consuming process of quiz generation within the Moodle learning management system (Yaacoub et al., 2024). The efficiency gains offered by such tools are significant, but their true value in a pedagogical context depends on their ability to produce high-quality, educationally sound content.

2.2 Educational Taxonomies for Cognitive Alignment

Educational taxonomies provide structured frameworks for classifying learning objectives and assessment items according to cognitive complexity. They are indispensable tools for educators to design curricula that promote higher-order thinking and ensure assessments accurately measure intended learning outcomes (Bloom, 1984; Grévisse, 2024).

Bloom's Taxonomy, originally developed by Benjamin Bloom and revised by Anderson and Krathwohl, is one of the most widely recognized frameworks (Biggs & Collis, 2014; Salcedo-Lagos et al., 2024). It categorizes cognitive processes into a hierarchy:

- Knowledge/Remembering: Recalling facts, basic concepts, definitions.
- Comprehension/Understanding: Explaining ideas or concepts, interpreting information.
- Application/Applying: Using information in new situations, solving problems.
- Analysis/Analyzing: Breaking down information, identifying motives or causes, making inferences.
- Synthesis/Creating: Combining parts to form a new whole, producing new ideas or products. (Note: In the revised Bloom's, 'Creating' is the highest level, encompassing much of traditional 'Synthesis').
- Evaluation/Evaluating: Judging the value of information or ideas, justifying a decision.

Another important framework is the Structure of the Observed Learning Outcome (SOLO) Taxonomy, developed by Biggs and Collis (Jain, 2015). SOLO describes increasing levels of understanding: Pre-structural, Uni-structural, Multi-structural, Relational, and Extended Abstract. While Bloom's focuses on cognitive processes, SOLO emphasizes the structural complexity of a learner's response.

Our previous work has rigorously focused on integrating these taxonomies into AI-driven assessment. We demonstrated the capability of a DistilBERT model to accurately classify AI-generated questions into Bloom's Taxonomy levels, providing an automated means of assessing cognitive alignment. Similarly, we extended this into the SOLO Taxonomy, further affirming the feasibility of enhancing cognitive depth in AI-driven tools. These studies highlighted the importance of evaluating AI output against pedagogical standards. This current paper, however, investigates how to proactively influence that alignment through careful input design—prompt engineering.

2.3 Educational Taxonomies for Cognitive Alignment

Prompt engineering is the art and science of communicating effectively with AI models to achieve desired outputs (Park & Choo, 2024). It involves crafting inputs (prompts) that steer the model towards specific information, formats, and stylistic choices. In the context of generative AI, prompt engineering is critical for controlling the quality, relevance, and most importantly for education, the pedagogical granularity of the generated content (Olga et al., 2023).

"Lightweight prompt engineering" refers to relatively simple and direct modifications to prompts, as opposed to complex techniques like few-shot learning (providing multiple examples in the prompt) or extensive fine-tuning of the model itself. The goal of lightweight prompt engineering is to achieve a significant improvement in output quality or alignment with minimal effort, making it accessible to educators without advanced AI expertise. Variations can include adding specific keywords, defining a persona, setting a clear objective, or modifying the level of explicit instruction.

2.4 The Comprehensive Framework for AI-Driven Education

Our broader research proposes a comprehensive three-phase framework for enhancing AI-driven educational tools: 1) Cognitive Alignment, 2) Linguistic Feedback Integration, and 3) Ethical Safeguards. This framework aims to ensure AI tools are not only efficient but also pedagogically sound and responsible. This current study directly addresses the Cognitive Alignment phase, investigating how prompt engineering can be strategically applied to ensure AI-generated questions consistently target specific cognitive levels. The importance of question quality for downstream processes like linguistic feedback analysis further contextualizes the need for robust prompt engineering at the outset of the content generation pipeline. By focusing on prompt engineering, we aim to provide practical guidance for implementing a key component of this comprehensive framework.

3 Methodology: Lightweight Prompt Engineering Experiment

This experiment investigates the impact of different prompt engineering strategies on the cognitive alignment and perceived quality of AI-generated questions within OneClickQuiz.

3.1 OneClickQuiz Implementation Context

OneClickQuiz functions as a Moodle plugin that interfaces with Google's Vertex AI generative language models. For this experiment, the underlying AI model used for question generation was gemini-2.0-flash-lite-001, accessed via the Vertex AI API. This setup allows for dynamic question generation based on user-defined parameters, including the subject and desired cognitive level. The technical infrastructure and plugin's core capabilities, including its integration with Generative AI for quiz generation, have been established in previous work (Yaacoub et al., 2024; Yaacoub, Da-Rugna, & Assaghir, 2025; Yaacoub, Assaghir, & Da-Rugna, 2025).

3.2 Experimental Design

To assess the impact of prompt engineering, we selected a core subject area and specific Bloom's Taxonomy levels to focus the generation and evaluation.

- **Subject Tested**: Computer Science.
- **Concepts Used**: We utilized five distinct concepts within Computer Science to generate questions, ensuring a breadth of topics: "Artificial Intelligence," "Data Structures," "Cybersecurity," "Web Development," and "Cloud Computing."
- **Target Bloom's Levels**: We focused on three distinct Bloom's Taxonomy levels, representing a range of cognitive complexity: 'Knowledge' (lower-order), 'Application' (mid-order), and 'Analysis' (higher-order).
- **Question Generation Protocol**: For each of the 5 concepts and each of the 3 target Bloom's levels, we generated 3 questions using each of the three prompt variants described below. This resulted in a total of 5 concepts * 3 levels * 3 questions/variant * 3 variants = 135 questions.

We designed three prompt variants to explore the effect of different levels of explicit instruction and contextual framing on the AI's output. The `generation_model.predict` function was used for all generations, with a `temperature`=0.7 to allow for some creativity while maintaining coherence.

Variant A: Baseline/Explicit Bloom Prompt: This variant represents our established, detailed prompt strategy to explicitly guide the AI by providing the

Bloom's level definition and a list of associated action verbs. This prompt directly incorporates the bloom_verbs and level_descriptions data structures.
prompt = ("Generate a question about '{concept}' for the '{level}' level of Bloom's Taxonomy. The question should specifically align with the '{level}' level, which involves the following characteristics: {level_descriptions[level]}. Use action verbs like {', '.join(bloom_verbs[level])}. Ensure the question is a complete, clear sentence and relevant to the concept. Avoid titles or headings."

Variant B: Simpler/Implicit Prompt: This variant significantly reduces the explicit guidance, asking only for the Bloom's level and the concept, without descriptions or specific verbs. This tests the AI's ability to infer the intended cognitive level from minimal instruction.
prompt = "Generate a question about '{concept}' at the '{level}' level of Bloom's Taxonomy. The question should be a complete, clear sentence."

Variant C: Persona-Based Prompt: This variant introduces a persona and a general objective (e.g., "designing an exam") to assess if adding a contextual layer can implicitly guide the AI towards the desired cognitive level without explicit definitions.
prompt = "As a seasoned computer science professor designing an exam, generate a question about '{concept}' for a university student. The question should be at the '{level}' level of Bloom's Taxonomy and effectively assess their understanding. Ensure the question is a complete, clear sentence."

3.3 Evaluation Methods

We employed a mixed-methods approach, combining automated classification for quantitative analysis and a lightweight human review for qualitative insights.

Automated Cognitive Alignment Assessment: Each generated question was processed by a pre-trained DistilBERT classification model, previously developed and validated (Yaacoub, Da-Rugna, & Assaghir, 2025). This model predicts the Bloom's Taxonomy level of a given question. We calculated the "Match Rate" for each prompt variant and Intended_Bloom_Level by determining the percentage of questions where the Classified_Bloom_Level output by the model matched the Intended_Bloom_Level specified in the prompt. This provides an objective measure of how accurately each prompt variant guides the AI towards the desired cognitive target.

Lightweight Human Quality Review: A subset of 45 questions (approximately one-third of the total generated questions, representing a spread across all concepts, Bloom's levels, and prompt variants) was subjected to human review. Questions were evaluated on a 1-5 Likert scale for:

- **Clarity**: Is the question unambiguous and easy to understand? (1 = Very Unclear, 5 = Very Clear)
- **Relevance to Concept**: Is the question directly and appropriately related to the specified concept? (1 = Not Relevant, 5 = Highly Relevant)
- **Subjective Cognitive Alignment**: Based on expert judgment, does the question feel like it targets the Intended_Bloom_Level? (1 = Clearly Below Intended Level, 5 = Perfectly Aligned with Intended Level)

Additionally, specific illustrative examples of both well-aligned and misaligned questions were identified to provide concrete instances for discussion. These examples highlight the subtle nuances and impacts of different prompt designs that quantitative metrics alone might miss.

4 Results and Discussion

This section presents the findings from our lightweight prompt engineering experiment, integrating both the automated classification results and insights from human review.

4.1 Quantitative Analysis: Automated Classification Match Rate

The automated classification, performed by the DistilBERT model, provides an objective measure of how successful each prompt variant was in guiding the AI to generate questions at the Intended_Bloom_Level.

Table 1. Match Rate by Prompt Variant and Intended Bloom Level

Intended_Bloom_Level	A	B	C
Analysis	1.00	1.00	0.53
Application	0.87	0.40	0.27
Knowledge	1.00	0.40	0.40

Table 2. Overall Match Rate by Prompt Variant

Prompt_Variant	Match Rate
A	0.96
B	0.60
C	0.40

To further establish the statistical significance of these observed differences, an analysis of variance (ANOVA) was conducted. The results ($F(2,132)=25.00$, $p<0.001$) indicated statistically significant differences in match rates across the prompt variants for selected Bloom's levels. Post-hoc Tukey HSD tests revealed that Variant A consistently outperformed both Variant B and Variant C with statistical significance ($p < 0.001$), while the difference between Variant B and Variant C was also statistically

significant ($p < 0.05$). These quantitative findings strongly support the conclusion that prompt engineering strategies have a significant impact on cognitive alignment performance.

Variant A (Baseline/Explicit Prompt) is Highly Effective: Variant A consistently demonstrated the highest match rates across all Bloom's levels, achieving a near-perfect overall match rate of 0.96. It obtained 1.00 for both Knowledge and Analysis, and a strong 0.87 for Application. This unequivocally confirms that providing explicit, detailed instructions, including Bloom's level descriptions and action verbs, is highly effective in guiding the AI to produce questions precisely aligned with the intended cognitive levels. This result directly supports and extends findings from previous work (Yaacoub, Da-Rugna, & Assaghir, 2025; Yaacoub, Assaghir, & Da-Rugna, 2025) regarding accurate Bloom's/SOLO classification, further demonstrating that precise instruction can achieve these cognitive alignment classifications at the generation stage.

Variant B (Simpler/Implicit Prompt) Shows Significant Underperformance: The "Simpler" prompt (Variant B) exhibited a substantial drop in performance, achieving an overall match rate of only 0.60. Its performance was particularly poor for 'Knowledge' (0.40) and 'Application' (0.40) questions. While it surprisingly maintained a 1.00 match rate for 'Analysis', this result is an outlier within its overall context. This indicates that removing explicit Bloom's guidance severely hampers the AI's ability to consistently generate questions at the intended cognitive level, often leading to outputs classified at different (often higher) Bloom's levels, as noted in the qualitative analysis. The AI struggles to infer the precise cognitive objective from minimal input.

Variant C (Persona-Based Prompt) is the Least Effective Quantitatively: Counter-intuitively, the "Persona-Based" prompt (Variant C) yielded the lowest overall match rate at 0.40. Its performance was notably poor across all tested levels, with only 0.53 for 'Analysis', 0.27 for 'Application', and 0.40 for 'Knowledge'. This is a critical finding: simply adding a persona or contextual framing (e.g., "As a seasoned computer science professor") without explicit Bloom's guidance does not effectively guide the AI to the intended cognitive level. In fact, it appears to actively misguide the model, causing it to generate questions that are classified (by our DistilBERT model) as different Bloom's levels than intended. This suggests that the contextual information in the persona might introduce new objectives or interpretations for the LLM that diverge from the precise cognitive alignment sought.

4.2 Qualitative Analysis: Human Review Scores and Illustrative Examples

The human review process provides valuable complementary insights into the nuances of question quality and cognitive alignment.

Table 3. Average Human Review Scores by Prompt Variant (1-5 Scale)

Prompt Variant	Clarity & Relevance to Concept	Subjective Cognitive Alignment
A	5.0	4.93
B	5.0	4.13
C	5.0	3.87

Table 4. Average Human Review Scores by Prompt Variant and Intended Bloom Level

Prompt Variant	Intended_Bloom_Level	Clarity & Relevance to Concept	Subjective Cognitive Alignment
A	Analysis	5.0	5.0
A	Application	5.0	4.8
A	Knowledge	5.0	5.0
B	Analysis	5.0	5.0
B	Application	5.0	3.8
B	Knowledge	5.0	3.6
C	Analysis	5.0	3.8
C	Application	5.0	3.8
C	Knowledge	5.0	4.0

Universal Clarity and Relevance: A striking finding is the remarkable consistency across all prompt variants in achieving high scores for Clarity (average 5.0) and Relevance to Concept (average 5.0). This suggests that generative AI models, even with minimal explicit instructions, are highly capable of generating questions that are grammatically correct, understandable, and directly pertinent to the given topic. The challenge, therefore, is not in basic question formation, but in precise cognitive alignment.

Subjective Cognitive Alignment Mirrors Quantitative Trends: The human ratings for Subjective Cognitive Alignment directly correlated with the automated classification match rates.

Variant A (Baseline): Consistently achieved near-perfect subjective alignment (overall average 4.93, with 5.0 for Knowledge and Analysis), indicating that human experts also perceived these questions to be excellent fits for their intended Bloom's level.

Variant B (Simpler): Showed a noticeable decline in subjective alignment (overall average 4.13), particularly for 'Knowledge' (3.6) and 'Application' (3.8). This confirms that removing explicit Bloom's guidance leads to outputs that are perceived as less aligned with the specific cognitive objective, often overshooting it.

Variant C (Persona-Based): Had the lowest subjective alignment (overall average 3.87), reflecting substantial human-perceived misalignment, especially for 'Analysis' (3.8) and 'Application' (3.8). This reinforces the quantitative finding that the persona, while adding context, does not effectively communicate or reinforce precise cognitive level expectations for the AI.

The following examples provide concrete evidence of how different prompt variants influence question quality and cognitive alignment.

Excellent Example (Confirming the Effectiveness of Detailed Prompts):
Question number: Q010
Intended Level: Application
Prompt Variant: A_Baseline
Generated Question: Imagine you are a city planner tasked with optimizing traffic flow using Artificial Intelligence. Apply your understanding of different AI algorithms to construct a hypothetical system that uses real-time data from traffic cameras and GPS devices to predict and avoid traffic congestion. Describe how this system would operate, allocate resources like traffic signal timing, and demonstrate its potential benefits to the city's residents.
Classified Level: Application
Observation: This sophisticated question requires applying knowledge within a hypothetical scenario, embodying the 'Application' level. The explicit prompting with verbs like "Apply," "construct," "operate," and "demonstrate" ensures precise alignment, validated by both automated classification and human review.

Question Requiring Improvement (Demonstrating Misalignment):
Question number: Q004:
Intended Level: Knowledge
Prompt Variant: B_Simpler
Generated Question: Explain the key components that define Artificial Intelligence and differentiate it from human intelligence.
Classified Level: Comprehension
Human Scores: Clarity=5.0, Relevance=5.0, Subjective Alignment=3.0
Observation: Despite an Intended_Bloom_Level of 'Knowledge', the question generated uses verbs like "Explain" and "differentiate," which fall into 'Comprehension'. This highlights how the AI, given a simpler prompt, tends to infer a higher cognitive demand than intended, leading to misalignment.

4.3 Discussion and Interpretation

Our findings strongly underscore the critical role of explicit prompt engineering in achieving precise cognitive alignment for AI-generated educational content within tools like OneClickQuiz. This highlights a crucial insight: while all prompt variants consistently produced questions rated highly for clarity and relevance, their ability to match the Intended_Bloom_Level varied dramatically, indicating that content might be clear and relevant but not necessarily pedagogically aligned.

The superior performance of the A_Baseline prompt (your original, detailed prompt) in both automated classification and human subjective alignment is a key takeaway. This indicates that investing in carefully crafted prompts that explicitly define the target Bloom's level, include relevant descriptions, and suggest appropriate action verbs, effectively guides the AI to produce questions that meet specific pedagogical objectives. This reinforces findings from our previous work on accurate Bloom's/SOLO classification (Yaacoub, Da-Rugna, & Assaghir, 2025; Yaacoub, Assaghir, & Da-Rugna, 2025), and furthermore, demonstrates that precise question generation can be achieved through focused prompt engineering.

Conversely, the significant drop in performance for the B_Simpler prompt highlights the limitations of relying on the LLM to implicitly infer cognitive levels. When specific guidance is removed, the AI tends to generate questions that "overshoot" the intended Bloom's level, producing questions that are too complex (e.g., 'Comprehension' instead of 'Knowledge', or 'Synthesis' instead of 'Application'). This presents a challenge for "lightweight" approaches if the precise cognitive targeting is a primary goal. While simpler prompts are easier to write, they sacrifice control over nuanced pedagogical outcomes, leading to less reliable results in a learning analytics context.

The most counter-intuitive result stemmed from the C_PersonaBased prompt. Despite the intuitive appeal of guiding the AI with a persona ("seasoned computer science professor"), this variant performed the worst in terms of automated classification accuracy and had the lowest average subjective alignment. This suggests that while a persona might add context or style, it does not effectively convey the cognitive constraints needed for precise Bloom's alignment, especially at lower levels. The LLM might interpret the persona's "expertise" as a directive to generate more complex or sophisticated questions, potentially overriding the implicit or weakly conveyed Bloom's level intent. This highlights that simply adding a persona might introduce unintended biases or new objectives that deviate from the educational goal of cognitive alignment. The concept of "Lightweight Prompt Engineering" needs to be carefully defined: it values simplicity in prompt construction but should not sacrifice precision in instruction.

These findings have direct implications for the design of smart learning environments and technology-enhanced education. For OneClickQuiz and similar AI-driven content generation tools, the ability to generate taxonomically aligned questions is paramount for effective learning design and assessment. Educators need tools that not only automate but also ensure pedagogically sound content. Our study demonstrates

that even within a "lightweight" framework, the most effective approach for achieving cognitive alignment relies on explicit, detailed prompts rather than implicit or contextual cues that can lead to unintended cognitive shifts. This contributes directly to the "Cognitive Alignment" phase of the comprehensive AI-driven education framework previously proposed (Yaacoub, Tarnpradab, Khumprom, et al., 2025), emphasizing that robust initial content quality directly benefits downstream learning analytics and assessment.

5 Conclusion and Future Work

This study investigated the impact of lightweight prompt engineering strategies on the cognitive alignment and perceived quality of AI-generated questions within OneClickQuiz, a Moodle plugin leveraging generative AI. Our comparative analysis across three prompt variants—a detailed baseline, a simpler version, and a persona-based approach—yielded crucial insights into optimizing AI for educational content generation.

Our findings unequivocally demonstrate that explicit, detailed prompts (Variant A) are significantly more effective in guiding the AI to generate questions precisely aligned with intended Bloom's Taxonomy levels, as validated by both automated classification and human expert review. While all variants produced clear and relevant questions, the simpler (Variant B) and persona-based (Variant C) prompts consistently failed to achieve accurate cognitive alignment, often resulting in questions that overshot or deviated from the target Bloom's level. Counter-intuitively, the persona-based prompt performed the worst quantitatively, suggesting that contextual framing without explicit cognitive constraints can mislead the AI, leading to less precise pedagogical outcomes.

This research underscores the critical importance of strategic prompt engineering in the development of pedagogically sound AI-driven educational tools. It provides practical guidance for educators and developers utilizing generative AI in smart learning environments: precision and explicitness in prompts are paramount for achieving accurate cognitive alignment. This is a core component of effective learning design and enhances the utility of learning analytics by ensuring the underlying data (generated questions) is of high pedagogical quality.

This study, while providing valuable insights, has several limitations inherent to its "lightweight" approach. The experiment was conducted with a specific LLM (Gemini 2.0 Flash Lite) and a limited set of Bloom's Taxonomy levels and concepts in a single domain (Computer Science). The generalizability of these findings to other generative AI models (e.g., GPT-4, Claude, Llama) or to diverse subject domains beyond computer science remains an area for future investigation. The human review, while insightful, was performed by a small number of individuals and on a subset of questions, potentially limiting the overall breadth of qualitative insights. Furthermore, the effectiveness of a prompt can be LLM-specific, meaning optimal prompts for one model may not directly transfer to others without re-evaluation.

Future research will explore several directions. Firstly, we plan to investigate more advanced prompt engineering techniques, such as few-shot learning (providing multiple examples in the prompt) or chain-of-thought prompting, to determine if they can further enhance alignment, particularly for more complex cognitive levels. Secondly, we will conduct larger-scale human expert reviews, possibly involving multiple reviewers and diverse subject matter experts, to validate findings more broadly and collect richer qualitative data. Thirdly, we intend to test these prompt strategies with different LLMs and across various subject domains to assess the generalizability of our findings. Finally, we aim to integrate prompt optimization features directly into OneClickQuiz, allowing educators to select from pre-optimized prompt templates or providing visual feedback on predicted cognitive alignment to guide their prompt creation, thereby operationalizing the "Cognitive Alignment" phase of our comprehensive AI-driven education framework.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.

Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming Education: A Comprehensive Review of Generative Artificial Intelligence in Educational Settings through Bibliometric and Content Analysis. *Sustainability*, *15*(17), Article 17. https://doi.org/10.3390/su151712983

Bao, Y. (2024). A comprehensive investigation for ChatGPTs applications in education. *Applied and Computational Engineering*, *35*, 116–122. https://doi.org/10.54254/2755-2721/35/20230377

Biggs, J. B., & Collis, K. F. (2014). *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.

Bloom, B. S. (1984). *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain* (2nd edition Edition). Addison-Wesley Longman Ltd.

Bule, J., & Peer, P. (2014). Technical, Legal, Economic and Social Aspects of Biometrics for Cloud Computing. *Journal of Information and Organizational Sciences*, *38*(2), 83–95.

Doughty, J., Wan, Z., Bompelli, A., Qayum, J., Wang, T., Zhang, J., Zheng, Y., Doyle, A., Sridhar, P.,

Agarwal, A., Bogart, C., Keylor, E., Kultur, C., Savelka, J., & Sakr, M. (2024). A Comparative Study of AI-Generated (GPT-4) and Human-crafted MCQs in Programming Education. *Proceedings of the 26th Australasian Computing Education Conference*, 114–123. https://doi.org/10.1145/3636243.3636256

Grévisse, C. (2024). Comparative Quality Analysis of GPT-Based Multiple Choice Question Generation. In H. Florez & M. Leon (Eds.), *Applied Informatics* (pp. 435–447). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-46813-1_29

Jain, S. (2015). Automated generation of programming language quizzes. *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*, 1051–1053. https://doi.org/10.1145/2786805.2807559

Lavidas, K., Voulgari, I., Papadakis, S., Athanassopoulos, S., Anastasiou, A., Filippidi, A., Komis, V., & Karacapilidis, N. (2024). Determinants of Humanities and Social Sciences Students' Intentions to Use Artificial Intelligence Applications for Academic Purposes. *Information*, *15*(6), 314.

Mayer, R. E. (2009). *Multimedia Learning* (Second Edi). Cambridge University Press.

Olga, A., Tzirides, Saini, A., Zapata, G., Searsmith, D., Cope, B., Kalantzis, M., Castro, V., Kourkoulou, T., Jones, J., da Silva, R. A., Whiting, J., & Kastania, N. P. (2023). *Generative AI: Implications and Applications for Education* (No. arXiv:2305.07605). arXiv. https://doi.org/10.48550/arXiv.2305.07605

Park, J., & Choo, S. (2024). Generative AI Prompt Engineering for Educators: Practical Strategies. *Journal of Special Education Technology*, 01626434241298954. https://doi.org/10.1177/01626434241298954

Salcedo-Lagos, P., Pinacho-Davidson, P., Pinninghoff J., M. A., Contreras A., R., Fuentes-Riffo, K., & Carrillo, M. F. (2024). Optimizing Didactic Sequences with Artificial Intelligence: Integrating Bloom's Taxonomy and Emotion in the Selection of Educational Technologies. In J. M. Ferrández Vicente, M. Val Calvo, & H. Adeli (Eds.), *Artificial Intelligence for Neuroscience and Emotional Systems* (pp. 509–517). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-61140-7_48

Shaikh, S., Daudpotta, S. M., & Imran, A. S. (2021). Bloom's Learning Outcomes' Automatic Classification Using LSTM and Pretrained Word Embeddings. *IEEE Access*, *9*, 117887–117909. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3106443

Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). A Predictive Model to Evaluate Students' Cognitive Engagement in Online Learning. *Procedia - Social and Behavioral Sciences*, *116*, 4844–4853. https://doi.org/10.1016/j.sbspro.2014.01.1036

Steingartner, W., & Novitzká, V. (2015). A new approach to operational semantics by categories. In T. Hunjak, V. Kirinić, & M. Konecki (Eds.), *Proceedings of the 26th Central European Conference on Information and Intelligent System (CECIIS 2015)* (pp. 247–254). University of Zagreb, Faculty of Organization and Informatics Varaždin.

Trust, T., Whalen, J., & Mouza, C. (2023). Editorial: ChatGPT: Challenges, Opportunities, and Implications for Teacher Education. *Contemporary Issues in Technology and Teacher Education*, *23*(1), 1–23.

Yaacoub, A., Assaghir, Z., & Da-Rugna, J. (2025). Cognitive Depth Enhancement in AI-Driven Educational Tools via SOLO Taxonomy. In K. Daimi & A. Al Sadoon (Eds.), *Proceedings of the Third International Conference on Advances in Computing Research (ACR'25)* (Vol. 1346, pp. 14–25). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-87647-9_2

Yaacoub, A., Assaghir, Z., Prevost, L., & Da-Rugna, J. (2025). *Analyzing Feedback Mechanisms in AI-Generated MCQs: Insights into Readability, Lexical Properties, and Levels of Challenge*. arXiv. https://doi.org/10.48550/ARXIV.2504.21013

Yaacoub, A., Da-Rugna, J., & Assaghir, Z. (2025). *Assessing AI-Generated Questions' Alignment with Cognitive Frameworks in Educational Assessment* (No. arXiv:2504.14232). arXiv. https://doi.org/10.48550/arXiv.2504.14232

Yaacoub, A., Haidar, S., & Da Rugna, J. (2024). OneClickQuiz: Instant GEN AI-Driven Quiz Generation in Moodle. *Conference on Sustainable Energy Education – SEED 2024*, 689–698.

Yaacoub, A., Tarnpradab, S., Khumprom, P., Assaghir, Z., Prevost, L., & Da-Rugna, J. (2025). *Enhancing AI-Driven Education: Integrating Cognitive Frameworks, Linguistic Feedback Analysis, and Ethical Considerations for Improved Content Generation* (No. arXiv:2505.00339). arXiv. https://doi.org/10.48550/arXiv.2505.00339

Children's Screen Time, Preoccupation with Technology, Social Media Use, and Playing Video Games as Possible Predictors of GPA – a Longitudinal Study

Maja Kućar, Andreja Brajša-Žganec, Tamara Bolić
Institute of Social Sciences Ivo Pilar
Trg Marka Marulića 19, 10000 Zagreb, Croatia
{mkucar, andreja.brajsa-zganec, tamara.bolic}@pilar.hr

Abstract. *Prior research has frequently linked digital technology use to academic outcomes, often highlighting small but negative associations between screen time and students' grade point averages (GPA). The goal of this study is to examine the unique contribution of screen time, preoccupation with technology, playing video games, and social media use in predicting changes in students' GPA among elementary school children. A total of 970 mother-child dyads from Croatia completed questionnaires in 2021 and 2023 as part of the CHILD-WELL project funded by the Croatian Science Foundation. At baseline, children were on average 11.04 years old (SD = 1.15), and 53.2% were female. A hierarchical regression analysis revealed that screen time was a small, negative, while preoccupation with digital technology was a small, positive predictor of changes in GPA measured two years later. Age was negatively associated with GPA and female students had slightly higher GPA. Overall, predictors explained a very low amount of variance above the control variables, and no convincing evidence was found that various digital technology use measures were related to academic performance.*

Keywords. screen time, social media, video games, GPA, academic achievement

1 Introduction

Digital technology, particularly smartphones and social media platforms, has become an integral part of students' daily lives. As digital engagement has increased, so too has scholarly interest in its potential impact on academic outcomes. A growing body of research suggests that digital technology, especially social media and screen time is often negatively associated with academic performance, as measured by grade point average (GPA) and other achievement indicators. For instance, a recent study conducted with early adolescents in North America found that the frequency of social media use, including platforms such as Facebook, Twitter, and Instagram, was associated with lower academic achievement (Gordon & Ohannessian, 2024). Similarly, a meta-analysis by Huang (2018) identified a small but statistically significant negative correlation between social media use and academic achievement. Interestingly, this analysis also noted that Facebook use alone showed no significant relationship with academic outcomes. Liu et al. (2017) reached similar conclusions, reporting a negative association between social media use and GPA, with stronger effects observed among female and college-aged students. Beyond social media, screen time has also been related to academic outcomes. For example, studies have shown negative associations between GPA and smartphone screen time (Sapci et al., 2021), general screen exposure (Cerniglia et al., 2021), and video game use (Weaver et al., 2013). In a longitudinal study by Cerniglia et al. (2021), screen time at age four was positively linked with emotional dysregulation and negatively associated with mathematics and literacy achievement by age eight, underscoring potential long-term effects. Paulich et al. (2021) also reported a moderate-size negative effect of increased screen time on academic performance of 9- and 10-year-old children. However, not all findings point in the same direction. Kućar et al. (2023) reported that while Instagram screen time negatively predicted GPA, the frequency of Facebook checking was a positive predictor, suggesting that platform-specific usage patterns may matter.

Similarly, Ventura et al. (2012) found no overall GPA differences based on general video game use. However, they did observe that students who engaged

in moderate levels (11–50 hours per week) of playing their favorite video game performed better academically than low-usage peers, emphasizing the importance of qualitative factors such as gameplay motivation and engagement.

Some authors argue that other measures and concepts, beyond the time spent on a specific app, frequency of checking a device, or general screen time should be involved in digital technology research. In other words, questions such as 'how much' and 'how often' are not sufficient in this rapidly evolving research area. Some researchers began to ask 'why', 'in what way', or 'for what purpose'. For example, Marty-Dugas et al. (2018) focused on absent-minded smartphone use, Bischof-Kastner et al. (2014) on various motivations for using digital technology (e.g., emotional coping vs. for fun/leisure), while Yildirim and Correia (2015) explored nomophobia - the fear of being out of mobile phone contact. Some authors examined a contemporary phenomenon of distracted users of smartphones in public spaces. The term 'smartphone zombie' was coined and gained recognition and attention, one of the reasons being the fact that it can be dangerous to engage in such behavior in traffic and crowded places (Pressey et al., 2024). Although such constructs can be very useful in research on adult population, they cannot be directly translated to a younger population, such as elementary school-aged children. While much research has focused on pathological digital media use, many children engage with digital technology in ways that reflect strong attachment rather than addiction. This non-pathological preoccupation, driven by interest and positive emotions toward devices and applications, may still influence their behavior and well-being. However, the relationship between such engagement and well-being remains underexplored, particularly in children, despite its potential significance through various pathways. Kotrla Topić et al. (2023) proposed a term preoccupation with digital technology, purposefully constructed for younger population in order to grasp their attachment to digital technology. Although the research between this construct and educational outcomes is still exploratory, authors state that a stronger preoccupation with digital technology may lead to less time for other activities, such as homework and studying. As originally defined by Kotrla Topić et al. (2023), preoccupation with digital technology reflects the cognitive and emotional importance that children attribute to digital technologies, rather than the amount of time spent or symptoms of problematic use. The construct is inherently non-pathological and was developed to capture the positive affect, personal relevance, and engagement children may experience in relation to digital tools. While the authors acknowledge that high preoccupation could reduce time for other activities (e.g., studying), they also emphasize that such digital involvement may reflect a normal and adaptive aspect of children's lives. This ambivalence is central to the construct, it may relate to well-being or academic outcomes through both constructive and disruptive pathways, depending on the broader context of use.

The concept of preoccupation with digital technology may be understood in light of emerging discussions around attachment towards digital objects. Preoccupation with digital technology can be understood as a form of emotional and cognitive attachment to digital tools, particularly relevant in contemporary childhood. This aligns with emerging extensions of attachment theory (Bowlby et al. 1992) into digital contexts, where children may form bonds with devices that offer predictability, responsiveness, and gratification (Koles & Nagy, 2021). From this perspective, digital objects can serve as emotional anchors or self-regulation aids, especially in a developmental stage where autonomy and exploration are prominent. Simultaneously, preoccupation with digital technology may reflect a child's familiarity and competence with digital environments, including their ability to navigate, manage, and utilize technology for both social and academic purposes. Research on digital literacy suggests that such skills are increasingly essential for academic performance and self-directed learning (Ng, 2012; Hatlevik et al., 2018). Thus, rather than signifying distraction or dysfunction, higher preoccupation may signal a child's adaptive engagement with digital tools that are increasingly embedded in educational contexts. As highlighted in recent work, this field is still developing, with few studies offering comprehensive frameworks to capture the psychological and behavioral complexities of users' relationships with digital media. Drawing from interdisciplinary perspectives, including psychology, sociology, and consumer behavior, research on digital attachment emphasizes the evolving and often deeply personal connection individuals form with digital objects. This aligns with the notion of preoccupation, which may reflect not just frequent use, but also emotional investment and cognitive engagement (Koles & Nagy, 2021).

Despite the extensive research linking digital technology use with academic outcomes, a notable gap persists: few studies have focused specifically on

children. Much of the current literature centers on adolescents, university students, or adults (e.g., Huang, 2018; Liu et al., 2017; Sapci et al., 2021), leaving open questions about how these findings generalize to younger populations. This underrepresentation limits our understanding of how early digital habits may influence academic trajectories over time.

1.1 Research goals

The goal of this study is to examine the unique contribution of screen time, preoccupation with digital technology, playing video games, and social media use in predicting changes in students' GPA measured two years later, above and beyond the control variables (GPA in previous time point, age, and gender).

2 Method

2.1 Participants

The sample consists of 970 mother-child dyads who completed the questionnaires in both research waves (2021 and 2023). The average age of children at baseline was M_{age}=11.04 (SD=1.15) and 53.2% of the sample was female. At the first time point, children attended third (N=246, 25.2%), fourth (N=264, 27%), fifth (N=250, 25.6%) and sixth (N=217, 22.2%) grade of elementary schools in two counties in Croatia (Osijek-Baranja and Varaždin county).

2.2 Instruments

2.2.1 Preoccupation with Digital Technology

Preoccupation with digital technology was assessed using the Preoccupation with Digital Technology Scale (Kotrla Topić et al., 2023). The scale consists of five items, each rated on a four-point Likert scale, where lower scores indicate lower preoccupation and higher scores indicate greater preoccupation with digital technology. The scale was designed as a brief measure to capture children's affective attachment to digital technology and the importance they attribute to it. Items are intended to reflect a single underlying construct representing the perceived significance of digital technology in the child's life, rather than differentiating between specific devices or activities. The total score is computed as the average of the five items, with higher scores reflecting higher levels of preoccupation. The items include: (1) "How important are devices (smartphone, laptop, tablet, and computer) to you in life?"; (2) "How bothered would you be if you were not allowed to use these devices?"; (3) "How happy does it make you to receive a new device, game, application, or game addition?"; (4) "Do you enjoy spending time with your smartphone, tablet, or computer?"; and (5) "Spending time with my smartphone, tablet, or computer is…" (rated from 1 = Very boring to 4 = Very interesting). Reliability analysis revealed good reliability (Cronbach α = .80).

2.2.2 Screen Time

Screen time was estimated by mothers, who reported the average number of hours their child spent using digital technology on weekdays and weekends separately (open-ended format). Mothers were asked: 'During a typical workday and during the weekend, how much time does your child spend on the following activity (please indicate the average number of hours per day) – using digital devices'. The two values were averaged to create a single screen time estimate. Responses indicating more than 12 hours of screen time per day were considered implausible and excluded from the analysis.

2.2.3 Social Media Use

Children self-reported their social media use by responding to a single item: "How often do you use social media like Facebook, Instagram, Snapchat, TikTok, etc.?" Responses were recorded on a four-point Likert scale (1 = Almost never, 2 = Rarely, 3 = A few times a week, 4 = Almost every day).

2.2.4 Playing Video Games

Children also self-reported the frequency of playing video games using a single item: "How often do you play video games on your laptop, tablet, smartphone, or computer?" Responses were provided on the same four-point Likert scale as above (1 = Almost never, 2 = Rarely, 3 = A few times a week, 4 = Almost every day).

2.2.5 Grade Point Average (GPA)

GPA was defined as the average of final grades across all subjects at the end of the last academic year. This information was obtained from school administrators or class teachers. Grades for each subject are between 1 (failed) to 5 (excellent) and the instruction was to provide a number with two decimal places (e.g., '4.45').

2.3 Procedure

The data used in this study are part of a broader research project called CHILD-WELL, focused on

child well-being in the family context, funded by the Croatian Science Foundation (project number HRZZ-IP-2019-04-6198). Prior to data collection, ethical approval was obtained from the institutional ethics committee and the Croatian Ministry of Science and Education. Written informed consent was obtained from parents of elementary school students across 15 participating schools. In addition, children provided their own consent before completing the questionnaires. The child questionnaire was administered in a group setting during regular class time and took approximately 45 minutes to complete. Mothers received their questionnaire via their children and were instructed to return the completed forms in sealed envelopes back to school. To ensure anonymity while enabling data matching, both children and parents generated a unique, anonymous code at the beginning of the questionnaire.

3 Results

Descriptive statistics (Table 1) show that the average GPA slightly decreased from time 1 (M = 4.78) to time 2 (M = 4.58). Participants reported moderate levels of preoccupation with digital technology, social media use and playing video games. The average reported screen time was 2.85, which in turn translates to 2 hours and 51 minutes. The sample had a mean age of 11.04 years, with a balanced gender distribution.

Table 1. Descriptive statistics

Variable	Min	Max	Mean	SD
GPA t1	2.78	5	4.79	0.35
GPA t2	2.67	5	4.59	0.49
PwDT t1	1	4	2.88	0.57
SM use t1	1	4	1.98	0.83
Screen time t1	0.3	12	2.85	1.51
VG t1	1	4	3.27	0.88
Age	8	14	11.02	1.14
Gender	1 male	2 female		

Note: GPA – grade point average, PwDT – preoccupation with digital technology, SM – social media, VG – video game

The correlation matrix (Table 1 in the Appendix) reveals that GPA at time 2 is strongly associated with GPA at time 1. Higher preoccupation with digital technology was positively correlated with greater social media use, screen time, and video game use, but negatively with gender (indicating higher in boys). GPA in t1 and t2 was negatively associated with screen time, social media use, and age, indicating that older students have lower GPA. GPA t1 showed a small, positive correlation with video games, but not in t2. In both time points, GPA showed a correlation with gender, indicating that female students have somewhat higher GPA.

Regression analysis was conducted in order to examine the unique contribution of each predictor to GPA at time 2, while controlling for the effects of other variables. Regression analysis predicting GPA at time 2 showed that prior GPA was the strongest predictor across all steps, explaining 61% of the variance. In the second step, age and gender were added. Age negatively predicted GPA, while gender had a positive effect, meaning that female students have somewhat higher GPA compared to male. In the final model (third step), preoccupation with digital technology positively predicted GPA, whereas screen time was a negative predictor of GPA. GPA t1, age, and gender remained significant predictors. Social media use and video game use did not significantly contribute to the model in the final step.

Table 2. Regression analysis

	Outcome: GPA t2		
	Step 1	Step 2	Step 3
Predictor	β	β	β
GPA t1	.78**	.76**	.75**
Age		-.08**	-.07**
Gender		.07**	.07**
PwDT t1			.08**
SM Use t1			-.04
Screen Time t1			-.07**
Video games t1			-.03
ΔF	1297.59	10.97	4.22
R^2	.61	.62	.63
ΔR^2		.01	.01

Note: GPA – grade point average, PwDT – preoccupation with digital technology, SM – social media

It is important to note that adding both control (age, gender) and digital technology variables only slightly increased the explained variance of the outcome. More precisely, after controlling for the outcome variable from a previous time point, control variables (age, gender) additionally explained only 1% of the variance, as well as adding four predictors related to digital technology use.

Due to the non-normal distribution of GPA scores, specifically, a large number of participants with a perfect GPA (approximately 50%), additional analyses were conducted using both logistic regression and rank-transformed regression. For the logistic

regression, GPA was dichotomized into "all excellent" vs. "not excellent," while in the rank-transformed regression, GPA scores were converted to ranks, allowing for a nonparametric estimation that reduces the impact of skewed distributions. These alternative models largely replicated the core findings from the original linear regression, though some differences emerged (Appendix, Table 2 and 3). For example, preoccupation with digital technology remained a small but significant positive predictor in the rank transformed model, but not in the logistic regression. Screen time remained a small, negative predictor in both alternative models, while using social media emerged as a small, negative predictor in both models. In all three models, predictors related to digital behavior explained a small amount of variance.

4 Discussion

The present study examined the predictive value of digital technology-related behaviors, specifically, screen time, preoccupation with digital technology, social media use, and video game use on changes in academic achievement (GPA) over a two-year period in a large sample of elementary school children. Importantly, the study addressed a gap in literature by focusing on a younger, understudied population, as most prior research has centered on adolescents and adults.

Consistent with some previous findings (Liu et al., 2017; Huang, 2018), screen time and social media use were negatively correlated with academic achievement. However, in the multivariate model, only screen time remained a significant negative predictor of GPA two years later, suggesting that overall exposure to screens may be more influential than specific digital activities such as social media use or gaming. Higher screen time can also mean less time for other activities such as sports or education, which can lead to negative changes in students' GPA. Faught et al. (2019) reported that adherence to guidelines on limiting excessive screen use is associated with higher academic achievement in a sample of Canadian preadolescents and adolescents. However, the literature on the effects of screen time on cognitive functioning and related outcomes remains limited and has faced growing criticism in recent years. Researchers have pointed out that the reported effect sizes in published studies are often very small, there is potential evidence of publication bias, and overall study quality in this field has been described as relatively low (Ophir et al., 2021; Orben et al., 2020; Wilmer et al., 2017). Therefore, screen time should perhaps be viewed as a rough proxy measure that captures a variety of underlying behaviors and lifestyle factors, rather than a direct causal influence on academic outcomes. Future research should aim to disentangle these complex relationships by using more nuanced and context-specific measures of digital technology use.

Interestingly, preoccupation with digital technology emerged as a small, but significant, positive predictor of later GPA, which is in line with Kotrla Topić et al. (2023), who stated that the preoccupation with digital technology scale was constructed as a counterbalance to constructs focusing solely on addiction or problematic use. Preoccupation with digital technology is not necessarily negative and may reflect a child's positive, meaningful engagement with digital environments. One possible explanation is that preoccupation reflects higher digital literacy. Children who are more engaged with technology may develop stronger skills in navigating digital tools, finding and evaluating information, and solving problems in online environments. These skills have become increasingly essential for completing schoolwork, especially as digital platforms become integrated into classroom and homework routines. Furthermore, preoccupation may reflect a form of attachment to digital objects, where digital devices provide both emotional gratification and a secure base for cognitive exploration. Drawing from attachment theory (Bowlby et al., 1992), if the attachment is not excessive or displacing real-world relationships, it can promote autonomy, self-regulation, and persistence. Children who feel emotionally connected to digital tools may also feel more competent and confident using them for academic tasks, particularly in environments where digital competence is essential. Finally, some aspects of the PwDT scale (e.g., enjoyment when receiving new digital devices) may indirectly capture socioeconomic conditions and which are known to be related to academic success. Children from more resourceful households are more likely to have regular access to such devices. These factors contribute both to higher PwDT scores and to more favorable academic conditions, suggesting that preoccupation with digital technology may, in part, reflect underlying digital capital and socioeconomic advantages. The absence of a significant effect of video game use on later GPA is in line with several mixed findings in the literature (Ventura et al., 2012; Weaver et al., 2013), indicating that the impact of gaming may be highly context-

dependent, influenced by content, duration, and the social or cognitive nature of the activity.

A major strength of this study lies in its longitudinal design and a large sample of mother-child dyads. The study used a multi-informant approach (parent- and child-reported data) and objective GPA records obtained from schools, increasing the reliability of the findings. The inclusion of both general and activity-specific digital technology indicators allowed for a nuanced understanding of how different digital behaviors uniquely relate to academic outcomes. Additionally, a strength of this research is its focus on children in middle childhood/early adolescence. It has been previously identified that studies are typically focused on adolescent/adult population or period of early childhood (Kućar & Šimleša, 2022).

Several limitations must be acknowledged. First, although the longitudinal design enables stronger causal inference than cross-sectional studies, the observational nature of the data precludes the conclusions about causality. Second, some measures relied on single-item self-reports (e.g. video games), which may reduce reliability and fail to capture the complexity of such behaviors. A limitation of this study is the fact that the GPA is strongly skewed towards higher values – around 50% of the children in this study had a GPA of 5.0, meaning they were graded with a maximum grade in all subjects. That is not an odd phenomenon in Croatia (Net.hr, 2024). This grade inflation trend, increasingly common in Croatian primary schools, raises concerns about the validity of GPA as a performance measure, as it may obscure meaningful differences in academic achievement. Moreover, higher-achieving students who are more likely to have supportive, well-functioning family environments are also disproportionately more likely to participate in research and remain engaged in longitudinal studies. This introduces a potential selection bias, whereby the sample may not fully represent the broader population of students, particularly those with lower academic performance or from less advantaged backgrounds. Future studies should include National or State exams as measures of academic achievement. Those types of tests are standardized and equal for all students in a certain country and are therefore more objective. Additionally, while the use of a multi-informant approach is generally a research strength, mothers' estimations of children's screen time are possibly underestimated due to social desirability. Moreover, mothers do not have full insight into all the contexts in which screen use occurs, such as time spent with peers, at school, or in other unsupervised settings. This partial visibility may affect the precision, accuracy and construct validity of the screen time variable and should be considered when interpreting the findings.

Lastly, the explained variance added by digital technology variables was small, highlighting the predominant role of prior academic achievement and suggesting that other contextual or individual factors likely play a more substantial role in academic development. Although two predictors reached statistical significance, effect sizes were small. Specifically, each additional hour of screen time per day at baseline was associated with a 0.07-point decrease in GPA one year later, and a one-point increase on the Preoccupation with digital technology (PwDT) scale predicted a 0.08-point increase in GPA. Age and gender also showed small effects, with older students and males tending to have slightly lower GPA scores. The low explained variance suggests that digital technology usage does not meaningfully account for differences in academic performance. Therefore, even though there is statistical significance, the practical implications about the relationship between digital technologies and GPA are limited.

Future research should investigate potential moderating or mediating mechanisms such as sleep quality, executive functioning, or parental monitoring—that may help explain the complex relationship between digital technology use and academic achievement. Exploring content-specific digital engagement (e.g., educational vs. entertainment use) and the quality of technology interactions would also provide more actionable insights. Ultimately, the findings underscore the importance of balanced digital media use and call for more targeted, developmentally appropriate guidelines for children's engagement with technology.

5 Conclusion

The findings of this study suggest that while certain aspects of digital technology use, such as overall screen time and preoccupation with digital technology, show small associations with academic performance (screen time – negative, preoccupation – positive), their practical impact on GPA appears minimal in elementary school children. Given the low amount of explained variance (1%) above control variables, digital technology use alone may not be a meaningful predictor of academic change during this developmental stage, highlighting the need to consider

broader contextual and individual factors in future research. Each additional hour of screen time at baseline was associated with a 0.07-point decrease in GPA one year later. The findings suggest that not all digital engagement is harmful, and in some cases, preoccupation with digital technology may support academic success. A one-point increase on the Preoccupation with digital technology (PwDT) scale predicted a 0.08-point increase in GPA. For parents and educators, this highlights the importance of monitoring the purpose and context of digital use. Encouraging students to use technology for goal-oriented tasks, information seeking, or organizational support may be more productive than imposing blanket limits. Still, the effect sizes are very modest, and digital behaviors alone should not be overemphasized as drivers of academic outcomes.

Acknowledgements

This paper draws on data collected as part of the CHILD-WELL project (*Child Well-Being in Family Context*), fully supported by the Croatian Science Foundation (HRZZ-IP-2019-04-6198). The study was developed in the context of the TEEN-WELL project (*Adolescents' Well-Being in Key Contexts*), which extends the scope of the original project to adolescent well-being. The TEEN-WELL project is financed by the European Union – NextGenerationEU and is carried out by the Ivo Pilar Institute of Social Sciences, Zagreb, Croatia, as part of the Institute's program funding for the period 1/1/2024–12/31/2027.

References

Bischof-Kastner, C., Kuntsche, E., & Wolstein, J. (2014). Identifying problematic Internet users: development and validation of the Internet Motive Questionnaire for Adolescents (IMQ-A). *Journal of medical Internet research, 16*(10), e230. https://doi.org/10.2196/jmir.3398

Bowlby, J., Ainsworth, M., & Bretherton, I. (1992). The origins of attachment theory. *Developmental Psychology, 28*(5), 759-775. https://doi.org/10.1037/0012-1649.28.5.759

Cerniglia, L., Cimino, S., & Ammaniti, M. (2021). What are the effects of screen time on emotion regulation and academic achievements? A three-wave longitudinal study on children from 4 to 8 years of age. *Journal of Early Childhood Research, 19(*2), 145-160. https://doi.org/10.1177/1476718X2096984

Faught, E. L., Qian, W., Carson, V. L., Storey, K. E., Faulkner, G., Veugelers, P. J., & Leatherdale, S. T. (2019). The longitudinal impact of diet, physical activity, sleep, and screen time on Canadian adolescents' academic achievement: An analysis from the COMPASS study. *Preventive medicine, 125,* 24-31. https://doi.org/10.1016/j.ypmed.2019.05.007

Gordon, M. S., & Ohannessian, C. M. (2024). Social media use and early adolescents' academic achievement: Variations by parent-adolescent communication and gender. *Youth & Society, 56*(4), 651-672. https://doi.org/10.1177/0044118X23118031

Hatlevik, O. E., Throndsen, I., Loi, M., & Gudmundsdottir, G. B. (2018). Students' ICT self-efficacy and computer and information literacy: Determinants and relationships. *Computers & Education, 118*, 107-119. https://doi.org/10.1016/j.compedu.2017.11.011

Huang, C. (2018). Social network site use and academic achievement: A meta-analysis. *Computers & Education, 119*, 76-83. https://doi.org/10.1016/j.compedu.2017.12.010

Koles, B., & Nagy, P. (2021). Digital object attachment. *Current Opinion in Psychology, 39*, 60-65. https://doi.org/10.1016/j.copsyc.2020.07.017

Kućar, M., Hainš, V. V., & Kovačić, R. (2023, May). Examining the Relationship Between Students' Social Media Usage, Smartphone Checking During Lessons/Studying, and Academic Achievement (GPA). In *2023 46th MIPRO ICT and Electronics Convention (MIPRO)* (pp. 728-733). IEEE. https://doi.org/10.23919/MIPRO57284.2023.10159991

Kućar, M., & Šimleša, S. (2022). Povezanost vremena provedenog pred zaslonom digitalnih uređaja i izvršnih funkcija-pregled literature. *Radovi Zavoda za znanstveni rad Varaždin*, (33), 245-277. https://doi.org/10.21857/y54jofvrxm

Liu, D., Kirschner, P. A., & Karpinski, A. C. (2017). A meta-analysis of the relationship of academic performance and Social Network Site use among adolescents and young adults. *Computers in Human Behavior*, *77,* 148–157. https://doi.org/10.1016/j.chb.2017.08.039

Marty-Dugas, J., Ralph, B. C., Oakman, J. M., & Smilek, D. (2018). The relation between smartphone use and everyday inattention. Psychology of Consciousness: Theory, *Research, and Practice, 5*(1), 46. https://doi.org/10.1037/cns0000131

Ng, W. (2012). Can we teach digital natives digital literacy?. *Computers & education, 59*(3), 1065-

1078. https://doi.org/10.1016/j.compedu.2012.04.016

Net.hr. (2024, September 11*). Imamo li previše odlikaša? Iz škola ih svake godine izađe 47 posto, neki tvrde – nije realno*. https://net.hr/danas/rtl-danas/imamo-li-previse-odlikasa-iz-skola-ih-svake-godine-izade-47-posto-neki-tvrde-nije-realno-3d9961e2-7065-11ef-be36-225a3ed1b99c

Ophir, Y., Rosenberg, H., & Tikochinski, R. (2021). What are the psychological impacts of children's screen use? A critical review and meta-analysis of the literature underlying the World Health Organization guidelines. *Computers in Human Behavior*, 124, 106925. https://doi.org/10.1016/j.chb.2021.106925

Orben, A., Weinstein, N., & Przybylski, A. K. (2020). Only holistic and iterative change will fix digital technology research. *Psychological Inquiry, 31*(3), 235-241. https://doi.org/10.1080/1047840x.2020.1820221

Paulich, K. N., Ross, J. M., Lessem, J. M., & Hewitt, J. K. (2021). Screen time and early adolescent mental health, academic, and social outcomes in 9-and 10-year old children: Utilizing the Adolescent Brain Cognitive Development℠(ABCD) *Study. PloS one, 16*(9), e0256591. https://doi.org/10.1371/journal.pone.0256591

Pressey, A., Houghton, D., & Istanbulluoglu, D. (2024). The problematic use of smartphones in public: the development and validation of a measure of smartphone "zombie" behaviour. Information *Technology & People, 37*(1), 479-501. https://doi.org/10.1108/itp-06-2022-0472

Sapci, O., Elhai, J. D., Amialchuk, A., & Montag, C. (2021). The relationship between smartphone use and students` academic performance. *Learning and Individual Differences, 89,* 102035. https://doi.org/10.1016/j.lindif.2021.102035

Topić, M. K., Brkljačić, T., & Brajša-Žganec, A. (2023). Digital technology use and preoccupation with digital technology as predictors of life satisfaction in children. *Child & family social work, 28*(4), 1131-1138. https://doi.org/10.1111/cfs.13032

Ventura, M., Shute, V., & Kim, Y. J. (2012). Video gameplay, personality and academic performance. *Computers & Education, 58*(4), 1260-1266. https://doi.org/10.1016/j.compedu.2011.11.022

Weaver, J., Kim, P., Metzer, R. L., & Szendrey, J. M. (2013). The impact of video games on student GPA, study habits, and time management skills: What's the big deal. *Issues in Information Systems, 14*(1), 122-128. https://doi.org/10.48009/1_iis_2013_122-128

Wilmer, H. H., Sherman, L. E., & Chein, J. M. (2017). Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning. *Frontiers in psychology, 8*, 605. https://doi.org/10.3389/fpsyg.2017.00605

Yildirim, C., & Correia, A. P. (2015). Exploring the dimensions of nomophobia: Development and validation of a self-reported questionnaire. *Computers in human behavior, 49*, 130-137. https://doi.org/10.1016/j.chb.2015.02.059

Appendix

Table 1. Correlation matrix

	GPA t1	GPA t2	PwDT t1	SM use t1	ST t1	VG t1	Age	Gender
GPA t1	1							
GPA t2	.79**	1						
PwDT t1	-0.02	-0.004	1					
SM use t1	-.19**	-.18**	.31**	1				
ST t1	-.16**	-.19**	.23**	.16**	1			
VG t1	.07*	.02	.43**	.16**	.11**	1		
Age	-.23**	-.26**	.18**	.15**	.22**	-.01	1	
Gender	.14**	.18**	-.10**	.03	-.08*	-.19**	-.03	1

Note: GPA – grade point average, PwDT – preoccupation with digital technology, SM – social media, VG – video games, ST – screen time

Table 2. Regression analysis with rank-transformed GPA

	Outcome: GPA t2 - rank		
	Step 1	Step 2	Step 3
Predictor	β	β	β
GPA t1 - rank	.74**	.71**	.70**
Age		-.08**	-.08**
Gender		.08**	.08**
PwDT t1			.10**
SM Use t1			-.05*
Screen Time t1			-.06*
Video games t1			-.02
ΔF	1004.47	11.28	4.57
R2	.55	.56	.57
ΔR2		.01	.01

Note: GPA – grade point average, PwDT – preoccupation with digital technology, SM – social media

Table 3. Regression analysis with GPA as a dichotomous variable ("all excellent" vs. "not excellent")

Predictor	B	Sig.	Exp(B)
GPA t1 (dichotomous)	2.24	.00	9.37
Age	-0.18	.03	0.84
Gender	0.55	.00	1.73
PwDT t1	0.30	.09	1.35
SM Use t1	-0.27	.02	0.77
ST t1	-0.20	.00	0.82
VG t1	0.08	.45	1.08
Constant	0.64	.49	1.90
Model Chi-square (df)		240.99 (7), p < .001	
Nagelkerke R^2		.32	
Cox & Snell R^2		.23	
Hosmer-Lemeshow Test (p)		.19	
Overall Classification Accuracy		73.8%	

Note: GPA – grade point average, PwDT – preoccupation with digital technology, SM – social media, ST – screen time

The AI Classroom: Systematic Literature Insights into Quality and Access in Education

Caitlin Braga, Adriana Aletta Steyn

Department of Informatics, University of Pretoria, 83 Lynnwood Rd,

Hartebeestpoort 362-Jr, Pretoria, 0083, South Africa0028, South Africa

u19015722@tuks.co.za, riana.steyn@up.ac.za

Abstract. *A significant gap in access to quality education persists between developed and developing countries. Artificial Intelligence (AI) offers the potential to bridge this divide by expanding access to quality educational resources. This systematic literature review (SLR) examines 40 articles sourced from eight academic databases. The review identifies key competencies necessary for the effective adoption of AI in higher education institutions (HEIs) in developing countries and explores the benefits of AI-driven education. The analysis is structured around the Technology-Organisation-Environment (TOE) framework, which highlights the interplay among technological, organisational, and environmental factors in the successful integration of AI. The study also emphasises the importance of policy frameworks and guidelines to ensure the responsible use of AI and enhance the experiences of both students and educators. By synthesising current research, this systematic literature review (SLR) provides actionable insights for higher education institutions (HEIs) and policymakers, supporting strategies to reduce the digital divide and enhance educational outcomes through the adoption of AI.*

Keywords. AI adoption, AI education, AI integration, developing countries, educational programs, policy framework, systematic literature review, tertiary institutions

1 Introduction

The world is currently witnessing the transformative power of digital technology in providing access to quality education (Mannuru et al., 2023; Steyn & Eybers, 2023). Access is granted through digitisation, ultimately enabling societal and economic growth (Mathrani et al., 2023). By increasing access to education, we simultaneously invest in human capital and economic development in a country (de Souza Zanirato Maia et al., 2023).

Artificial Intelligence is a series of computational systems that can mimic human decision-making processes (Zawacki-Richter et al., 2019). These machines think like us and for us, providing the opportunity to streamline the mundane tasks that fill our lives.

Tertiary Institutes use these intelligent technologies to promote collaboration and distance learning, enabling studying to take place whenever and wherever (Velinov et al., 2021). AI in higher education is currently empowering students by tailoring the learning experience to suit their individual needs best, promoting customisable schooling (Labadze et al., 2023). These techniques contribute to achieving Sustainable Development Goals (SDGs) by ensuring inclusive and equitable quality education for all (Opesemowo & Adekomaya, 2024).

However, despite the depth of knowledge on digital technologies used in education, we are still presented with an existing divide within the Global South due to the lack of knowledge and skills regarding the capabilities of AI in Higher Educational Institutes (HEIs) (Mathrani et al., 2022). With the rapid adoption of AI worldwide, rural communities continue to lack proficiency due to limited access and biases (Sekwatlakwatla, 2023). The benefits of these technologies are not equally well understood (Mannuru et al., 2023), highlighting the need for investments and support to ensure that all students, regardless of their circumstances, benefit from these technologies before we exacerbate the digital divide.

This Systematic Literature Review (SLR) aims to take inspiration from the advancements achieved using these technologies in HEIs in the Global North (Aly, 2022), providing a blueprint for the necessary capabilities to achieve similar results in the Global South. The primary objective of this study is to highlight the requirements for adopting Artificial Intelligence (AI) within Higher Educational Institutions (HEIs) and how it can promote accessible, high-quality education in developing countries, sparking a wave of hope and optimism for a future of global educational transformation. This SLR identifies the obstacles that need to be overcome in developing nations to enable this change. This paper aims to answer the following research questions:

- RQ1: What competencies are essential for effective AI adoption in higher education?
- RQ2: What are the benefits of using AI in the Global South's HEIs that lead to access to quality education?

This SLR aims to expand the body of knowledge by providing insights for Higher Education Institutions (HEIs) to leverage the transformative potential of AI in delivering accessible, high-quality education in developing countries. The study is structured as follows: the second section provides a detailed background on the topic of AI used in educational settings, followed by an explanation of the research methodology employed in this systematic literature review (SLR). This is followed by the research findings, after which a detailed discussion of the findings obtained is presented. Section six then presents a conceptual contribution analysed using the Technology–Organisation–Environment (TOE) framework. Finally, this paper concludes with an outline of the key findings identified and their implications, with recommendations for future research.

2. Background

2.1 Application of AI in HEIs

Artificial Intelligence (AI) promotes education in many ways. One example of the use of AI in HEIs is personalised learning and feedback enforced by adaptive systems (Zawacki-Richter et al., 2019). These adaptive and intelligent systems customise educational content according to the learner's preferences (Maier, 2022). Additionally, these systems automate previously manual tasks, allowing educators to spend more time with their students and providing each individual with more focused attention (Baykasoğlu et al., 2018).

Telepresence Robotics is an additional application of AI that enhances education. This technology supports remote students by providing access to videoconferencing and locomotion features (Lei et al., 2022). Teachers are provided with the necessary support to connect with their students (Leoste et al., 2022). The term "Telepresence" gained popularity during the COVID-19 pandemic (Velinov et al., 2021). Language and grammar are simplified using Generative Artificial Intelligence (GAI), which has had a significant impact in the field of education (Mannuru et al., 2023).

GAI refers to a form of AI that generates content based on a provided prompt (Ming Liu, 2023). These services are available on platforms like ChatGPT and CoPilot. Research has proven that this technology improves language writing performance by offering automated suggestions to enhance word choice, sentence structure, and the overall coherence of the text (Maier, 2022).

These incredible AI technologies have not and will not stop there. We have already seen our very first AI Teacher, a new concept of 'machine teachers' that references non-human teachers, which are expected to play diverse and dynamic roles in education. Jill Watson is the first-ever AI teacher and was introduced to an online Knowledge-Based Artificial Intelligence (KBAI) class. Jill was designed to answer all students' questions within online forums (Kim et al., 2020).

2.2 Challenges and Opportunities of AI Education in Developing Countries

AI has promoted growth in developing countries like South Africa. It is currently being used to address accessibility issues while enhancing the relationships between students and educators (Treve, 2021) by creating an environment that fosters critical thinking, removes language barriers, and more. This highlights the potential of AI intelligence to solve issues in the field of education (Khan et al., 2022), leading to efficient, adaptable, and effective learning systems.

Currently, many students and educators in developing communities struggle with Adoption Anxiety due to the lack of digital skills required to use and understand these technologies (Jatileni et al., 2024). "AI Readiness" refers to the level of understanding and competencies required by individuals to critically evaluate AI technologies (Long & Magerko, 2020) and make informed decisions. In the context of education and training, AI Readiness is not only about learning AI but also about preparing various fields to understand and leverage its capabilities. Experts have noted that the lack of AI proficiency among students and instructors hinders the effective integration of AI into the education sector (Zhai et al., 2021).

AI enhances information sharing; however, it is crucial to address data privacy concerns and algorithmic bias in the context of a developing country (Popenici & Kerr, 2017). These concerns should be addressed with an AI policy framework which provides risk mitigation and opportunities for the use of AI (Chan, 2023).

In developing countries, existing structural issues exist due to unequal access to digital media and supporting services. Notably, female students have less access to digital media, highlighting the digital divide. It is crucial to inform policymakers about the need to plan initiatives that bridge the digital divide and establish equitable, gender-sensitive learning policies (Mathrani et al., 2022).

To ensure a successful and responsible implementation of AI in academic settings, careful consideration and systematic preparation are essential. Ethics in Artificial Intelligence (AI) encompasses privacy protection and bias mitigation, ensuring the ethical and responsible use and development of AI technologies (Hermansyah et al., 2023). AI policies in higher education should address the potential risks and opportunities associated with AI technologies (Chan, 2023).

3 Methodology

By identifying competencies, we aim to empower developing countries to stay ahead in the field of AI Education and learn from the education sector in Developed This study utilises a Systematic Literature Review (SLR) to thoroughly identify and analyse core competencies for effective AI adoption in higher education. Countries. The process follows Dickson's (2023) protocol with an emphasis on transparency and replicability at every stage. The primary objective is to identify both the competencies required for AI readiness and to explore the benefits within higher education institutions.

Comprehensive searches were conducted across the following databases: SpringerLink, ScienceDirect, Sage Publications, Web of Science, ProQuest, and Wiley. The search combined terms related to higher education (e.g., universities, colleges) AND artificial intelligence (e.g., AI adoption, AI tools) AND academic programs or teaching methodologies.

The inclusion criteria included in this SLR:

- Relevance: The study should prioritise AI Education by identifying necessary competencies for adoption in higher educational institutes.
- Publication Date: Due to the rapid evolution of AI and to ensure the accuracy and relevance of the information used, only articles that have been published in the last five years are considered (2020 – 2024).
- Language: The studies must be published in English.
- Population: In line with the focus on developing countries and the Global South, only studies that present data, case studies, or empirical findings from these regions were considered. Only one article, published in the USA, was included due to its relevance to the topic and its contribution to developing countries.
- Publication Type: acceptable publications include peer-reviewed journal articles and conference proceedings.

Each article retrieved from the search was evaluated by analysing the title and abstract, based on the inclusion and exclusion criteria. An adaptation of the PRISMA diagram is used to illustrate the selection process (Moher et al., 2010). The PRISMA diagram is presented in Fig. 1 below.

A total of 314 articles were collected from various databases. This was followed by a screening process that involved reviewing titles and abstracts, as well as full-text analysis, to assess eligibility. After removing duplicates, 309 publications were screened for relevance to AI adoption in HEIs. Based on the research questions, 134 articles were excluded, leaving 175 for full-text review. A further 135 were excluded for not meeting criteria, resulting in 40 articles included in the final SLR. In the data analysis part of this systematic literature review (SLR), a thematic analysis was conducted on the 40 academic papers summarising main themes and their corresponding recurring codes. This thematic analysis follows an inductive approach with main codes and sub-themes.

4 Findings

The findings of this SLR are composed through the identification of relevant and popular themes discussing the adoption of AI in tertiary institutions.

Looking at the years of publication, four (4) papers were published in 2020 and 2021, ten (10) in 2022, twelve (12) in 2023, and the remaining ten (10) papers were published in 2024. This effectively showcases the expanding body of knowledge on the application of AI technologies in higher education, highlighting the growing interest in discussion since 2021.

Common ideas and themes identified in the included papers (Byrne, 2022) are presented in Table 1.

The thematic analysis successfully identifies many competencies required for the successful adoption of AI in HEIs. Additionally, the analysis highlights several benefits of using AI tools that can be applied in HEIs in developing countries. The popularity of these themes, as presented in multiple papers, displays their interdependency, as digital infrastructure and governance policies are necessary for the effective use of AI tools. The findings are further synthesised according to the research questions.

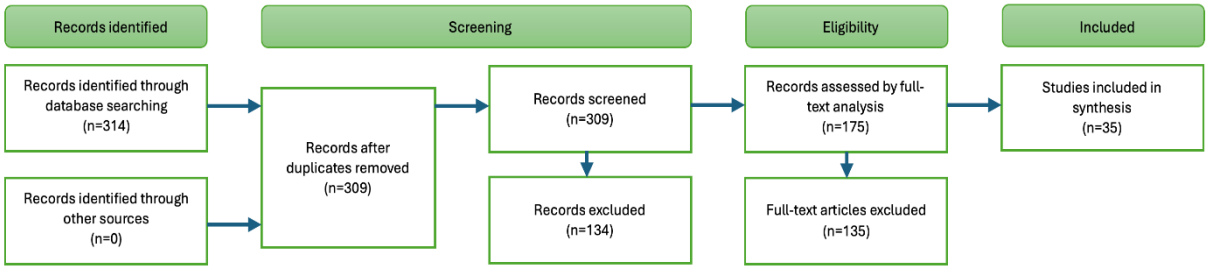


Figure 1. PRISMA Flowchart

Table 1. Thematic analysis themes

Main Themes	Sub themes	Extracted data from articles
Digital Access	Digital Infrastructure	• Quality of internet connections • Access to digital devices • Access to supporting services
Application of AI in HEIs	AI tools	• Personalised student learning • AI student performance evaluation • Remote learning Improved language with generative AI
AI Compliance Requirements	AI ethics	• Reliability, accuracy & accountability • Data privacy
	AI readiness	• Student and Educator AI familiarity • Adoption anxiety
	Policy frameworks & guidelines	• AI Governance Policy Framework

Research Question 1: What competencies are essential for effective AI adoption in higher education?

32 out of 40 papers included in this SLR discuss different competencies required for the successful adoption of AI in HEIs and can be grouped into two categories: Digital Access and AI Compliance Requirements; four (4) articles discuss 'Quality Internet Connections,' six (6) articles cover 'Access to Digital Devices,' and two (2) articles explore 'Access to Support Services.' This highlights the necessity for digital infrastructure to ensure digital access competency.

A total of twenty-five (25) out of forty (40) articles discuss competencies included in AI compliance requirements to ensure the successful application of AI in HEIs (Table 2).

Table 2. Components of AI Compliance Requirements

Competency	Focus Area
AI ethics	Reliability, accuracy & accountability; data privacy
AI readiness	Student AI familiarity; educator AI familiarity; absence of adoption anxiety
Policy frameworks & guidelines	AI Governance Policy Framework

Research Question 2: Which benefits of AI applications in the Global South would help achieve access to quality education?

Typical beneficial applications of AI in HEIs highlighted in the present papers for this SLR include:

- Personalised Student Learning (15 Papers)
- AI Student Performance Evaluation (8 Papers)
- Remote Learning (3 Papers)
- Improved Language through Generative Artificial Intelligence (5 Papers)

5 Discussion

The findings highlighted in the previous section are discussed in relation to the research questions.

Research Question 1: What competencies are essential for effective AI adoption in higher education?

Achieving AI readiness and ensuring ethical compliance in Higher Education Institutions (HEIs) in developing countries hinges on a combination of infrastructure-based, ethical, pedagogical, and policy-level strategies. The literature reveals interconnected domains - digital access, AI ethics, AI literacy, and governance structures - each with its own set of competencies, challenges, and implementation mechanisms

Digital Infrastructure. Digital access refers to the differences in the attainability of digital infrastructures, such as the internet and digital devices, among individuals, households, and even businesses (Soomro et al., 2020). Varying levels of digital access arise due to different reasons, including social and economic status. For example, female students in rural areas often reported lower access to digital services, suggesting that personal characteristics may be a factor (Mathrani et al., 2022). Strategies to promote AI adoption begin with bridging digital infrastructure gaps. Ensuring reliable internet, access to devices, and user support services are consistently identified as prerequisites (Gupta & Bhaskar, 2020; Jaiswal, 2021).

COVID-19 highlighted the challenges faced by countries in the Global South when transitioning to online learning, due to poor internet connections and a lack of support services (Maphosa, 2023). This led to students incurring high expenses by purchasing data to participate in learning, highlighting a significant disparity in digital access between developed and developing countries (Sekwatlakwatla, 2023). Despite these challenges, research has indicated that Artificial Intelligence (AI) has the potential to bridge this gap by providing opportunities for quality education (Jaiswal, 2021). However, Gupta and Bhaskar (2020) emphasise that the successful adoption of AI relies on the availability of digital technologies and infrastructure, which should be provided by Higher Educational Institutions (HEIs).

To ensure a successful AI adoption in Higher Education, in terms of digital access, the following competencies have been identified:

- Ensure quality internet connections
- Ensure access to digital devices

- Ensure access to supporting services

Mechanisms proposed in the literature include institutional investments in infrastructure, subsidised data packages, and mobile-based AI platforms as scalable solutions. HEIs must adopt a proactive role in provisioning equitable digital environments.

The subcategories are now discussed in conjunction with the corresponding competencies.

AI Ethics. Many AI tools are being introduced into the education sector, raising concerns that Generative Artificial Intelligence (GAI) is not always accurate and reliable (Labadze et al., 2023). This highlights the importance of students double-checking work generated by chatbots. AI may also lead to a decline in students' writing and critical thinking skills if they become reliant on these automated tools (Chan, 2023), making it important for students to interact with these tools responsibly and with accountability. AI can also bring about privacy concerns, as students may unknowingly share personal information (Tamanna & Sinha, 2024), which can be collected and misused.

Key ethical competencies concerning the use of AI, extracted from multiple academic sources, include:

- Ensure the reliability and accuracy of AI tools used in HEIs
- Promote accountability when using AI tools in HEIs
- Ensure AI tools promote data privacy

Mechanisms such as mandatory ethics training, AI-use policies, and digital literacy curricula have been proposed (Labadze et al., 2023). Institutional adoption of privacy-compliant tools and transparent algorithmic practices is crucial for ensuring ethical alignment.

AI Readiness. Artificial Intelligence (AI) Readiness can be defined as the ability to leverage AI technologies and tools to your advantage (Luckin et al., 2022). However, it is not just about learning what AI is; it also involves using it and familiarizing oneself with it, ultimately improving one's mental model of AI technologies and developing a comprehensive understanding of these, as well as the societal implications that come with this technology (Sperling et al., 2024), creating a growing responsibility for both educators and students. When discussing the implications of Artificial Intelligence in Higher Education, many academic papers highlight the anxiety that accompanies this new venture, as educators fear changes to their job roles (Hemachandran et al., 2022; Shahid et al., 2024). Familiarising oneself with these technologies can reduce the anxiety associated with the unknown factor of using them (Tamanna & Sinha, 2024). However, the popularity of discussions on training needed to prepare students and educators for AI readiness is not supported by the required resources. This could lead to limitations in the adoption of AI in HEIs in developing countries, particularly in the absence of relevant governing guidelines and adequate funding.

The following competencies are identified for the promotion of AI Readiness in HEI AI adoption:

- Student AI familiarity
- Educator AI familiarity
- Absence of adoption anxiety

Proposed mechanisms include structured AI-readiness programs, peer-supported learning environments, and national frameworks supporting HEI AI training pipelines.

Policy Frameworks and Guidelines. The final enabler of AI readiness and compliance is a robust governance framework. Policy guidelines are a set of structural frameworks which support the development of understanding Artificial Intelligence (AI) in the education sector (Luckin et al., 2022). AI's potential to revolutionise the education sector (Chan, 2023) highlights the importance of guidelines to form a clear understanding of how to implement these technologies (Linderoth et al., 2024), ultimately creating a systematic approach to decision-making on the implementation of these technologies. It is also crucial to establish these frameworks to determine where institutions can contribute by providing the necessary funding and infrastructure to implement AI effectively (Rahiman & Kodikal, 2024).

The following strategy focus areas have been identified to fulfil the competency of developing policy frameworks and guidelines:

- AI adoption guidelines for HEI use
- Governance of AI technologies in HEIs
- Funding for digital Infrastructure.

Effective mechanisms include the co-creation of AI guidelines with stakeholders, integration of AI into institutional strategic plans, and alignment with Sustainable Development Goals.

Research Question 2: What are the benefits of using AI in the Global South's HEIs that lead to access to quality education?

AI Tools. Artificial Intelligence (AI) is a technology that has the potential to transform every sector of our lives (Hemachandran et al., 2022), particularly in the realm of education and learning. Mannuru et al. (2023) describe Generative Artificial Intelligence (GAI) as a powerful tool for enhancing communication skills and removing language barriers. GAI has several strong points, including suggesting improvements in grammar and formatting (Alqahtani et al., 2023), as well as assessing the quality of research literature and identifying reliable sources.

AI has promoted distance learning (Mustopa et al., 2024), enabling students to maintain a social presence without being physically present in a classroom (Leoste et al., 2022), which could benefit students in developing countries by providing them with access to quality education from around the world. However, challenges may arise if there are issues with the quality and speed of internet connections.

A notable creation is the introduction of AI Teachers or 'Machine Teachers' (Kim et al., 2020), designed to answer student questions and provide a virtual reality learning experience (Opesemowo & Adekomaya, 2024). AI has enhanced the teaching

system by providing Personalised Student Learning (Labadze et al., 2023), offering tailored feedback according to individual needs (Chan, 2023), ultimately improving the student's experience and fostering inclusivity (Alam et al., 2023).

Educators also benefit from the capabilities of Artificial Intelligence. AI-driven Student Performance Evaluation systems can grade work while providing feedback (Filgueiras, 2023), streamlining mundane tasks, and giving educators more time for one-on-one instruction (Tamanna & Sinha, 2024). Additionally, AI can predict learning outcomes through big data analysis (Crompton & Burke, 2023) and identify the students who are struggling (Filgueiras, 2023).

The following benefits obtained from the use of AI tools in HEIs, as identified in multiple research sources included in this study, are as follows:

- Personalised student learning
- Automated student performance evaluation
- Remote learning
- Improved language skills and reduced language barriers using Generative Artificial Intelligence (GAI).

6 Conceptual contribution

To further conceptualise the competencies required to aid the adoption of AI in HEIs of developing countries, a Technology–Organisation–Environment (TOE) framework is used (Malik et al., 2021). The TOE framework model is presented in Fig. 2 below.

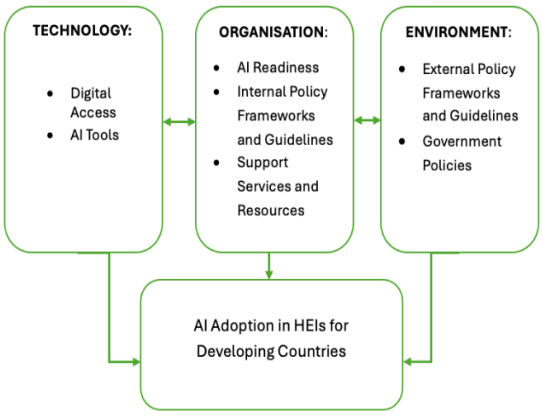


Figure 2. TOE Framework Conceptual Model

The TOE framework displays the various technological, organisational, and environmental competencies required for the adoption of AI in HEIs of developing countries and display their interdependencies.

The technology factor discusses the necessity for AI tools and digital access, which includes competencies such as access to quality internet connections, digital devices and supporting services.

The organisational factor highlights the need for AI readiness, internal policy guidelines, and relevant support services, including maintenance and training services, for the effective use of AI tools. The environmental factor discusses the need for external policy frameworks that govern the use of AI in HEIs and the conformance of government policies, such as data privacy laws. The interdependencies displayed in this TOE framework underscore the need for competencies to coexist simultaneously for the successful adoption of AI in HEIs of developing countries.

7 Conclusion

This paper employed a Systematic Literature Review (SLR) approach, where a total of 40 previously published academic papers were collected from various online databases and subjected to several screening phases to determine their relevance to answering our research questions. The research questions aimed to identify the key competencies required for the effective adoption of Artificial Intelligence (AI) in Higher Education Institutions (HEIs), specifically in developing countries (RQ1), and to highlight the potential benefits obtained from its implementation in accessing quality education in developing countries (RQ2). A thematic analysis was applied to group similar points of discussion in a more meaningful manner, grouping codes under corresponding themes and sub-themes.

This SLR has successfully investigated and identified the key competencies required for the effective adoption of AI in higher education institutions in developing countries, as well as the potential benefits. Through an in-depth analysis of the selected 40 academic papers, we have found two categories of competencies required for the successful adoption of AI in HEIs being Digital Access, composed of the digital infrastructural competencies, and the AI Compliance Requirements category composed of the sub-competency categories including: AI ethics, AI readiness and Policy Frameworks and guidelines that need to be present for the successful implementation of AI in HEIS and to obtain the benefits from using AI technologies in HEIs to access quality education in developing countries.

The research highlights an interlinking relationship between technology infrastructure, organisational infrastructure and environmental infrastructure through the conceptualisation using the TOE framework. The main challenge of the digital divide presents disparities in access to digital devices, along with limits on quality internet connections and the availability of supporting services. These create a significant barrier to the successful adoption of AI in HEIs in developing countries. This divide is emphasised by economic and personal characteristics

getting in the way, where female students face further limitations in obtaining access.

This review identified several AI tools that have the power to revolutionise and increase access to quality education in the Global South, including personalised student learning, automated student performance evaluation, and improved language using GAI. These tools present promising opportunities to expand on quality and accessible education in developing countries by creating a space to learn from any location while removing language barriers. This effective implementation relies on the presence of digital infrastructure and devices required for the use of these AI tools.

The SLR highlights the need for policy guidelines and frameworks that govern the adoption and use of these technologies in higher education. Furthermore, this review highlights the need for both students and educators to familiarise themselves with these technologies to lessen the effects of AI anxiety or resistance to change. This emphasises the need for training programs on how to adopt and use these tools effectively.

Although this SLR identified the need for training programs to increase AI readiness, there is a lack of research on how these programs can be effectively implemented. It is suggested that further research investigates the development of these training programs.

8 References

Alam F., Lim M.A., Zulkipli I.N. (2023). Integrating AI in Medical Education: Embracing Ethical Usage and Critical Understanding. Front Med (Lausanne). Oct 13;10:1279707. doi: 10.3389/fmed.2023.1279707

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., bin Saleh, K., Alowais, S. A., Alshaya, O. A., Rahman, I., Al Yami, M. S., & Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Research in Social and Administrative Pharmacy, 19(8), 1236-1242. ttps://doi.org/https://doi.org/10.1016/j. sapharm.2023. 05.016

Aly, H. (2022). Digital transformation, development and productivity in developing countries: is artificial intelligence a curse or a blessing? Review of Economics and Political Science, 7(4), 238-256. https://doi.org/10.1108/REPS-11-2019-0145

Baykasoğlu, A., Özbel, B. K., Dudaklı, N., Subulan, K., & Şenol, M. E. (2018). Process mining based approach to performance evaluation in computer-aided examinations. Computer Applications in Engineering Education, 26(5), 1841-1861. https://doi.org/https://doi.org/10.1002/cae.21971

Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. Quality & Quantity, 56(3), 1391-1412. https://doi.org/10.1007/s11135-021-01182-y

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. International Journal of Educational Technology in Higher Education, 20(1), 38. https://doi.org/10.1186/s41239-023-00408-3

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: the state of the field. International Journal of Educational Technology in Higher Education, 20(1), 22. https://doi.org/10.1186/s41239-023-00392-8

de Souza Zanirato Maia, J., Bueno, A. P. A., & Sato, J. R. (2023). Applications of Artificial Intelligence Models in Educational Analytics and Decision Making: A Systematic Review. World, 4(2), 288-313. https://www.mdpi.com/2673-4060/4/2/19

Dickson, M. G. C. A. B. R. (November 15, 2023). Doing a Systematic Review: A Student's Guide (Third ed.).

Filgueiras, F. (2023). Artificial intelligence and education governance. Education, Citizenship and Social Justice, 17461979231160674. https://doi.org/10.1177/17461979231160674

Gupta, K., Priya; Bhaskar, Preetti. (2020). Inhibiting and motivating factors influencing teachers' adoption of ai-based teaching and learning solutions: prioritisation using analytic hierarchy process. Journal of Information Technology Education: Research, 19, 693-723.

Hemachandran, K., Verma, P., Pareek, P., Arora, N., Rajesh Kumar, K. V., Ahanger, T. A., Pise, A. A., & Ratna, R. (2022). Artificial intelligence: A universal virtual tool to augment tutoring in higher education. Computational Intelligence and Neuroscience, 2022(1), 1410448.

Hermansyah, M., Najib, A., Farida, A., Sacipto, R., & Rintyarna, B. S. (2023). Artificial intelligence and ethics: Building an artificial intelligence system that ensures privacy and social justice. International Journal of Science and Society, 5(1), 154-168.

Jaiswal, A. A., Joe. (2021). Potential of Artificial Intelligence for transformation of the education system in India. International Journal of Education and Development using Information and Communication Technology (IJEDICT), 17(1), 142-158.

Jatileni, C. N., Sanusi, I. T., Olaleye, S. A., Ayanwale, M. A., Agbo, F. J., & Oyelere, P. B. (2024). Artificial intelligence in compulsory level of education: Perspectives from Namibian in-service teachers. Education and information technologies, 29(10), 12569-12596.

Khan, M. A., Khojah, M., & Vivek. (2022). Artificial intelligence and big data: The advent of new pedagogy in the adaptive e-learning system in the higher educational institutions of Saudi Arabia. Education Research International, 2022, 1-10.

Kim, J., Merrill, K., Xu, K., & Sellnow, D. D. (2020). My Teacher Is a Machine: Understanding Students' Perceptions of AI Teaching Assistants in Online Education. International Journal of Human–Computer Interaction, 36(20), 1902-1911. https://doi.org/10.1080/10447318.2020.1801227

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. International Journal of Educational Technology in Higher Education, 20(1), 56. https://doi.org/10.1186/s41239-023-00426-1

Lei, M., Clemente, I. M., Liu, H., & Bell, J. (2022). The Acceptance of Telepresence Robots in Higher Education. International Journal of Social Robotics, 14(4), 1025-1042. https://doi.org/10.1007/s12369-021-00837-y

Leoste, J., Virkus, S., Talisainen, A., Tammemäe, K., Kangur, K., & Petriashvili, I. (2022). Higher education personnel's perceptions about telepresence robots [Original Research]. Frontiers in Robotics and AI, 9. https://doi.org/10.3389/frobt.2022.976836

Linderoth, C., Hultén, M., & Stenliden, L. (2024). Competing visions of artificial intelligence in education—A heuristic analysis on sociotechnical imaginaries and problematizations in policy guidelines. Policy Futures in Education, 22(8), 1662-1678. https://doi.org/10.1177/14782103241228900

Long, D. and Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376727

Luckin, R., Cuhurova, M., Kent, C. and du Boulay, B. Empowering educators to be AI-ready. (2022). Computers and Education: Artificial Intelligence 3: https://doi.org/10.1016/j.caeai.2022.100076

Maier. (2022). Personalized feedback in digital learning environments: Classification framework and literature review. Computers and Education: Artificial Intelligence, 3. https://doi.org/info:doi/10.1016/j.caeai.2022.100080

Malik, S., Chadhar, M., Vatanasakdakul, S., & Chetty, M. (2021). Factors Affecting the Organizational Adoption of Blockchain Technology: Extending the Technology–Organization–Environment (TOE) Framework in the Australian Context. Sustainability, 13(16), 9404. https://doi.org/10.3390/su13169404

Mannuru, N. R., Shahriar, S., Teel, Z. A., Wang, T., Lund, B. D., Tijani, S., Pohboon, C. O., Agbaji, D., Alhassan, J., Galley, J., Kousari, R., Ogbadu-Oladapo, L., Saurav, S. K., Srivastava, A., Tummuru, S. P., Uppala, S., & Vaidya, P. (2023). Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development. Information Development, 0(0). https://doi.org/10.1177/02666669231200628

Maphosa, V. M., Mfowabo. (2023). African higher Education institution's response to COVID-19: A bibliometric analysis and visualisation study. Cogent Education, 10(2). https://doi.org/10.1080/2331186X.2023.2273002

Mathrani, A., Sarvesh, T., & Umer, R. (2022). Digital divide framework: online learning in developing countries during the COVID-19 lockdown. Globalisation, Societies and Education, 20(5), 625-640. https://doi.org/10.1080/14767724.2021.1981253

Ming Liu, Y. R., Lucy Michael Nyagoga, Francis Stonier, Zhongming Wu, Liang Yu. (2023). Future of education in the era of generative artificial intelligence: Consensus among Chinese scholars on applications of ChatGPT in schools. Future in Educational Research, 1(1). https://doi.org/info:doi/10.1002/fer3.10

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2010). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. International journal of surgery, 8(5), 336-341.

Mustopa, M., Nasikhin, N., Chamami, R., Nihayah, H., Habibullah, M., Romadlon;, & Manshur, A. (2024). Challenges in Artificial Intelligence Development in Higher Education in China, India, and Indonesia: International Students' Perspectives. International Journal of Learning, Teaching and Educational Research, 23(2), 354-373.

Opesemowo, O. A. G., & Adekomaya, V. (2024). Harnessing Artificial Intelligence for Advancing Sustainable Development Goals in South Africa's Higher Education System: A Qualitative Study. International Journal of Learning, Teaching and Educational Research, 23(3), 67-86.

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. Research and Practice in Technology Enhanced Learning, 12(1), 22. https://doi.org/10.1186/s41039-017-0062-8

Rahiman, H. U., & Kodikal, R. (2024). Revolutionizing education: Artificial intelligence

empowered learning in higher education. Cogent Education, 11(1), 2293431. https://doi.org/10.1080/2331186X.2023.2293431

Sekwatlakwatla, S., Prince; Malele, Vusumuzi. (2023). A Bibliometric Analysis of Generative Artificial intelligence Chatbots in Higher Education: A case study of African countries collaborating with developing nations. International Journal of Education and Development using Information and Communication Technology (IJEDICT), 19(3), 39-49.

Shahid, M. K., Zia, T., Bangfan, L., Iqbal, Z., & Ahmad, F. (2024). Exploring the relationship of psychological factors and adoption readiness in determining university teachers' attitude on AI-based assessment systems. The International Journal of Management Education, 22(2), 100967. https://doi.org/https://doi.org/10.1016/j.ijme.2024.100967

Soomro, K. A., Kale, U., Curtis, R., Akcaoglu, M., & Bernstein, M. (2020). Digital divide among higher education faculty. International Journal of Educational Technology in Higher Education, 17, 1-16.

Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in teacher education: A scoping review. Computers and Education Open, 6, https://doi.org/https://doi.org/10.1016/j.caeo.2024.100169

Steyn, A.A. and Eybers, S. (2023). Programmatic assessment to Unlocking the power of soft skills through a Micro-credential lense. Proceedings of the 2023 AIS SIGED International Conference on Information Systems Education and Research. 12. https://aisel.aisnet.org/siged2023/12.

Tamanna, M., & Sinha, B. (2024). A conceptual analysis of artificial intelligence (AI) on academic opportunities and challenges: a case study based on higher educational institutions in Bangladesh. Quality Assurance in Education, ahead-of-print(ahead-of-print). https://doi.org/10.1108/QAE-03-2024-0050

Treve, M. (2021). What COVID-19 has introduced into education: Challenges facing higher education institutions (HEIs). Higher Education Pedagogies, 6(1), 212-227.

Velinov, A., Koceski, S., & Koceska, N. (2021). A review of the usage of telepresence robots in education. Balkan Journal of Applied Mathematics and Informatics, 4(1), 27-40.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? International Journal of Educational Technology in Higher Education, 16(1), 1-27.

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. Complexity, 2021, 8812542. https://doi.org/10.1155/2021/8812542.

Attitudes Toward AI in Academic Assessment: A Survey of Croatian Higher Education Teachers

Jurica Grzunov, Ljiljana Zekanović-Korona
University of Zadar
Department of Applied Communication Studies
Dr. Franje Tuđmana 24i, Zadar, Croatia
{jgrzunov, ljkorona}@unizd.hr

Zdravko Kunić
Algebra Bernays University
Department of Information Systems and Business Analytics
Gradišćanska 24, Zagreb, Croatia
zdravko.kunic@algebra.hr

Abstract. *This paper examines how Croatian university teachers perceive the use of AI tools in academic assessment and feedback. A national survey (N = 111) shows that AI is mainly used for technical tasks such as plagiarism detection and language improvement, while its use in more complex or creative assessment remains limited. Teachers see some benefits in efficiency but express concerns about fairness, accuracy, and the quality of AI-generated feedback. Many feel unprepared and lack institutional support. The findings highlight the need for clearer guidelines and targeted training to support responsible and meaningful use of AI in higher education assessment.*

Keywords. artificial intelligence, academic assessment, higher education, teacher attitudes, large language models

1 Introduction

The growing use of artificial intelligence (AI) in higher education is changing how learning is planned, delivered, and assessed. One of the most important developments is the rise of generative AI, especially large language models (LLMs) such as ChatGPT, Claude, and Gemini. These tools can generate and improve written content in real time, which makes them useful for academic assessment and feedback. Their availability has increased expectations about efficiency, personalisation, and managing large volumes of student work. At the same time, they raise concerns about fairness, authorship, academic honesty, and the reliability of automated feedback.

As AI becomes more common in education, universities are trying to define its role in existing teaching and assessment systems. Some focus on using AI to support formative assessment and help students learn better. Others put more emphasis on detecting cheating or plagiarism, which can lead to strict or cautious use. In both cases, the teacher remains essential, not only for creating assessments, but also for interpreting AI results, giving feedback, and ensuring quality.

Research shows that teacher attitudes strongly influence how technologies like AI are used. Confidence, digital skills, institutional support, and the specific subject area all affect how teachers approach AI tools in their work. Even though international research is growing, there is still little evidence about how university teachers in specific countries view and use AI, especially in practice.

This paper addresses that gap by presenting the results of a national survey conducted among university teachers in Croatia. It explores how AI tools are used in assessment and feedback, what benefits and risks teachers see, and what personal and institutional factors shape their experiences. The study also examines the connection between AI use and factors such as digital literacy, self-assessed AI competence, and the presence of institutional guidelines. The aim is to offer useful insights for universities, policy makers, and researchers who are working to integrate AI in higher education in a responsible and effective way.

2 Challenges and perspectives of assessment in higher education in the context of AI

The emergence of ChatGPT and similar tools has sparked debate around academic integrity, while also highlighting the educational potential of such technologies. Media discourse analyses in countries like Australia, the United States, and the United Kingdom reveal that public discussions tend to focus on integrity issues, with limited attention given to students' perspectives (Sullivan, Kelly & McLaughlan, 2023). This imbalance suggests a need for a more inclusive understanding of how AI is shaping academic practices.

Although various recommendations encourage more open and pedagogically sound approaches, many institutions continue to prioritize detection and surveillance. Some scholars argue that this approach

can be counterproductive. Rather than relying on a "catch and punish" model, they advocate for fostering critical understanding of AI and using its potential to support transformative learning. This shift also includes promoting student autonomy and ethical responsibility (Kramm & McKenna, 2023).

A useful contribution to this discussion is provided by Forde-Leaves, Walton and Tann (2023), who propose a framework for studying assessment. Their model helps identify strategic and operational roles within the educational system and provides a comprehensive understanding of the complexity of assessment processes, including the role of AI tools. Similarly, Xia et al. (2024) emphasize the importance of transforming assessment practices to support self-regulated learning and uphold academic integrity. They stress the need for professional development, interdisciplinary cooperation, and educational policies aligned with technological change.

On a broader scale, more universities and international organizations are engaging with the ethical implications of AI in education. Key principles, such as transparency, accountability, inclusiveness, and human oversight, are increasingly highlighted. Research suggests that the most effective institutional responses are grounded in normative frameworks that are applied flexibly in teaching practice and clearly communicate the role of AI in curricula and assessment (Dabis & Csáki, 2024).

At the same time, traditional assessment models, especially summative ones, are becoming increasingly difficult to apply in digital and online environments. In response, a growing number of scholars advocate for "assessment for learning" approaches, which use AI and learning analytics to support collaboration, flexibility, and personalized learning (Williams, 2023). In this context, formative assessment becomes more relevant, as it offers immediate feedback and reduces student stress (Hooda et al., 2020).

Studies focusing on the application of AI in specific educational contexts further illustrate the complexity of the topic. For example, Amin (2023) demonstrates that AI can enhance foreign language instruction by enabling personalized learning, instant feedback, and automated assessment. However, the study also warns of the importance of maintaining ethical standards, equal access, and the human dimension of education. Likewise, Dong (2023) reports that using AI tools in academic writing classes can help improve student skills and increase engagement, opening new possibilities for pedagogical innovation.

Nonetheless, concerns persist, particularly regarding the detection of AI-generated content. Alexander, Savvidou and Alexander (2023) note that both educators and detection tools struggle to reliably identify such content, especially in cases involving partially AI-generated texts. This finding highlights the need for new evaluation criteria and ongoing teacher training.

Generative AI tools also raise questions about educational principles, especially given their ability to produce high-quality content in very short periods. This prompts the need to rethink curriculum design and assessment strategies (Bennett & Abusalem, 2024). In line with this, there is increasing support for learning approaches that develop higher-order thinking and encourage active student participation. Kurtz et al. (2024) propose a model for adopting GenAI technologies based on four stages, which include AI literacy, teacher empowerment, and student collaboration, while also addressing risks such as bias, inaccuracy, and dependency.

These developments have significantly influenced higher education pedagogy. Although concerns about plagiarism remain valid, some authors argue that generative AI can be integrated into assessment design to support deeper learning and metacognitive skills (Eager & Brunton, 2023). In practice, AI tools are already influencing how instructors evaluate student work. Farazouli et al. (2024) found that educators tend to be more cautious and stricter when they suspect AI-generated content, which underlines the need to update quality criteria.

As AI becomes more embedded in educational settings, it is crucial to examine how different types of assessment are affected. While students often express concerns about creativity and authenticity, educators are generally more open to using AI when it promotes critical thinking. These contrasting views highlight the need for inclusive policies and clear guidelines that involve both groups (Smolansky et al., 2023).

Despite advances in AI detection technology, its effectiveness remains limited. Studies show that neither educators nor software tools can consistently identify AI-generated content, reinforcing the argument for rethinking assessment strategies and providing training for all stakeholders (Perkins et al., 2024).

Reviews of institutional policies at leading universities show that many still lack clear strategies for addressing the challenges of GenAI. Nevertheless, some positive examples of integrating AI into assessment design are emerging (Moorhouse, Yeo & Wan, 2023).

Parker et al. (2024) argue that uncertainty in using AI tools is especially visible among early-career academics and teaching assistants, who often lack institutional guidance and clarity regarding professional responsibility. They also indicate that those with prior experience using ChatGPT are more skilled at identifying AI-generated texts, which suggests that experience and training are essential for developing digital literacy and maintaining academic integrity.

Research confirms that generative AI can support problem-solving and critical thinking but also carries the risk of misuse. Ogunleye et al. (2024) stress that task design plays a key role in preventing negative outcomes and ensuring responsible use. To guide this

process, Schön et al. (2023) propose a conceptual model that includes three dimensions: people, learning, and organization. This model emphasizes the need for new competencies, ethical guidelines, and institutional adaptations, as well as a broader awareness of the values shaping education today.

One major challenge in this evolving context is trust. When students are unsure about what is allowed and teachers do not clearly explain their expectations, this can create a sense of surveillance and distrust. Luo (2024) argues that open communication and transparent assessment policies are essential for restoring confidence between students and educators.

Even though the European Union is leading in the development of ethical AI frameworks, applying these principles in educational systems remains difficult. Tubella, Mora-Cantallops and Nieves (2023) emphasize the need for educational strategies and accessible resources to ensure the responsible use of AI in teaching and learning.

The impact of ChatGPT on essay writing demonstrates that it is possible to produce high-quality texts, but also that problems such as citation consistency and reproducibility remain. This supports the need for assessment models that focus not only on knowledge but also on competences and practical skills, aligned with lifelong learning (Kolade, Owoseni & Egbetokun, 2024).

Despite risks of misuse, many scholars advocate for the ethical use of generative AI in education. With responsible teaching and character development, authentic assessment practices can support the constructive use of AI rather than excluding it altogether (Crawford, Cowling & Allen, 2023).

Educators' views on AI remain diverse, which reflects the complexity of its integration into higher education. Research from an Australian university found that many instructors already use AI in teaching, particularly to adapt assessment tasks and stay current with technological developments (Lee et al., 2024). These findings point to the need for ongoing professional development and stronger institutional support.

This trend is also visible in primary and secondary education. Chiu (2024) shows that GenAI tools are increasingly used at earlier education levels, although public and expert debates mostly focus on higher education. Interviews with teachers and school leaders highlight the need to revise educational policies, introduce new professional standards, and support interdisciplinary teaching about AI.

Institutional policies play a central role in shaping these developments. Khlaif et al. (2024), in their study of educators in the Middle East, found that while teachers recognize the benefits of AI, such as increased engagement and reduced workload, they also express concerns about academic honesty and students' essential skills. The authors recommend stronger ethical frameworks and targeted professional development.

From a broader pedagogical perspective, Escotet (2024) cautions that AI can only enhance education when used to support personalization and feedback, not as a replacement for pedagogical judgment. While AI can help in content creation, data analysis, and administrative tasks, its use must remain rooted in human values and professional responsibility.

In the context of language education, Dong (2023) found that AI writing tools support student skill development, offer timely and personalized feedback, and improve engagement. These findings support the transformation of teaching methods, especially in academic writing in English as a foreign language.

Riapina (2024) emphasizes the benefits of personalized learning, adaptive teaching strategies, and interdisciplinary collaboration. At the same time, she highlights the importance of ethical guidelines to ensure fairness and transparency in the use of AI in education.

To sum up, the role of generative AI in higher education is complex and evolving. These technologies are not just technical tools, but also bring important questions about values in education, ethical responsibility, power structures and the goals of teaching and learning. To respond to these challenges, it is important to develop clear theoretical foundations, supportive institutional policies and teaching practices that enable responsible and meaningful use of AI. Balancing the potential of technology with human-centred values will be essential for maintaining academic integrity, encouraging deeper learning and preparing students for the challenges of a rapidly changing world.

3 Methodology

The empirical part of this study was based solely on a quantitative research design, with a structured questionnaire serving as the primary instrument for data collection. The aim was to explore the attitudes and experiences of university teachers regarding the use of AI tools for automated assessment and feedback in higher education.

Participation was voluntary and anonymous, and it was clearly stated at the beginning of the questionnaire that the research was conducted exclusively for academic purposes. The questionnaire consisted of three main sections. The first section gathered demographic data (such as age, gender, academic position, employment status, years of teaching experience, and field of study), along with self-assessments of digital and AI literacy. The second section was addressed only to respondents who had previously used AI tools, asking about the types of tools used, the teaching activities in which they were applied, and the nature of assignments where AI was involved in the assessment process. The third section presented a series of statements on the perceived advantages, limitations, and challenges related to AI-

supported assessment, rated using a 5-point Likert scale.

The questionnaire was distributed online during May 2025, by sending a call for participation to the official contact email addresses (126 in total) of all public and private higher education institutions in the Republic of Croatia, as listed on their official websites. These contacts typically represented institutional coordinators or administrative offices and were asked to further distribute the survey internally among teaching staff. In this way, at least theoretically, all academic staff in the Croatian higher education system had the opportunity to participate, provided the invitation was shared internally by their institution.

The collected data were analysed using both descriptive and inferential statistical methods, and the most relevant findings are presented and discussed in the following chapters.

4 Descriptive analysis of survey results

This chapter presents the descriptive analysis of the survey conducted among university teachers. The analysis is structured according to the thematic sections of the questionnaire and includes an overview of participant socio-demographic characteristics, prior experiences with AI tools, specific educational activities supported by these tools, and respondents' attitudes, perceived challenges, and institutional context. A total of 111 valid responses were analysed, offering an empirical foundation for understanding the current state of AI use in assessment practices within Croatian higher education. In line with the principle of anonymity guaranteed to participants, the survey did not include a question about institutional affiliation, and respondents were informed that answering questions related to their academic rank and disciplinary field was optional. Despite this, most participants voluntarily provided information about their current position and field of work, which allowed for a general overview of the sample without compromising confidentiality. This approach was taken to encourage honest responses on a potentially sensitive topic, while respecting participants' privacy.

4.1 Socio-demographic characteristics

The sample of participants in this study consisted predominantly of women (73.87%), which may reflect the gender composition of certain academic fields, as well as the generally higher participation of women in educational research. The age categories were defined to reflect common generational groupings used in social research, and to facilitate comparative analysis of responses in relation to digital familiarity and adoption tendencies. The age distribution shows that the largest group of respondents were between 40 and 49 years old (35.14%), followed by those aged 50 to 59 (28.83%) and 30 to 39 (26.13%). Younger respondents (under 30, 3.60%) and older ones (60+, 6.31%) were significantly less represented. The academic ranks of respondents varied, with associate professors being the most common (29.09%), followed by assistant professors (18.18%), teaching assistants (16.36%), full professors (10.91%), and tenured full professors (7.27%). Senior lecturers (7.27%), lecturers (4.55%), and postdoctoral researchers (5.45%) were also included. Almost all respondents (93.58%) were permanently employed at public higher education institutions, suggesting that their experiences and attitudes are shaped primarily within the public education sector.

When it comes to teaching experience, the largest share of participants reported between 11 and 20 years of experience (35.14%), followed by those with 21 to 30 years (30.63%). Less represented were those with 0–5 years (16.22%), 6–10 years (12.61%), and more than 30 years (5.41%) of experience. Most respondents came from the social sciences (31.53%) and humanities (16.22%), followed by technical sciences (14.41%), biotechnical sciences (14.41%), and biomedical and health sciences (10.81%). Other fields, such as natural sciences (4.50%), the arts (1.80%), and interdisciplinary sciences (6.31%), were represented to a lesser extent.

Most respondents rated their general digital literacy as high, selecting 4 (52.25%) or 5 (25.23%) on a 5-point scale (mean = 4.02; SD = 0.71; N = 111). However, self-assessed literacy in using AI tools was somewhat lower (mean = 3.20; SD = 1.02; N = 111), with the majority selecting 3 (39.64%) or 4 (29.73%). Only a small number of respondents rated their AI literacy as very high (5 – 9.01%), while 6.31% gave the lowest rating (1). These findings suggest that while the sample is generally digitally literate, there remains a lack of confidence and skills in using specific AI tools, which provides important context for interpreting the rest of the study.

The final question in the first section asked, "Have you ever used artificial intelligence (AI) tools for assessment or generating feedback?" This question played a key role in determining the path through the questionnaire. Of the 111 respondents, 61 (54.95%) answered "yes," while 50 (45.05%) answered "no." Based on their answers, respondents were directed to different parts of the survey. Those who answered affirmatively proceeded to the second section, where they were asked to specify which AI tools they used, and for which activities and types of tasks. Those who answered negatively skipped this section and continued directly to the third section, which focused on general attitudes and perceived challenges related to AI tools. This structure allowed for collecting data tailored to each respondent's experience, gaining detailed insights from AI users and perceptions from non-users.

4.2 Experience with AI tools for assessment and feedback

The second section was presented only to respondents who indicated prior use of AI tools for assessment or feedback purposes. A total of 61 respondents answered the first and second questions in this section, while 60 answered the third question. This part of the survey explored the specific tools, teaching activities, and types of tasks in which AI had been applied in their professional practice.

The first set of items asked participants to identify which AI tools they had used. Results show a clear dominance of large language models (LLMs). Of the 61 respondents, 59 (96.72%) reported using tools such as ChatGPT, Claude, or Gemini, indicating near-universal adoption of LLMs in academic evaluation. This high usage rate may be explained by the accessibility, flexibility, and ease of use of these tools across different educational contexts. In contrast, specialized tools such as Gradescope or FeedbackFruits were not used at all (0%), possibly reflecting limited institutional support, poor integration into existing systems, or a lack of awareness. Moderate use was reported for tools like Turnitin Draft Coach/AI Writing Feedback (29.51%), AI features in LMS platforms like Moodle (26.22%), and Microsoft-integrated AI tools (e.g., Copilot) (19.67%). These findings suggest that some educators rely on AI features embedded within familiar platforms. Less frequently used tools included Google Classroom AI (14.75%) and various niche tools such as Otter.ai, Lumi, MagicSchool AI, Curipod, and Brisk (1.64% to 4.92%). Overall, these results indicate that LLMs are currently the most widely used and accessible form of AI support in assessment, while more advanced or domain-specific tools are still in early stages of adoption. Although the survey did not explicitly ask whether respondents used free or subscription-based versions of AI tools, some comments in the open-ended responses indicated that access to paid versions (such as ChatGPT Plus) was limited. This distinction may influence the range of features used and overall user experience, which represents a relevant variable for future research.

The second group of items focused on the types of teaching and evaluative activities in which AI tools had been employed. The most frequent response was plagiarism detection (65.57%), suggesting that educators see clear value in AI for technical, automated tasks. Other common uses included stylistic or language analysis of student writing (36.07%), generating feedback suggestions (32.79%), and recommending additional materials based on student performance (32.79%). These activities suggest a gradual expansion of AI into formative teaching practices. Fewer respondents reported using AI for customizing feedback (31.15%), developing rubrics (19.67%), and analysing quiz/test responses (16.39%). Even fewer used AI for automated essay grading (11.48%) or short-answer quiz evaluation (14.75%), which may reflect hesitancy to relinquish control over interpretative assessment tasks to AI systems. Rare uses included audio/video content evaluation (8.20%), generating performance reports (14.75%), and creating tailored learning materials (18.03%). These lower rates likely reflect technical limitations, lack of tool integration, or insufficient training. In summary, AI tools are most commonly used for technical and supportive roles in assessment, particularly plagiarism detection, language analysis, and feedback generation. More complex uses, such as direct grading or creative task evaluation, remain limited, reflecting cautious trust in AI as a standalone assessment agent. In most cases, AI is seen as a supplementary tool, not a replacement for professional judgment. However, the study did not collect specific data on how teachers technically interacted with these systems. For example, whether they uploaded entire student assignments or worked with shorter excerpts using prompts. This limits deeper insight into practical workflows and ethical considerations, and will be addressed in future research.

The third set of items explored the instructional contexts in which these tools were used. The results indicate that AI tools are most frequently used for text-based tasks involving analysis, interpretation, and expression. The most common context was language and communication tasks, such as summarization, translation, or editing, as reported by 66.67% of respondents (40 out of 60). This suggests that educators recognize the strength of generative AI in processing and evaluating language-based content. AI is most often used by respondents for research papers and analytical writing (40%), including tasks like information structuring, citation checks, and stylistic improvement. Reflective writing (33.33%) and creative tasks (30%) were also commonly supported, indicating that AI is valued in open-ended tasks that emphasize expression and originality. In contrast, use in multiple-choice quizzes (28.33%) and open-ended quiz questions (10%) was lower, as was use in essay grading (28.33%). Interestingly, AI was least used in multi-phase projects (10%), oral presentations (11.67%), and portfolio assessments (6.67%). A modest share of respondents (20%) also reported using AI for administrative tasks. These findings suggest that AI is more applicable to structured, text-based assessments than to long-term or unstructured tasks, which are still difficult to automate in higher education. Overall, AI tools are most often used for tasks involving textual processing and analysis, while more complex or multidimensional tasks remain less explored.

4.3 Attitudes toward AI tools for assessment and feedback

The third and final section of the questionnaire examined university teachers' attitudes, perceived

challenges, and institutional context related to the use of AI tools for assessment and feedback.

The first set of items assessed to what extent participants agreed with various statements regarding the efficiency, fairness, objectivity, and educational value of AI tools in student evaluation processes. Respondents rated their level of agreement with each statement using a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). A total of 103 respondents completed this question, offering valuable insights into current perceptions of AI tools in the context of higher education assessment. The most positively rated statement was "AI increases the efficiency of student assessment" (mean = 3.09; SD = 0.98), indicating moderate optimism about its usefulness. In contrast, the lowest-rated statement was "AI-generated feedback is more useful to students than teacher feedback" (2.17; SD = 0.99), reflecting significant scepticism about AI's capacity to replace human pedagogical insight. Respondents also expressed low agreement with statements that AI tools are more accurate (2.47; SD = 1.03), fair (2.55; SD = 1.09), or more objective (2.87; SD = 1.07) than teachers. Only the statement about plagiarism detection received a score above the neutral midpoint (3.47; SD = 1.14), showing broader confidence in AI for technical tasks. Statements about the role of AI in the evaluation process, whether as a first step (2.50; SD = 1.15) or last (2.85; SD = 1.20), were also rated relatively low, indicating uncertainty about AI's pedagogical place. Similarly, respondents expressed mostly neutral views regarding whether AI is more suitable for formative than summative assessment (2.88; SD = 0.89).

The second part of this section focused on perceived challenges. A total of 109 respondents completed this question. The greatest concern was that AI tools might provide inaccurate or unfair evaluations (3.98; SD = 0.93), followed by the fear that AI fails to understand context, creativity, or nuance (3.61; SD = 0.97). Lack of clear institutional policies was also a major issue (3.82; SD = 1.16), as was the risk of losing control over the assessment process (3.45; SD = 1.04) and concerns over data privacy (3.36; SD = 1.24). A significant portion of respondents felt insufficiently trained (3.14; SD = 1.25) and lacking access to necessary technical support (2.91; SD = 1.24). The time investment required to learn these tools was also seen as a potential barrier (2.76; SD = 1.09). Some reported colleague or student resistance (3.07; SD = 1.12), and others were uncertain about how to apply AI tools in their current systems (3.08; SD = 1.21) or believed that the tools are not suitable for their type of evaluation (3.12; SD = 1.11). A portion of respondents did not see clear benefits (2.98; SD = 1.18) or viewed AI feedback as generic and unhelpful (3.29; SD = 1.01).

The following question was a single-item statement aimed to capture whether university teachers feel the need for additional training to use AI tools effectively and responsibly in the context of assessment and feedback. Interestingly, this was the highest-rated item in the entire section (3.99; SD = 1.22; N = 111), clearly highlighting a strong demand for structured professional development in this area.

The final question asked whether participants' institutions promote responsible AI use in education. Of the 111 respondents, 57.66% answered no, indicating a general absence of institutional support or initiative. A further 24.32% said that formal policies were in progress, while only 10.81% stated that their institution had a binding official policy in place. Other responses (e.g., "I don't know," "not familiar") were minimal but pointed to broader uncertainty and a lack of communication around institutional expectations and guidelines. This lack of formal institutional support may contribute to hesitation or inconsistent adoption of AI tools in educational assessment. Moreover, it highlights the urgent need for clearer communication, strategic leadership, and policy development to ensure responsible and effective integration of AI in teaching and evaluation practices.

5 Discussion of key findings and emerging patterns

Beyond the descriptive overview, additional statistical analysis was conducted to explore the relationships between demographic factors, digital and AI literacy, institutional support, and perceptions of AI use in assessment and feedback. This section presents the key patterns and correlations that emerged from the data and highlights relevant implications.

5.1 Institutional contexts, privacy concerns, and perceptions of AI fairness

The analysis shows that institutional policies and contexts strongly influence how university teachers use and perceive AI tools in assessment. Respondents from institutions with formal or developing AI policies were more likely to use these tools (60% vs. 52%), though they reported lower average ratings of creativity in assessment, suggesting that such policies may introduce constraints that reduce perceived autonomy.

At the same time, the presence of institutional guidelines was associated with lower concern about student data privacy (mean = 3.13 vs. 3.49), indicating that institutional support can alleviate ethical uncertainties. Perceptions of AI fairness and accuracy varied across academic disciplines and levels of teaching experience, greater trust was expressed by those in fields with well-defined evaluation criteria, such as biomedicine (Table 1), and by less experienced teachers.

Table 1. Average teacher ratings of AI evaluation across scientific fields (source: authors' work)

Scientific Field	Fairer	More accurate	More consistent	More objective
Arts	3.0	3.0	3.0	3.0
Biomedicine and Health Sciences	2.9	3.0	3.4	3.3
Biotechnical Sciences	2.8	2.6	3.4	3.2
Social Sciences	2.7	2.5	2.8	2.8
Technical Sciences	2.6	2.6	3.1	3.2
Interdisciplinary Sciences	2.5	2.2	3.2	2.8
Humanities	2.0	2.0	2.1	2.3
Natural Sciences	1.8	1.8	2.0	2.6
Overall (mean)	**2.54**	**2.46**	**2.88**	**2.9**

Among non-users of AI tools, the most frequently cited barriers were fears of unfair or inaccurate evaluation, lack of institutional guidance, and concerns about student privacy. A clear relationship was also found between privacy concerns and perceptions of fairness: the more concerned teachers were about data privacy, the less fair they considered AI evaluation to be. These findings underscore the need for comprehensive institutional policies and transparent communication strategies to foster trust and promote the responsible use of AI in educational assessment.

5.2 Correlational insights

To better understand the relationships between different attitudes toward AI tools in assessment, a Spearman correlation matrix was calculated based on all variables rated by the respondents. From the full matrix, several statistically notable and theoretically relevant correlations were identified and are discussed below.

A moderate positive correlation ($\rho = 0.50$) was found between respondents' general digital literacy and their self-assessed literacy in using AI tools. This suggests that those who feel more confident in their general digital skills are also more likely to feel capable of using AI in their teaching practice. This finding supports Parker et al. (2024), who argue that previous experience with AI tools like ChatGPT contributes to a greater sense of competence. It also highlights the importance of continuous training that includes both technical and pedagogical dimensions of AI use, as recommended by Xia et al. (2024).

Another significant finding is the correlation between the belief that AI increases efficiency in assessment and the belief that it can recognize context, creativity, and language nuances in student work ($\rho = 0.61$). Respondents who see AI as efficient are also more likely to believe that it can handle complex aspects of student responses. This shows a high level of trust in AI tools and aligns with findings by Dong (2023) and Eager & Brunton (2023), who suggest that AI can support deeper forms of learning and assessment when it is properly integrated into course design.

A particularly strong correlation ($\rho = 0.88$) was found between the belief that AI evaluates more fairly than teachers and the belief that it evaluates more accurately. This suggests a strong belief that AI can outperform human assessors in terms of objectivity. While this confidence may come from negative experiences with human inconsistency or bias, it also contradicts studies warning about the limitations of AI in detecting plagiarism or understanding context (Perkins et al., 2024; Alexander, Savvidou & Alexander, 2023). Still, it fits within the arguments by Riapina (2024) and Kolade et al. (2024), who support the use of AI as a complement rather than a replacement for teacher judgment, especially in more routine or factual tasks.

Another interesting correlation ($\rho = 0.53$) was found between the belief that AI can recognize subtle aspects of student writing and the belief that it can provide useful feedback. This supports the idea that AI tools may have a meaningful role in formative assessment. The finding is consistent with Hooda et al. (2020), who emphasize the value of immediate and personalized feedback, and with authors like Williams (2023) and Bennett & Abusalem (2024), who argue for "assessment for learning" approaches. This also suggests that teachers appreciate AI not just for speed and automation, but for its perceived contribution to the learning process.

Across the full matrix, several variables related to perceived fairness, objectivity, accuracy, and consistency of AI assessment were highly correlated. This suggests that respondents tend to group these dimensions into one overall positive impression of AI. Such patterns can be interpreted as general enthusiasm, but they may also reflect a lack of clear differentiation between different aspects of assessment quality. As Kurtz et al. (2024) point out, teacher training needs to help educators critically understand both the potential and limitations of AI tools.

One of the most relevant findings relates to the correlation between "I do not have enough knowledge about AI tools" and "I do not know how to include AI in my assessment practices" ($\rho = 0.62$). This confirms that lack of knowledge, not resistance, is a major reason why teachers do not use AI. This supports calls from Moorhouse, Yeo & Wan (2023) and Parker et al. (2024) for greater access to professional development and institutional support. Without such support, digital inequalities between educators may deepen.

Another correlation ($\rho = 0.59$) shows that those who do not know how to use AI in their own assessment systems are also more likely to say that AI tools are not useful. This shows how much perception of usefulness depends on feeling in control and

competent. Tubella et al. (2023) argue that even the best technologies remain unused without clearly defined strategies, guidelines, and support structures.

It is also worth noting the connection between two specific beliefs: that AI is more suitable for open/essay-type questions, and that it is more appropriate for formative rather than summative assessment ($\rho = 0.59$). This suggests that teachers who see AI as helpful in complex or creative student tasks are also more likely to see it as a tool for feedback and learning, not final grading. This supports the shift away from purely summative models toward "assessment for learning," as suggested by Williams (2023) and Eager & Brunton (2023). In this context, AI is seen as a support (not a substitute) for teachers.

In summary, these findings show that teachers' attitudes toward AI in assessment are interrelated and shaped by their digital confidence, familiarity with technology, and level of trust in its capabilities. High correlations between various positively perceived aspects of AI suggest that expectations may be high, but not always critically examined. As noted by Farazouli et al. (2024) and Alexander, Savvidou & Alexander (2023), it is important to balance optimism with caution and to promote the responsible and thoughtful use of AI in higher education. Institutions need to support this process by offering practical training, clear policies, and space for open discussion about the benefits and limits of AI in assessment.

6 Conclusion

The findings of this study show that AI tools are starting to find their place in Croatian higher education, but their use in assessment remains cautious and selective. Most university teachers use AI tools, especially large language models, for relatively technical tasks such as detecting plagiarism or improving student language. However, their use in more interpretative or creative forms of assessment is still not that common. This suggests that while AI is seen as helpful in supporting certain tasks, it is not yet trusted to take a more central role in academic evaluation.

Teachers in the study express moderate confidence in AI's usefulness, especially in terms of saving time and improving efficiency. However, they also express significant concerns about fairness, accuracy, and the quality of AI-generated feedback. Many are unsure whether AI can understand context, creativity, or nuance in student work, which are all elements that are essential to meaningful assessment. Importantly, most respondents report a lack of institutional guidelines and feel they do not have enough training or support to use AI tools effectively.

Correlational data from the survey reveal that teachers who feel more confident in their AI skills are more likely to see its benefits. However, general digital literacy alone does not predict AI adoption, pointing to the need for more targeted training focused specifically on educational and ethical use of AI. The absence of clear policies also contributes to uncertainty, raising concerns about data privacy, professional responsibility, and the risk of relying on AI without fully understanding its limitations.

These findings suggest that without institutional guidance and structured professional development, the integration of AI into assessment will remain fragmented and conservative. Educators require not only technical training but also support in designing pedagogically meaningful assessment strategies that include AI as a complement, not a replacement for human judgment. At the same time, institutions must ensure that AI is used responsibly, transparently, and in ways that align with academic values.

This study offers an important first look at the state of AI-supported assessment in Croatian universities. However, it also has limitations. The sample is self-selected and may not fully represent all academic fields or institutions. The survey design is primarily quantitative and does not capture deeper qualitative insights into teachers' reasoning and experiences. Future research should include interviews or focus groups, examine disciplinary and generational differences, and explore how students experience and respond to AI-based assessment. These aspects will be essential for developing ethical, inclusive, and pedagogically sound uses of AI in higher education. In addition, future studies should aim to collect more detailed data on how AI tools are used in concrete teaching and assessment practices, and distinguish between different types of access and tool functionalities. A more granular demographic and disciplinary breakdown of participants will also be important for assessing patterns of adoption and perception across diverse academic contexts.

References

Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting AI-generated writing in second language education in higher education. *The Journal of Teaching English with Technology*, 23(2), 25–43. https://doi.org/10.56297/BUKA4060/XHLD5365

Amin, M. Y. M. (2023). AI and Chat GPT in language teaching: Enhancing EFL classroom support and transforming assessment techniques. *International Journal of Higher Education Pedagogies*, 4(4), 1-15.

Bennett, L., & Abusalem, A. (2024). Artificial Intelligence (AI) and Its Potential Impact on the Future of Higher Education. *Athens Journal of Education*, 11(3), 195-212.

Chiu, T. K. (2024). The impact of Generative AI (GenAI) on practices, policies and research

direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 32(10), 6187-6203.

Crawford, J., Cowling, M., & Allen, K.-A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice*, 20(3). https://doi.org/10.53761/1.20.3.02

Dabis, A., & Csáki, C. (2024). AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI. *Humanities and Social Sciences Communications*, 11(1). https://doi.org/10.1057/s41599-024-03526-z

Dong, Y. (2023). Revolutionizing academic English writing through AI-powered pedagogy: Practical exploration of teaching process and assessment. *Journal of Higher Education Research*, 4(2).

Eager, B., & Brunton, R. (2023). Prompting higher education towards AI-augmented teaching and learning practice. *Journal of University Teaching and Learning Practice*, 20(5), 1-19.

Escotet, M. Á. (2024). The optimistic future of Artificial Intelligence in higher education. *Prospects*, 54(3), 531-540.

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2024). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, 49(3), 363-375.

Forde-Leaves, N., Walton, J., & Tann, K. (2023). A framework for understanding assessment practice in higher education. *Assessment & Evaluation in Higher Education*, *48*(8), 1076-1091.

Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, 2022(1). https://doi.org/10.1155/2022/5215722

Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A., ... & Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences*, 14(10).

Kolade, O., Owoseni, A., & Egbetokun, A. (2024). Is AI changing learning and assessment as we know it? Evidence from a ChatGPT experiment and a conceptual framework. *Heliyon*, 10(4).

Kramm, N., & McKenna, S. (2023). AI amplifies the tough question: What is higher education really for? *Teaching in Higher Education*, 28(8), 2173-2178.

Kurtz, G., Amzalag, M., Shaked, N., Zaguri, Y., Kohen-Vacs, D., Gal, E., ... & Barak-Medina, E. (2024). Strategies for integrating generative AI into higher education: Navigating challenges and leveraging opportunities. *Education Sciences*, 14(5).

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., ... & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, 6.

Luo, J. (2024). How does GenAI affect trust in teacher-student relationships? Insights from students' assessment experiences. *Teaching in Higher Education*, 30(4), 991–1006. https://doi.org/10.1080/13562517.2024.2341005

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, 5.

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). Higher education assessment practice in the era of generative AI tools. *Journal of Applied Learning & Teaching*, 7(1). https://doi.org/10.37074/jalt.2024.7.1.28

Parker, L., Carter, C., Karakas, A., Loper, A. J., & Sokkar, A. (2024). Graduate instructors navigating the AI frontier: The role of ChatGPT in higher education. *Computers and Education Open*, 6. https://doi.org/10.1016/j.caeo.2024.100166

Perkins, M., Roe, J., Postma, D., McGaughran, J., & Hickerson, D. (2024). Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. *Journal of Academic Ethics*, 22(1), 89-113.

Riapina, N. (2024). Teaching AI-enabled business communication in higher education: A practical framework. *Business and Professional Communication Quarterly*, 87(3), 511-521.

Schön, E.-M., Neumann, M., Hofmann-Stölting, C., Baeza-Yates, R., & Rauschenberger, M. (2023). How are AI assistants changing higher education? *Frontiers in Computer Science*, 5. https://doi.org/10.3389/fcomp.2023.1208550

Serrano, M. M., O'Brien, M., Roberts, K., & Whyte, D. (2018). Critical Pedagogy and assessment in higher education: The ideal of 'authenticity'in learning. *Active Learning in Higher Education*, 19(1), 9-21.

Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator and

student perspectives on the impact of generative AI on assessments in higher education. *Proceedings of the Tenth ACM Conference on Learning @ Scale*. https://doi.org/10.1145/3573051.3596191

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1), 1-10. https://doi.org/10.37074/jalt.2023.6.1.17

Tubella, A. A., Mora-Cantallops, M., & Nieves, J. C. (2023). How to teach responsible AI in higher education: Challenges and opportunities. *Ethics and Information Technology*, 26(3). https://doi.org/10.1007/s10676-023-09733-7

Williams, P. (2023). AI, analytics and a new assessment model for universities. *Education Sciences*, 13(10).

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21(1).

Enhancing Learning Experiences: The role of mixed reality in Higher Education

Kelebogile M Dibetso, Adriana Aletta Steyn
University of Pretoria, 83 Lynnwood Rd
Department of Informatics
Hartebeestpoort 362-Jr, Pretoria, South Africa, 0083
u20496703@tuks.co.za, Riana.steyn@up.ac.za

Abstract. *Immersive technologies, such as Augmented Reality, Virtual Reality, and Mixed Reality, are increasingly adopted in higher education for their ability to enhance learning through interactive, experiential environments. This systematic literature review, following PRISMA guidelines and covering 45 papers from 2020 to 2024, examines the impact of mixed reality on student engagement, personalized and collaborative learning, and its limitations. Findings indicate that these technologies enhance learning outcomes and satisfaction; however, they face barriers such as high costs, infrastructure demands, and cybersickness. The review provides recommendations to address these challenges and identifies future research needs, particularly in under-resourced educational settings.*

Keywords. Augmented reality; extended reality; higher education; immersive learning; mixed reality; student engagement; virtual reality.

1 Introduction

The rapid advancements in technology have brought about changes in various aspects of our lives, including the way we live, communicate with each other, work, and learn. As these innovative technologies evolve, the world remains connected through network nodes, increasing accessibility to these advanced technologies. This kind of accessibility provides a platform for different people to utilise these technologies. Current technologies changing how we learn are immersive technologies. These technologies have the characteristic to enhance the way we live and how we learn (Harjana et al., 2023). Immersive technologies encompass augmented reality (AR), virtual reality (VR), mixed reality (MR), and the umbrella term extended reality (XR). These kinds of technologies have the functionality of blending the physical and virtual worlds.

The capabilities of immersive technologies differ; each technology has unique strengths that make it a suitable choice for specific applications, depending on the case study. With each unique strength, the common goal remains to blend the physical and virtual worlds to create an interactive environment.

Mixed Reality (MR) integrates components from Augmented Reality (AR) and Virtual Reality (VR), resulting in a hybrid reality where physical and digital objects coexist. Extended reality (XR) is an immersive technology that involves blending virtual and real-world elements to create a virtual environment where digital and physical objects interact (McMillan et al., 2017). Mixed reality is also defined as one of the emerging technologies that aids in facilitating teaching and learning processes (Suryodiningrat et al., 2021). This technology enables students to interact with 3D models and simulations, promoting a deeper understanding of disciplines ranging from science and technology to the arts (Baltezarevic & Baltezarevic, 2024).

The use of this technology in education has demonstrated an improvement in the quality of teaching and learning experiences (Harjana et al., 2023). This immersive technology creates realistic scenarios that enhance students' cognitive abilities (Petruse et al., 2024). Despite the potential benefits of adopting this technology in higher education, several limitations exist, including health concerns, student engagement issues (Steyn et al., 2024), and accessibility barriers.

This paper explores the role of Mixed reality technologies in higher education. It contributes to understanding the potential of how these technologies can enhance learning experiences and improve educational curricula despite various challenges. The objective is to provide educators in higher education with insights and recommendations on the adoption of extended reality for academic advancement. The guiding research question is: What role can immersive technologies play in Higher Education? The outline of this paper begins with an introduction that establishes the research problem and objectives. This is followed by a literature review on extended reality, which includes virtual reality, augmented reality, and mixed reality, as well as the adoption of this technology in higher education and its limitations. The research method, selection criteria, findings, discussion, and

conclusion are presented, followed by a recommendation for future research.

2 Literature review

The use of Mixed Reality (MR) technology in higher education has sparked considerable interest, particularly because it promises to enhance learning experiences across various disciplines. Chen et al. (2017) revealed that MR has significant promise in medical education, particularly in training and surgical planning. The authors contend that MR can give essential patient-specific data, thereby improving the training landscape for medical practitioners. However, they also acknowledge the technological constraints and expenses that have hampered wider acceptance, paving the way for future breakthroughs in this field.

Birt and Cowling (2018) analysed the use of MR in design education using a comparison visualisation methodology. Their findings suggest that MR technologies enhance student outcomes by fostering reflective learning and participatory teaching methods. They argue for a departure from traditional approaches, emphasizing the need for further study to determine which technologies best serve the diverse learning modes in higher education, a point echoed by various authors (Steyn et al., 2024a; Steyn et al., 2024b). MR applications need to be tailored to maximize educational benefits (Chen et al., 2017).

He & Perlin (2020) studied the usefulness of MR in classrooms utilizing the Chalktalk software platform. The study examined how MR can enhance engagement and retention during face-to-face encounters, providing empirical data to support its inclusion in educational methods. This study supports the idea that MR can enhance traditional learning settings by making them more engaging and effective.

Misirlis et al. (2023) explored students' plans to use Meta-Education technology, including MR applications. They highlight both the potential benefits, such as improved learning settings and gamification, and the obstacles posed by technical adoption, infrastructure, and cultural differences. Their findings indicate that while MR has enormous promise in education, there are still considerable barriers to adoption.

2.1 Evolution of technology in education

MR technologies have advanced significantly in recent years, with improvements in both hardware and software enabling educational institutions to utilize them as effective teaching aids. Augmented reality and virtual reality technologies have gained popularity in higher education due to their ability to engage students through experiential learning and immersive simulations. These technologies are commonly used in STEM fields, where visualization and interaction with complex systems are critical (Banjar et al., 2023; Hughes et al., 2005).

Augmented Reality (AR) overlays digital content in the real world, allowing students to perceive and interact with 3D objects while remaining in their physical context. This method is beneficial in subjects such as biology and engineering, where augmented reality can provide visual representations of molecular structures, anatomical models, and mechanical systems (Valladares Ríos et al., 2023). Virtual Reality (VR) immerses students in totally virtual surroundings, allowing them to explore and engage with information in ways that are not feasible in the real world. In medical education, virtual reality (VR) will enable students to practice surgical procedures, explore human anatomy, and mimic clinical settings. This allows students to apply their practical learning skills without the hazards associated with real-life practice. Virtual reality is also making substantial contributions to the humanities, providing immersive experiences in archaeology and history (Valladares Ríos et al., 2023).

2.2 The adoption of mixed reality in higher education

MR technologies are utilized in various higher education disciplines, including STEM fields such as science, engineering, and medicine. These technologies offer interactive simulations that help students grasp complex systems, thereby enhancing information retention and practical application (Marks & Thomas, 2022; Alexander et al., 2019).

Virtual reality simulations in medical education enable students to practice operations and learn anatomy without the use of cadavers or real-life patients, hence increasing confidence and technical abilities. Augmented reality augments lectures and demonstrations by presenting 3D renderings of body parts, allowing students to learn anatomy interactively (Baniasadi et al., 2020).

In the arts and social sciences, a mixed reality environment enables students to visit historical locations or engage in digital cultural activities virtually. According to studies, students who studied ancient civilizations using virtual reality had a greater awareness and knowledge of historical events and societal systems than those who used textbooks alone (Banjar et al., 2023).

Mixed reality offers numerous significant advantages in educational contexts, including increased engagement and motivation, enhanced comprehension of complex topics, and support for collaborative learning. The immersive aspect of mixed-reality settings draws students in and encourages them to engage actively in their studies. For example, in physics classes, students who used virtual reality simulations retained 25% more complicated concepts than those who learned through lectures and textbooks. MR-enhanced collaborative learning has been

demonstrated to improve students' collaboration and problem-solving skills (Munfaridah et al., 2021).

2.3 Challenges of adopting MR

Mixed Reality (MR) has the potential to improve or even revolutionize learning experiences in higher education. However, despite its numerous benefits, the broad adoption of mixed reality is hindered by several external challenges, including high prices, infrastructure limitations, and the readiness of educators and students to utilize the technology (Banjar et al., 2023).

The high cost of hardware and software, especially in underdeveloped countries, can significantly contribute to the integration of these immersive technologies in higher education. Underdeveloped countries have several socio-economic issues which result in an unequal standard of living. This factor limits equal student accessibility to the use of mixed reality, which hinders inclusive growth in education (Kezia Herman & Bordoloi, 2024). Technical impediments, such as a lack of reliable and fast internet connections, equipment that does not cater to all students, and a lack of user-friendly interfaces, hinder the effectiveness of mixed reality in lecture rooms, which frustrates both educators and students. Educators must possess specific technical skills and know how to incorporate MR into their teaching methods effectively. Training must be available to upskill educators (Harjana et al., 2023) (Blooma et al., 2022) to ensure that they are equipped to deliver quality educational content.

According to Harjana et al. (2023), the key challenges to the adoption of mixed reality include a lack of funding, a lack of understanding among users on how to operate immersive technologies, and a slow adaptation among students and teachers. These challenges raise significant concerns. The study found that more than 90% of students consider learning with mixed reality to be more effective than traditional methods. Despite these challenges, the potential benefits of mixed reality in higher education are too significant to ignore. Mixed reality can transform the way students learn and foster collaboration between students and their educators in ways that traditional methods cannot. Mixed reality can provide content for different education sectors. It is necessary to address these issues so that higher education can fully reap the benefits of the power of mixed reality (Birt et al., 2018) to create engaging and effective learning environments.

3 Methodology

This study uses a systematic literature review to synthesize current knowledge, identify research gaps, and offer recommendations (Dinter et al., 2021). Following established guidelines (Braun & Clarke, 2006), relevant literature was sourced from major databases via the University of Pretoria's library, which includes ProQuest, Science Direct, ACM Digital Library, IEEE, and Emerald Insight.

The following search string was used on the identified databases: ("Mixed Reality") AND ("Higher Education") OR (tertiary education) AND ("Virtual Reality") AND ("Extended Reality") AND ("Augmented Reality").

The review included only peer-reviewed, open-access research articles and conference papers published in English between 2020 and 2024, filtered by the relevant keywords. Non-peer-reviewed research papers were excluded. This is to avoid grey literature and maintain accuracy and reliability.

Using the PRISMA flowchart, 239 records were identified, duplicates and irrelevant studies were removed, and 42 papers meeting all criteria were included in the final review (Fig. 1).

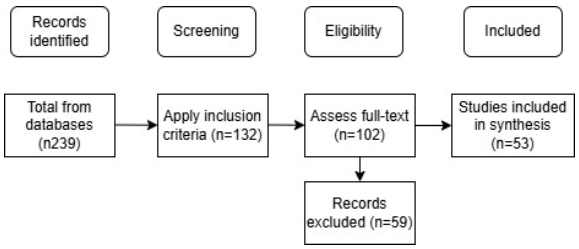


Figure 1. PRISMA flowchart

Data extraction, guided by the PRISMA flowchart, selected articles aligned with the study's objectives and criteria. Thematic analysis was used to systematically identify codes and themes (Braun & Clarke, 2006; Clarke et al., 2015). With support from generative AI (ChatGPT, 2024), key themes emerged: 'Enhancing Learning Experiences' (with subthemes 'Interactive Learning' and 'Personalized Learning') and 'External Challenges' (including 'Financial Constraints' and 'XR Limitations'). Retrieved articles were grouped accordingly.

4 Findings

The discussion of the adoption of mixed reality or similar technologies in higher education has become quite popular in recent years.

Table 1. Themes identified

Main Theme	Sub theme	References
Enhancing Learning Experiences	Interactive learning	(Andalib & Muntazar, 2024), (Banjar & Campbell, 2022), (Banjar et al., 2023), (Bermejo et al., 2023), (Dong et al., 2023), (Dutt et al., 2022), (Harjana et al., 2023), (Hensen & Klamma, 2022), (Hutson & Olsen, 2022), (Izard et al., 2021), (Jiawei & Mokmin, 2023), (Jin et al., 2022), (Li, 2023), (Lohfink et al., 2022), (Meccawy, 2023), (Pan et al., 2021), (Pellas et al., 2021), (Perçuku & Minkovska, 2022), (Salloum et al., 2023), (Sofiadin, 2023), (Wang & Li, 2024)
	Personalised learning	(Alkhudair et al., 2024), (Jiang et al., 2024) ,(Moser et al., 2024), (Peters et al., 2021), (Petruse et al., 2024), (Suryodiningrat et al., 2021), (Tang et al., 2020), (Staddon, 2020)
	Financial constraints	(Antón-Sancho et al., 2023), (Ewais et al., 2022), (Ironsi, 2023), (Laine & Lee, 2024), (Marks & Thomas, 2022), (Nesenbergs et al., 2021)
External Challenges	XR limitations	(Baxter & Hainey, 2020), (Beetul et al., 2023), (Duan, 2024), (Kezia Herman & Bordoloi, 2024),(Lai & Cheong, 2022), (Li et al., 2021), (Obeidallah et al., 2023), (Radianti et al., 2020), (Hajirasouli & Banihashemi, 2022)

Regarding geographic location and types of papers, Fig. 2 and Table 2 provide a visual representation of the number of articles published on the topic of mixed reality adoption in higher education from 2020 to mid-2024. The trend in these figures suggests that more papers are being published to address the topic, indicating a growing interest in the adoption of mixed reality technologies in higher education.

Table 2. Papers per conference or Journal

Year	Conference papers	Journal papers
2020	0	4
2021	3	4
2022	5	5
2023	9	4
2024	8	1

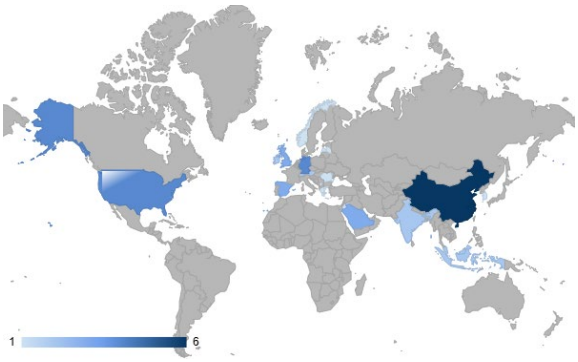


Figure 2. Geographical location

The papers were published in different countries. The country with the most published papers is China with the least being the countries in the light-shaded blue. No papers were identified in the countries shaded in the grey colour.

The papers discovered address the use of extended reality technologies can be used in different educational fields.

Fig. 3 indicates that extended reality technologies are commonly used STEM fields in education.

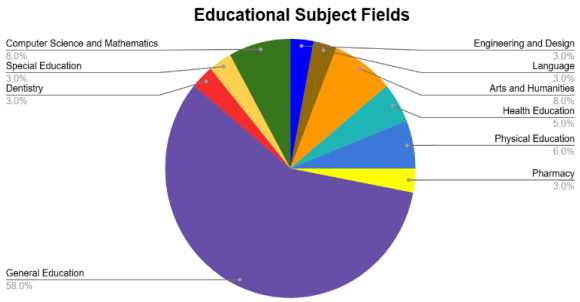


Figure 3. Various fields in education using XR

5 Discussion

5.1 Enhancing learning experiences

Mixed reality technologies promote interactive and personalized learning experiences. Students can utilize these immersive technologies to enhance their understanding of subject matter in technical fields such as engineering, architecture, medicine, geography, and science (Andalib & Muntazar, 2024). The virtual simulations enable students to interact with complex 3D model prototypes, providing them with a real-life experience that is not available with traditional methods (Andalib & Muntazar, 2024). Interactive learning through virtual simulations supports embodied learning theory, which allows students to learn by applying cognition and physical interaction with objects (Jin et al., 2022).

These simulations improve students' geometric analysis and creativity skills (Harjana et al., 2023). Immersive technologies are primarily applied in subjects across both STEM and the Bachelor of Arts

fields. They are commonly used in science and technology studies (Dong et al., 2023; Banjar et al., 2023; Meccawy, 2023) due to the immersive qualities of these fields. Art is still predominantly presented using traditional teaching methods in classrooms (Jiawei & Mokmin, 2023); however, art content is also offered using virtual reality (VR). Extended is used to arrange a virtual tour of art museums for Bachelor of Arts students (Hutson & Olsen, 2022) and archaeological museums for Humanities students (Lohfink et al., 2022). A virtual tour allows users to view, engage with, and interact with the artwork, thereby witnessing an authentic virtual experience through a learning process (Lohfink et al., 2022).

In health and medical studies, extended reality tools are enabling health students to interact with holograms rather than physical bodies (Bermejo et al., 2023); a similar concept can be applied in ICT coursework. Villena-Taranilla et al. (2022), cited in Dong et al. (2023), reported that immersive technologies are used in the Bachelor of Arts field by displaying historical tales and historical artefacts to students (Dong et al., 2023). Extended reality is applied in sports training. Putranto et al. (2023) outlined how VR in sports training can enhance players' performance analysis, prevent injuries, and improve training skills through various training scenarios. This demonstrates that extended reality technologies are leveraged to enhance the interactive learning experience of students across multiple fields of education.

Undergraduate computer science students were enrolled in a practical course to apply their skills using MR concepts on software methodologies and development lifecycles (Hensen & Klamma, 2022). The students were divided into groups to collaborate on the project requirements.

Tang et al. (2020) investigated the learning effectiveness of students who enrolled in the design subject. One group of students studied using HoloLens, a mixed-reality headset that enables digital content to overlay with the real-life environment. Students used HoloLens technology to visualize the digital model of an aircraft turbofan. This demonstrates how students can benefit from personalized learning using mixed reality features; this content delivery enhanced how information is provided through the visualisation of the digital model. Students were able to interact with the digital model in real time through hand gestures, voice commands, and gaze (Tang et al., 2020), thereby enhancing their learning. An alternative to the use of headsets is Light Field Display (LFD), a technology that projects 3D images through light in multiple directions. Students can view the pictures without wearing headsets (Pan et al., 2021), achieving interactive learning through simulation. Active participation enhances students' understanding of the subject content (Harjana et al., 2023).

The use of extended reality technologies provides a platform for students to engage in class interactively (Banjar et al., 2023) using 2D and 3D environments. Mixed reality enables real-time collaboration, allowing multiple users to interact within the same virtual space and exchange feedback with one another (Wang & Li, 2024). This dynamic interaction promotes deeper learning and more effective collaboration (Perçuku & Minkovska, 2022). The Metaverse is a virtual world that enables connections between the virtual and offline worlds (Salloum et al., 2023). Higher educational institutions can be part of the Metaverse, allowing their stakeholders to connect. Traditionally, lectures are delivered by one lecturer to a small audience, which compels all parties to stay in place in the lecture room. The Metaverse offers users a "cyber-physical" experience, enabling them to navigate seamlessly between physical and online environments (Salloum et al., 2023). This kind of environment makes learning more engaging, interactive, and interesting for students in higher education.

Students exposed to virtual reality-enhanced learning experiences display higher levels of engagement and enhance their critical thinking and problem-solving skills (Pellas et al., 2021).

The HoloLens and Oculus mixed-reality technology is used to deliver human gross anatomy laboratory content to medical students (Jiang et al., 2024; Izard et al., 2021). Through visual simulations, students were able to perform surgical operations while simultaneously receiving feedback on the surgical steps to take and patient information (Jiang et al., 2024). Students were able to learn without the risks associated with surgery. The active use of such immersive technology enables users to enhance their skills, thereby reducing the likelihood of human error (Jiang et al., 2024) and increasing the rate of successful operations.

Extended reality environments cater to the diverse learning styles of users. This is achieved through the active development of the training content in an extended reality Learning Content Management System (LCMS) based on an individual's user data (Moser et al., 2024). The integration of generative AI with the extended reality environment can enhance the generation of content, thereby improving the adaptive learning experiences of users (Moser et al., 2024).

Expanding on the benefits of extended reality allows students to personalize their learning experiences. Students have their preferred learning methods, and an extended reality, such as MR, is flexible enough to accommodate each student's learning methods (Suryodiningrat et al., 2021). This allows students to interact in the classroom in their way. A mixed reality system enhances student interaction, which improves students' confidence in learning and their self-efficacy as they learn at their own pace.

Studies suggest that age may have an influence on students' learning outcomes using technology-enhanced learning (Staddon, 2020). Older students outperformed younger students when using a technology-enhanced learning system (Staddon, 2020),

indicating that age is a factor to be considered when designing and implementing technology-enhanced learning systems (Petruse et al., 2024) to accommodate different age groups in higher education.

A user-centred design (UCD) approach ensures that users' needs are accounted for from the beginning of the design process (Petruse et al., 2024) until the end of the design lifecycle. User-centred design ensures that augmented reality environments are designed to meet users' needs, enhance the end-user experience when using the environment, and accommodate their preferences (Petruse et al., 2024). These components ensure meaningful learning experiences when using augmented reality environments. To understand a user's needs, a persona needs to be developed (Petruse et al., 2024) to inform the design of interfaces in an augmented reality environment. Failure to include users' needs may result in students being confused about how to use the environment, which defeats the goal of enhancing students' learning experiences (Petruse et al., 2024).

The adoption of extended reality has several benefits, including enhanced student engagement, interactive learning, collaborative learning, and personalized learning; however, how do educators measure the effectiveness of extended reality technologies in their lecture rooms? Sofiadin (2023) provided students with a quiz to evaluate their understanding of learning activities using an XR application.

The results show a significate difference in the students' performance. 58 students used the XR app learning experience, 95% of the students scored a grade A (81–100), and the remaining 5% scored a grade B (71–80). As opposed to the 63 students who did not use the XR app learning experience, only 14% scored a grade A, 67% scored a grade B and 19% scored a grade C (41-60).

The results show that more students who used the XR application obtained a mark between 81-100 than the students who did not utilise the XR application. The difference in the number of students is quite significant, validating that immersive technologies have the potential to improve students' achievements.

Learning difficulty is an umbrella term for learning disorders that affect one's capability to learn (Dutt et al., 2022); examples include Dyslexia, Dyspraxia, and Perceptual Disabilities. Students who live with a learning difficulty get below-average marks compared to students of their age who do not live with any learning difficulty (Dutt et al., 2022). Learning disabilities can put a strain on how students grasp information, leading to demotivation and reduced academic performance. Extended reality technologies can provide support for individuals in need (Dutt et al., 2022), including those who are visually or hearing-impaired. Students can use interactive learning environments to enhance their learning outcomes.

Extended Reality (XR) technologies, including Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), have demonstrated significant potential benefits in enhancing learning experiences through interactive and personalized education. These benefits include increased student engagement and motivation. These immersive technologies allow students to interact with 3D models using embodied learning. This learning helps students understand educational content in depth. Students can apply the skills they learned in the virtual world in real-world scenarios. Extended reality technologies have the flexibility to cater to diverse learning styles, allowing students to learn at their own pace while applying personalized learning. While existing literature has outlined the benefits and challenges of immersive technologies, there is limited guidance on how institutions can integrate these technologies pedagogically. To address this gap, this paper proposes a conceptual model designed to guide the implementation of immersive technologies in higher education. The details of the model are later discussed in the paper.

5.2 External challenges

The following section reflects the findings discussed under the theme of external challenges, focusing on the sub-themes of financial constraints and XR limitations. Mixed reality technologies require specialized hardware, including virtual reality head-mounted displays, high-performance computers, interactive devices, and virtual reality software. The cost of a virtual reality headset can range from $100 to $1,000 (Marks & Thomas, 2022), depending on the quality and quantity required for purchase. Virtual reality head-mounted displays require a connection to high-performance computers that can handle the intensive processing of virtual reality environments. In particular, organising a functional virtual reality lab equipped with 26 units of quality hardware at the University of Sydney costs over AU$100,000 (Marks & Thomas, 2022). While these figures provide a proper estimate, the cost of virtual reality hardware has been subject to fluctuation since 2022, due to technological advancements and inflation.

Additionally, facility and operational costs are incurred, which include classroom space and the Internet. Classrooms must be designed to allow free movement during virtual interactions with the virtual reality head-mounted displays. Small classrooms cause a lack of oxygen flow, which results in a hot and stuffy environment (Marks & Thomas, 2022); thus, redesigning classrooms adds to the financial burden. Slow internet connectivity can hinder the sharing of virtual reality content; good infrastructure is required to support high internet connectivity for streaming or to download content for all users (Marks & Thomas, 2022). A critical challenge to the adoption of immersive technologies in higher education is the readiness of infrastructure in higher educational institutions (Ewais et al., 2022). The lack of Internet,

inadequate IT infrastructure, and the cost of hardware and licensing maintenance create a limitation (Antón-Sancho et al., 2023).

The use of immersive technologies in Tanzania is quite minimal (Kezia Herman & Bordoloi, 2024), with some activities of augmented reality in the tourism space; however, there are few opportunities for the engagement of augmented reality in higher education. Other African countries could experience the same challenge, which is why no articles were found published in any African countries relating to the research question. The lack of good infrastructure is a barrier that makes it challenging to ensure that extended reality technologies are accessible (Kezia Herman & Bordoloi, 2024).

The costs of adopting immersive technologies could increase educational inequalities (Ironsi, 2023), which, unfortunately, stunts the growth of digital inclusion, as not all students may have the financial means to have access to such equipment.

The skill gap between trained educators who are proficient in using immersive technologies is also a challenge experienced when incorporating extended reality (Kezia Herman & Bordoloi, 2024). This challenge is a result of the limited adoption of immersive technologies, specifically augmented reality, in higher educational institutions in India (Hajirasouli & Banihashemi, 2022). In India, 80.6% of faculty and students have never attended any training on augmented reality technologies (Kezia Herman & Bordoloi, 2024). Inadequate augmented reality and virtual reality training programs can lead to poor collaborative learning experiences for both educators and students, as educators may struggle to deliver content efficiently, and students may struggle to grasp the concepts being taught. Higher education institutions should organise extended reality-based training materials that educators can enrol in (Duan, 2024) to learn how to navigate the basic operating skills necessary for preparing their classes. Additionally, they should invite experts to the sessions to provide their advice on real scenarios (Duan, 2024). Additionally, to combat this challenge, students must familiarise themselves with the use of immersive technologies at their own pace (Nesenbergs et al., 2021) by also integrating traditional learning methods.

Prolonged use of mixed reality equipment can cause users to experience cybersickness (Laine & Lee, 2024). Cybersickness is a visually induced form of motion sickness that causes symptoms such as discomfort, dizziness, headache, and dry eyes, among others (Laine & Lee, 2024). Individuals who have epilepsy, frequent headaches, and vertigo are more susceptible to cybersickness (Beetul et al., 2023), symptoms caused by prolonged use for more than an hour. Fortunately, users do not experience cybersickness to the same degree (Beetul et al., 2023). The recommended practice to minimise the symptoms of cybersickness is not to use immersive technology devices for more than two hours (Beetul et al., 2023).

Baxter and Hainey (2020) asked 83 participants for their opinions on the challenges associated with adopting virtual reality technology. One participant mentioned that they were unable to use the head-mounted displays because they wear glasses (Baxter & Hainey, 2020). The equipment's inability to accommodate users who wear glasses is a limitation. This raises a concern about other physical limitations users may have that could restrict their ability to wear and use extended reality head-mounted displays. The current designs of immersive technological products are used more for entertainment-oriented than education-oriented purposes (Obeidallah et al., 2023; Radianti et al., 2020). These devices should have the functionality to allow users to edit the presented content (Radianti et al., 2020). The simulation functionality of immersive technologies does not necessarily appeal to all students. A study reported that some students expressed that the simulation does not replicate the sports training experiences of the real world (Li et al., 2021). This may hinder students from applying the skills learned in the virtual world to real-life experiences.

Extended reality offers numerous benefits; however, educators must ensure that they incorporate extended reality into their teaching methods to address known problems (Lai & Cheong, 2022). Ineffective implementations occur when there is a misalignment between the teaching objectives and the technological solutions provided. Mikropoulos and Natsis (2011), as cited by Lai and Cheong (2022), revealed that the use of virtual reality in education between 1999 and 2009 lacked a clear direction on how technology is used to solve pedagogical challenges (Lai & Cheong, 2022). It is essential to develop a curriculum that continues to deliver high-quality educational content, incorporating extended reality (Obeidallah et al., 2023).

While extended reality technologies offer exciting enhancements in education, several challenges must be addressed. Challenges include cost, infrastructure, educator training, accessibility, and health considerations. Higher educational institutions should focus on ways to overcome these barriers to have a seamless adoption of MR reality in academic curricula.

The adoption of Mixed Reality (MR) technologies in higher education holds great potential for enhancing learning by merging physical and virtual environments. This conceptual model, Fig. 4, outlines an approach to adopting mixed reality or other immersive technologies in higher education, highlighting key areas of focus which will be discussed further.



Figure 4: Conceptual model

Curriculum Development and Alignment

Higher education should integrate immersive technologies into the curriculum to maintain subject relevance. An educational advisory board, guided by experts, can modify and align the curriculum accordingly. Once adopted, the board should continuously optimize this approach to keep content relevant.

Infrastructure and Accessibility

To overcome financial barriers, institutions should seek funding and partner with governments and tech providers to lower costs. Equipment must be inclusive for all students, and classrooms redesigned for flexible, interactive learning.

Educator and Student Training

Successful adoption requires training educators to use mixed reality tools and adapt teaching methods effectively. Students also need training to engage with virtual content and personalize their learning.

Addressing Health, Safety, and Inclusivity

Higher education institutions should set safety guidelines for immersive technology use, including usage limits, breaks, and cybersickness management. Content should follow universal design principles and include accessibility features like voice commands and touch interfaces to support all students.

Strategic Partnerships and Collaboration

Higher education institutions should partner with industry leaders for internships and workshops, keeping mixed reality tools relevant and providing practical experience. Regional collaboration can also foster knowledge exchange and resource sharing, enriching the educational ecosystem.

6 Conclusion

This systematic review highlights the potential of mixed, virtual, and augmented reality to enhance learning in higher education through interactive simulations and 3D visualization. These technologies improve engagement, critical thinking, collaboration, motivation, and personalized learning, serving as valuable complements to traditional methods.

However, the review also reveals an apparent inequality in the global adoption of extended reality in higher education. Most of the reviewed studies originate from countries with developed economies, where institutions benefit from their country's advanced infrastructure, reliable connectivity, and financial resources. In comparison, regions in Africa remain underrepresented, both in research output and in adoption of extended reality, mainly due to financial constraints and a lack of infrastructural development.

Future research should address the adoption of extended reality in African higher education, where literature is scarce, to develop frameworks suited to the continent's socio-economic context, to ensure equitable access to innovative learning opportunities globally.

7 References

Alexander, B., Ashford-Rowe, K., Barajas-Murph, N., Dobbin, G., Knott, J., McCormack, M., Pomerantz, J., Seilhamer, R., & Weber, N. (2019). Horizon Report 2019 Higher Education Edition.

Alkhudair, N., Alrayes, F., Alsehli, D., & AlRayes, S. (2024). Virtual reality in experiential pharmacy education: A quasi-experimental study. Saudi Pharmaceutical Journal, 32(5), 102028. https://doi.org/https://doi.org/10.1016/j.jsps.2024.102028

Andalib, S. Y., & Muntazar, M. (2024). Co-Created Virtual Reality (VR) Modules in Landscape Architecture Education: A Mixed Methods Study Investigating the Pedagogical Effectiveness of VR. Education Sciences, 14(6), 553. https://doi.org/https://doi.org/10.3390/educsci14060553

Antón-Sancho, Á., Vergara, D., & Fernández-Arias, P. (2023). Impact of the Digitalization Level on the Assessment of Virtual Reality in Higher Education. International Journal of Online Pedagogy and Course Design, 13(1), 1-19. https://doi.org/https://doi.org/10.4018/IJOPCD.314153

Baltezarevic, I., & Baltezarevic, R. (2024). The role of mixed reality in the education sector. Bastina, 34, 471-480. https://doi.org/10.5937/bastina34-48803

Baniasadi, T., Ayyoubzadeh, S., & Mohammadzadeh, N. (2020). Challenges and Practical Considerations in Applying Virtual Reality in Medical Education and Treatment. Oman Medical Journal, 35, e125-e125. https://doi.org/10.5001/omj.2020.43

Banjar, A., & Campbell, A. G. (2022, 30 May-4 June 2022). Doctoral Colloquium—The Potential of Mixed-Reality Technology for Motivating Dentistry Students in Higher Education. 2022 8th International Conference of the Immersive Learning Research Network (iLRN),

Banjar, A., Xu, X., Iqbal, M. Z., & Campbell, A. (2023). A systematic review of the experimental studies on the effectiveness of mixed reality in higher education between 2017 and 2021. Computers & Education: X Reality, 3, 100034. https://doi.org/https://doi.org/10.1016/j.cexr.2023.100034

Baxter, G., & Hainey, T. (2020). Student perceptions of virtual reality use in higher education [Student perceptions of virtual reality]. Journal of Applied Research in Higher Education, 12(3), 413-424. https://doi.org/https://doi.org/10.1108/JARHE-06-2018-0106

Beetul, A. V., Rajabalee, Y. B., & Santally, M. I. (2023, 31 May-2 June 2023). Augmented, Virtual and Mixed Realities and their potential in Teaching and Learning: A Systematic Literature Review. 2023 IST-Africa Conference (IST-Africa),

Bermejo, B., Juiz, C., Cortes, D., Oskam, J., Moilanen, T., Loijas, J., Govender, P., Hussey, J., Schmidt, A. L., Burbach, R., King, D., Colin, O. C., & Dunlea, D. (2023). AR/VR Teaching-Learning Experiences in Higher Education Institutions (HEI): A Systematic Literature Review. Informatics, 10(2), 45. https://doi.org/https://doi.org/10.3390/informatics10020045

Birt, J., & Cowling, M. (2018). Assessing mobile mixed reality affordances as a comparative visualization pedagogy for design communication. Research in Learning Technology, 26, 1-25. https://doi.org/10.25304/rlt.v26.2128

Birt, J., Štromberga, Z., Cowling, M., & Moro, C. (2018). Mobile Mixed Reality for Experiential Learning and Simulation in Medical and Health Sciences Education. Information (Switzerland), 9, 31. https://doi.org/10.3390/info9020031

Blooma, M. J., Kurian, J. C., Fitzgerald, R., & Goh, D. H.-L. (2022). Students' Learning Experience in a Mixed Reality Environment: Drivers and Barriers. Commun. Assoc. Inf. Syst., 50, 28.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

ChatGPT, O. (2024).

Chen, L., Day, T. W., Tang, W., & John, N. W. (2017). Recent Developments and Future Challenges in Medical Mixed Reality 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), https://doi.ieeecomputersociety.org/10.1109/ISMAR.2017.29

Clarke, V., Braun, V., & Hayfield, N. (2015). Qualitative Psychology: A Practical Guide to Research Methods. Thematic analysis., 3, 222–248.

Dong, W., Zhou, M., Zhou, M., Jiang, B., & Lu, J. (2023). An overview of applications and trends in the use of extended reality for teaching effectiveness: an umbrella review based on 20 meta-analysis studies. The Electronic Library, 41(5), 557-577. https://doi.org/10.1108/EL-11-2022-0257

Duan, F. (2024). Application of Multimedia Technology in PE Teaching in Colleges and Universities. International Journal of Information and Communication Technology Education, 20(1), 1-14. https://doi.org/https://doi.org/10.4018/IJICTE.350184

Dutt, S., Sharma, R., Suyal, P., & Thapliyal, M. (2022, 16-17 Dec. 2022). An Investigative Study of Long-Term Implication of Extended Reality Applications in Educational Pursuits for Learning-Disabled Population: A Recommendation Approach. 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART),

Ewais, A., Zein, S., & Hamed, G. (2022). Need Analysis for Higher Educational Institutions for using Virtual Reality-TESLA Project: Staff Willingness and Readiness for Using VR in Teaching. International Journal of Emerging Technologies in Learning (Online), 17(22), 216-231. https://doi.org/https://doi.org/10.3991/ijet.v17i22.34355

Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. Health Information & Libraries Journal, 26(2), 91-108. https://doi.org/https://doi.org/10.1111/j.1471-1842.2009.00848.x

Hajirasouli, A., & Banihashemi, S. (2022). Augmented reality in architecture and construction education: state of the field and opportunities. International Journal of Educational Technology in Higher Education, 19(1), 39. https://doi.org/10.1186/s41239-022-00343-9

Harjana, M. N. S., Saputra, H. Y., & Tho, C. (2023). A Review of the Potential Use of Mixed Reality Learning Methods in Comparison to Traditional Learning Methods. Procedia Computer Science, 227, 734-742.

https://doi.org/https://doi.org/10.1016/j.procs.2023.10.578

Hensen, B., & Klamma, R. (2022, 30 May-4 June 2022). A Mixed Reality Teaching Course for Formal Higher Education. 2022 8th International Conference of the Immersive Learning Research Network (iLRN),

Hughes, C. E., Stapleton, C. B., Hughes, D. E., & Smith, E. M. (2005). Mixed reality in education, entertainment, and training. IEEE Computer Graphics and Applications, 25(6), 24-30. https://doi.org/10.1109/MCG.2005.139

Hutson, J., & Olsen, T. (2022). Virtual Reality and Art History: A Case Study of Digital Humanities and Immersive Learning Environments. Journal of Higher Education Theory and Practice, 22(2), 50-65. https://www.proquest.com/scholarly-journals/virtual-reality-art-history-case-study-digital/docview/2644080815/se-2?accountid=14717

Ironsi, C. S. (2023). Investigating the use of virtual reality to improve speaking skills: insights from students and teachers. Smart Learning Environments, 10(1), 53. https://doi.org/https://doi.org/10.1186/s40561-023-00272-8

Izard, S. G., Vicent, C. V., Méndez, J. A. J., & Palau, R. (2021). Virtual Reality In Higher Education: An Experience With Medical Students: Research into how Virtual Reality can be used as a powerful training tool for medicine students Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality, Salamanca, Spain. https://doi-org.uplib.idm.oclc.org/10.1145/3434780.3436539

Jiang, X., Liu, X., & Wang, Y. (2024). Utilization of Microsoft HoloLens Augmented Reality Technology in Health Care Virtual Simulation. Procedia Computer Science, 243, 278-285. https://doi.org/https://doi.org/10.1016/j.procs.2024.09.035

Jiawei, W., & Mokmin, N. A. M. (2023). Virtual reality technology in art education with visual communication design in higher education: a systematic literature review. Education and Information Technologies, 28(11), 15125-15143. https://doi.org/https://doi.org/10.1007/s10639-023-11845-y

Jin, Q., Liu, Y., Yarosh, S., Han, B., & Qian, F. (2022). How Will VR Enter University Classrooms? Multi-stakeholders Investigation of VR in Higher Education Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, <conf-loc>, <city>New Orleans</city>, <state>LA</state>, <country>USA</country>, </conf-loc>. https://doi-org.uplib.idm.oclc.org/10.1145/3491102.3517542

Kezia Herman, M., & Bordoloi, R. (2024). Augmented reality for inclusive growth in education: the challenges. Asian Association of Open Universities Journal, 19(1), 88-100. https://doi.org/https://doi.org/10.1108/AAOUJ-09-2023-0113

Lai, J. W., & Cheong, K. H. (2022). Adoption of Virtual and Augmented Reality for Mathematics Education: A Scoping Review. IEEE Access, 10, 13693-13703. https://doi.org/10.1109/ACCESS.2022.3145991

Laine, T. H., & Lee, W. (2024). Collaborative Virtual Reality in Higher Education: Students' Perceptions on Presence, Challenges, Affordances, and Potential. IEEE Transactions on Learning Technologies, 17, 280-293. https://doi.org/10.1109/TLT.2023.3319628

Li, D., Chao, Y., & Gu, Y. (2021). Research on College Physical Education and Sports Training Based on Virtual Reality Technology. Mathematical Problems in Engineering, 2021. https://doi.org/https://doi.org/10.1155/2021/6625529

Li, W. (2023, 15-17 Sept. 2023). Simulating Virtual Environment and Experience for Training, Exergaming, and Edutainment in eXtended Reality (XR): A Survey. 2023 International Conference on Computer Applications Technology (CCAT),

Lohfink, M. A., Miznazi, D., Stroth, F., & Müller, C. (2022, 17-21 Oct. 2022). Learn Spatial! Introducing the MARBLE-App - A Mixed Reality Approach to Enhance Archaeological Higher Education. 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct),

Marks, B., & Thomas, J. (2022). Adoption of virtual reality technology in higher education: An evaluation of five teaching semesters in a purpose-designed laboratory. Education and Information Technologies, 27(1), 1287-1305. https://doi.org/https://doi.org/10.1007/s10639-021-10653-6

McMillan, K., Flood, K., & Glaeser, R. (2017). Virtual reality, augmented reality, mixed reality, and the marine conservation movement. Aquatic Conservation: Marine and Freshwater Ecosystems, 27, 162-168. https://doi.org/10.1002/aqc.2820

Meccawy, M. (2023). Teachers' prospective attitudes towards the adoption of extended reality technologies in the classroom: interests and concerns. Smart Learning Environments, 10(1), 36. https://doi.org/https://doi.org/10.1186/s40561-023-00256-8

Misirlis, N. Nikolaidis, Y and Sabidussi, A. (2023). Should I use metaverse or not? An investigation of university students behavioral intention to use MetaEducation technology. IADIS International Journal on Computer Science and Information Systems, 18(1), 18-29.

Moser, T., Wolfartsberger, J., Sorko, S. R., & Naim, B. A. (2024, 6-10 May 2024). Extended Reality Based Education and Training for Human-Centric Industry 5.0 Skill Enhancement. NOMS 2024-2024 IEEE Network Operations and Management Symposium,

Munfaridah, N., Avraamidou, L., & Goedhart, M. (2021). The Use of Multiple Representations in Undergraduate Physics Education: What Do we Know and Where Do we Go from Here? Eurasia Journal of Mathematics, Science and Technology Education, 17, em1934. https://doi.org/10.29333/ejmste/9577

Nesenbergs, K., Abolins, V., Ormanis, J., & Mednis, A. (2021). Use of Augmented and Virtual Reality in Remote Higher Education: A Systematic Umbrella Review. Education Sciences, 11(1), 8. https://doi.org/https://doi.org/10.3390/educsci11010008

Obeidallah, R., Ayat Al, A., & Qutishat, D. (2023). Challenges of Extended Reality Technology in Higher Education: A Review. International Journal of Emerging Technologies in Learning (Online), 18(14), 39-50. https://doi.org/https://doi.org/10.3991/ijet.v18i14.39871

Pan, X., Zheng, M., Xu, X., & Campbell, A. G. (2021). Knowing Your Student: Targeted Teaching Decision Support Through Asymmetric Mixed Reality Collaborative Learning. IEEE Access, 9, 164742-164751. https://doi.org/10.1109/ACCESS.2021.3134589

Pellas, N., Mystakidis, S., & Kazanidis, I. (2021). Immersive Virtual Reality in K-12 and Higher Education: A systematic review of the last decade scientific literature. Virtual Reality, 25(3), 835-861. https://doi.org/https://doi.org/10.1007/s10055-020-00489-9

Perçuku, A., & Minkovska, D. (2022, 30 May-2 June 2022). The Application of eXtended Reality for Learning in Higher Education. 2022 10th International Scientific Conference on Computer Science (COMSCI),

Perlin, Z. H. a. K. (2020). Exploring the Effectiveness of Face-to-face Mixed Reality for Teaching with Chalktalk.

Peters, M., Simó, L. P., Amatller, A. M., Liñan, L. C., & Kreitz, P. B. (2021, 17 May-10 June 2021). Toward A User-Centred Design Approach for AR Technologies in Online Higher Education. 2021 7th International Conference of the Immersive Learning Research Network (iLRN),

Petruse, R. E., Grecu, V., Gakić, M., Jorge Martin, G., & Mara, D. (2024). Exploring the Efficacy of Mixed Reality versus Traditional Methods in Higher Education: A Comparative Study. Applied Sciences, 14(3), 1050. https://doi.org/https://doi.org/10.3390/app14031050

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. Computers & Education, 147, 103778. https://doi.org/https://doi.org/10.1016/j.compedu.2019.103778

Salloum, S., Amina Al, M., Khaled Younis, A., Shwedeh, F., Aburayya, A., Mohammed Rasol Al, S., & Rana Saeed, A.-M. (2023). Sustainability Model for the Continuous Intention to Use Metaverse Technology in Higher Education: A Case Study from Oman. Sustainability, 15(6), 5257. https://doi.org/https://doi.org/10.3390/su15065257

Sofiadin, A. (2023, 25-27 Oct. 2023). Extended Reality for Experiential Learning from Students' Perspective. 2023 IEEE 8th International Conference on Engineering Technologies and Applied Sciences (ICETAS),

Staddon, R. (2020). Bringing technology to the mature classroom: age differences in use and attitudes. International Journal of Educational Technology in Higher Education, 17, 11. https://doi.org/10.1186/s41239-020-00184-4.

Steyn AA, van Slyke C, Dick G, Twinomurinzi H, Amusa LB (2024a) Student intentions to continue with distance learning post-COVID: An empirical analysis. PLoS ONE 19(1): e0293065. https://doi.org/10.1371/journal.pone.0293065.

Steyn, A. A., Adebesin, F., & Mennega, N. (2024b). Empowering Business Students for the Fourth Industrial Revolution: A Hands-On Pedagogical Approach. Proceedings of the 53rd Annual Conference of the Southern African Computer Lecturers' Association (SACLA 2024), Gqeberha, South Africa.

Suryodiningrat, S. P., Prabowo, H., Meyliana, & Hidayanto, A. N. (2021, 24-25 Nov. 2021). Mixed Reality System for Teaching and Learning: A Systematic Literature Review. 2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE),

Tang, Y. M., Au, K. M., Lau, H. C. W., Ho, G. T. S., & Wu, C. H. (2020). Evaluating the effectiveness of learning design with mixed reality (MR) in higher education. Virtual Real., 24(4), 797–807. https://doi.org/10.1007/s10055-020-00427-9

Valladares Ríos, L., Acosta-Diaz, R., & Santana-Mancilla, P. C. (2023). Enhancing Self-Learning in Higher Education with Virtual and Augmented Reality Role Games: Students' Perceptions. Virtual Worlds, 2(4), 343-358. https://www.mdpi.com/2813-2084/2/4/20

van Dinter, R., Tekinerdogan, B., & Catal, C. (2021). Automation of systematic literature reviews: A systematic literature review. Information and Software Technology, 136, 106589. https://doi.org/https://doi.org/10.1016/j.infsof.2021.106589

Wang, Q., & Li, Y. (2024). How virtual reality, augmented reality and mixed reality facilitate teacher education: A systematic review. Journal of Computer Assisted Learning, 40(3), 1276-1294. https://doi.org/https://doi.org/10.1111/jcal.12949

Between Innovation and Reality: Teachers' Attitudes Toward ICT in Teaching

Krešimir Pavlina, Ana Pongrac Pavlina, Anita Modrušan
University of Zagreb
Department of Information and Communication Sciences / Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10000 Zagreb, Croatia
{kpavlina,apongrac,amodrusan}@ffzg.unizg.hr

Abstract. *This study explores primary school teachers' experiences and perceptions about implementation of ICT in everyday educational practice. Between April and May 2023, 132 teachers (M age=43.44; SD=10.04; 93.9% female) were asked to rate their agreement with 17 statements about the use of digital technology, artificial intelligence and digital content in teaching. Study was designed as a cross-sectional study with a convenience sample. Results show that most teachers are confident using digital technologies and recognize their benefits in personal and professional life. Yet there are still major challenges - insufficient time to develop digital materials, unstable access to student devices and a preference for traditional teaching resources over digital content. Attitudes toward artificial intelligence are mixed, with some teachers expressing enthusiasm and others voicing concerns about its potential impact on job security. These findings indicate both teacher enthusiasm as well as barriers to digitalization of primary education and call for improved infrastructure and targeted professional support for successful implementation of ICT in education.*

Keywords. primary education, technology integration, barriers to digitalization, educational innovation, teachers' perceptions

1 Introduction

Digitalisation and artificial intelligence (AI) in primary education have created new opportunities as well as challenges for educators and students. In just the past decade, digital tools have transformed traditional classroom instruction into dynamic, student-centred experiences (Walter, 2024). The COVID-19 pandemic further accelerated this digital transformation, highlighting both the potential of technology to sustain educational continuity and the persistent barriers to equitable access and effective implementation.

Meta-analysis by Soriano-Sánchez (Soriano-Sánchez, 2025) demonstrates the significant positive impact of Information and Communication Technologies (ICT) on Natural Science learning in primary school students, particularly in supporting those with special educational needs. The findings highlight that interactive and adaptive digital technologies enhance academic performance, foster active participation, and increase intrinsic motivation by personalizing learning experiences and addressing diverse student needs. Interactive platforms, educational software, and AI-powered applications individualize instruction, formative assessment and real-time feedback to meet a variety of learning styles and develop critical 21st century skills including creativity, problem-solving and digital literacy. AI technologies have shown promise in primary education for personalizing learning, automating administrative tasks and providing adaptive support for students with varying abilities.

Digital technology adoption in primary education is dependent on several interrelated factors. The most important predictors of technology uptake and effective classroom use is teachers' digital competence and attitudes (Yu et al., 2023). Pezo reviewed recent literature on the adoption of ICT by primary school teachers, finding that while ICT use in primary schools has mixed effects on academic performance, its successful implementation depends on equitable access to devices and the internet, development of digital skills, curricular adaptation and adequate infrastructure. The study warns that without these conditions, poor ICT integration can hinder rather than support learning, potentially leading to frustration among the educational community (Pezo, 2024). The relationship between ICT use in schools and specific cognitive outcomes like memory, attention, and executive functioning is complex and multi-faceted. While ICT can enhance engagement and provide learning experiences that may support memory recall and motivation (Bautista-Vallejo et al, 2020), concerns persist regarding its potential to diminish sustained attention and impact executive functions such as working memory and inhibitory control (Maeneja et al, 2025). Ultimately, the impact often depends on the quality of the digital content, the pedagogical approach and whether technology is integrated thoughtfully to complement, rather than simply replace, traditional learning methods. Socioeconomic disparities can

further exacerbate the digital divide, affecting both teachers' and students' opportunities to benefit from educational technology (Wang et al., 2024).

Another level of complexity is introduced by attitudes regarding AI in education. Even though a lot of educators are aware of how AI might improve instruction, worries about job security, moral dilemmas and the changing dynamics of teacher-student relationships still exist. According to recent studies, it is critical to have professional development and policy frameworks that address these issues and give educators the tools they need to successfully and ethically incorporate AI into their teaching (Garcia et al., 2025).

Contextual variables like school leadership, digital culture, and digital content quality also influence technology integration. An effective professional development model characterized by sustained, collaborative, and contextually relevant training has been identified as essential for enhancing teachers' capacities to use digital tools effectively (Røkenes et al., 2022). More broadly, the existence of digital as well as traditional teaching resources reflects a pragmatic attitude of primary school teachers who tend to see technology as an addition to rather than an alternative to established pedagogical practices.

While the overarching benefits and common barriers to digital technology integration in education are extensively documented across European and global contexts (e.g., as highlighted by comprehensive frameworks like DigCompEdu (Ghomi et al., 2019), the nuances of implementation and the lived realities of teachers can vary significantly based on specific national and regional contexts. General surveys, while valuable for broad comparisons, often mask the granular challenges and unique attitudinal complexities faced by educators in diverse educational ecosystems.

This study aims to bridge this analytical gap by offering a granular perspective from primary school teachers in Croatia. Despite Croatia's efforts towards digital transformation in education (e.g., through initiatives like the "e-Schools" program), specific local conditions can shape the actual "innovation-to-reality" pathway.

Much literature exists on digital technology adoption in education, but gaps remain in our understanding of the holistic experiences of primary school teachers - including attitudes, competences, infrastructural barriers and the emerging influence of AI. Addressing such gaps informs educational policy, guides professional development and ensures equitable digital transformation of primary education.

The aim of this research was to examine primary school teachers' attitudes toward various aspects of ICT integration in classroom instruction. This study investigates the experiences, perceptions, and challenges primary school teachers face regarding digital technology in their daily school activities, specifically exploring their current technology use, attitudes toward digital innovation and AI, and prevailing barriers to digital integration within primary school classrooms.

2 Methodology

The study addresses the integration of ICT in primary education. It investigates teachers' experiences, perceptions and challenges regarding digital technology application in everyday school activities.

2.1 Participants

The study employed a cross-sectional design with a convenience sample of 132 primary school teachers who completed the survey between April and May 2023. The survey was distributed through teacher groups in several primary schools in Karlovac County (Slunj, Eugen Kvaternik in Rakovica, Švarča, and Dragojla Jarnević Primary Schools) and via teacher groups on Facebook. The predominant demographic of respondents in the sample was female (93.9%), exceeding the national average of 85%. The teachers who participated in the survey had the following age distribution: 9.8% were between the ages of 21 and 30; 31.1% were between the ages of 31 and 40; 31.1% were between the ages of 41 and 50; 25% were between the ages of 51 and 60; and 3% were older than 60 years. The average age was 43.44 years with a standard deviation of 10.04 years.

The distribution of the surveyed teachers' teaching experience was as follows: 12.9% of teachers had zero to five years of experience, 15.2% had six to ten years, 37.1% had eleven to twenty years, 22.0% had twenty to thirty, and 12.9% had thirty to forty years of teaching experience. On average, teachers had 17.03 years of teaching experience with a standard deviation of 9.94 years.

2.2 Instrument

The research employed a quantitative survey instrument consisting of 17 statements rated on a 5-point Likert scale (1=Strongly disagree to 5=Strongly agree). The statements were constructed to assess both attitudes and reported behaviours regarding ICT use.

An exploratory factor analysis was conducted on these 17 items to examine attitudes and practices related to ICT in teaching, revealing three factors. The internal consistency of Factor 1 (reflecting the usefulness and frequency of digital technology in the classroom) was excellent, with McDonald's ω of 0.89 (95% CI: 0.86, 0.92) and Cronbach's α of 0.88 (95% CI: 0.85, 0.92). Similarly, Factor 2 (capturing teachers' preference for and creation of digital teaching materials) exhibited very good internal consistency, with McDonald's ω of 0.84 (95% CI: 0.79, 0.88) and Cronbach's α of 0.84 (95% CI: 0.79, 0.89).

2.3 Procedure

The study was designed as a cross-sectional survey. Data were collected through voluntary participation via self-administered questionnaires distributed to teachers in April and May 2023. Participation in the survey was entirely voluntary and all responses were collected anonymously to ensure confidentiality. Prior to participation, all respondents were informed about the study's purpose, data usage, and their right to withdraw at any time, with their completion of the survey signifying their informed consent.

3 Results

This section presents the key findings from our survey, detailing primary school teachers' reported patterns of ICT use and their attitudes towards it. The subsequent figures and tables provide a comprehensive overview of the collected data, offering granular insights into the innovation-to-reality pathway in Croatian primary education.

Table 1 presents the descriptive statistics for all 17 survey items. Among these, the item #1 "Digital technology makes my life easier" (Mean = 4.29, SD = 0.82) received the highest average score, closely followed by item #3 "I think I am good at handling new digital technologies for both private and professional purposes" (Mean = 4.26, SD = 0.83). This indicates a strong positive perception among teachers regarding the general use of digital technology and their confidence in using it. Conversely, the item #12 "My students regularly bring school tablets that are functional and charged, so we can use digital technology in class without any problems" (Mean = 2.01, SD = 1.23) had the lowest mean score, highlighting a significant and pervasive challenge related to student device reliability. Similarly, item #16 "I often conduct digital knowledge assessments" (Mean = 2.11, SD = 1.24) also scored very low, suggesting limited adoption of digital assessment tools.

Table 1. Descriptive statistics

Item	Statement	Mean	Std. deviation	Skewness
1.	Digital technology makes my life easier	4.29	0.82	-1.35
2.	I often use digital technology for relaxation and entertainment	3.93	1.03	-0.85
3.	I think I am good at handling new digital technologies for both private and professional purposes	4.26	0.83	-1.42
4.	I am pleased with the progress of artificial intelligence	2.93	1.21	-0.11
5.	I think that the development of artificial intelligence could threaten my job	2.57	1.24	0.23
6.	I often use digital technology in the teaching process	4.05	0.90	-0.74
7.	I believe that digital technology significantly improves the teaching process	3.89	0.96	-0.68
8.	It is easier to conduct classes with the use of digital technology	3.86	1.00	-0.79
9.	Students learn more when the material is taught and practiced with the help of digital technology	3.43	1.06	-0.48
10.	I have enough time to design digital teaching materials for my students	2.79	1.22	0.13
11.	I like to explore and use new digital tools that can improve my teaching	3.47	1.21	-0.32
12.	My students regularly bring school tablets that are functional and charged, so we can use digital technology in class without any problems	2.01	1.23	1.08
13.	I use digital technology every day in the teaching process	3.66	1.16	-0.59
14.	I often use ready-made digital materials provided by the textbook publisher I use (IZZI, e-sfera, and similar)	3.58	1.30	-0.54
15.	I use digital teaching materials (games, quizzes) more frequently than workbooks and worksheets to practice the content.	3.11	1.15	-0.15
16.	I often conduct digital knowledge assessments	2.11	1.24	0.86
17.	I give preference to digital teaching content over traditional ones	2.50	1.23	0.35
	Factor 1 - Use of digital technology in the classroom	*3.82*		
	Factor 2 - Preference for use and creation of digital teaching materials	*2.80*		

An exploratory factor analysis was conducted on 17 Likert-type items to examine attitudes and practices related to ICT in teaching. The Kaiser-Meyer-Olkin measure was 0.896, indicating excellent suitability for factor analysis. Bartlett's test was significant ($\chi^2(136) = 1109.93$, $p < .001$), confirming that inter-item correlations were sufficient. Table 2 presents the pattern matrix loadings obtained from this analysis.

Table 2. Results of factor analysis

	Factor 1	Factor 2	Factor 3	Uniqueness
7. I believe that digital technology significantly improves the teaching process	1.059			0.123
8. It is easier to conduct classes with the use of digital technology	0.836			0.238
9. Students learn more when the material is taught and practiced with the help of digital technology	0.667			0.534
6. I often use digital technology in the teaching process	0.651			0.274
14. I often use ready-made digital materials provided by the textbook publisher I use (IZZI, e-sfera, and similar)	0.528			0.788
13. I use digital technology daily in the teaching process	0.505			0.496
1. Digital technology makes my life easier	0.441			0.371
16. I often conduct digital knowledge assessments		0.987		0.308
15. I practice the material more often using digital teaching materials (quizzes, games) than with workbooks and worksheets		0.736		0.494
17. I give preference to digital teaching content over traditional ones		0.691		0.399
10. I have enough time to design digital teaching materials for my students		0.569		0.568
11. I like to explore and use new digital tools that can improve my teaching		0.518		0.519
3. I think I am good at handling new digital technologies for both private and professional purposes			1.028	0.085
2. I often use digital technology for relaxation and entertainment				0.824
4. I am pleased with the progress of artificial intelligence				0.660
5. I think that the development of artificial intelligence could threaten my job				0.993
12. My students regularly bring school tablets that are functional and charged, so we can use digital technology in class without any problems				0.894

Notes:
Factor 1 – Use of digital technology in the classroom
Factor 2 – Preference for use and creation of digital teaching materials

Using principal axis factoring and Promax rotation, three factors were extracted, explaining 45.9% of the total variance. The pattern matrix loadings are displayed in Table 2. Factor 1 reflected the usefulness and frequency of use of digital technology in the classroom. Factor 2 captured teachers' preference for and creation of digital teaching materials. Item 3 ("I think I am good at handling new digital technologies for both private and professional purposes"), which initially loaded as a separate factor, exhibited significant negative skewness (Skewness = -1.42), suggesting a strong ceiling effect where most teachers reported very high digital self-efficacy. From a psychometric standpoint, a single item cannot constitute a valid factor. Consequently, due to its limited discriminatory power and psychometric invalidity as a standalone factor, Item 3 will be excluded from subsequent factor and correlational analyses.

Table 3. Correlation analysis

	Factor 1 *(Use of digital technology in the classroom)*	Factor 2 *(Preference for use and creation of digital teaching materials)*	Age	Teaching experience
Factor 1 *(Use of digital technology in the classroom)*	1.00	0.71	-0.01	0.03
Factor 2 *(Preference for use and creation of digital teaching materials)*	0.71	1.00	0.15	0.07
Age	-0.01	0.15	1.00	0.79
Teaching experience	0.03	0.07	0.79	1.00

The correlation analysis reveals strong positive interrelationships among the identified factors. Factor 1 showing a particularly high correlation with Factor 2 (r=0.71). Conversely, there were negligible correlations between these technology-related factors and demographic variables such as age and teaching experience. Notably, as expected, age and teaching experience demonstrated a very strong positive correlation (r=0.79), indicating that older teachers generally possess more years of experience.

The data for statement "Digital technology makes my life easier" in the Fig. 1 indicates widespread agreement with 53 teachers who agreed and 61 strongly agreeing (86.5%), while only 3 respondents disagreed (2.3%). This indicates that digital technology is considered a facilitator for personal and professional life. Improved efficiency, easier communication and simplified administrative tasks help teachers see technology as an asset in everyday life.

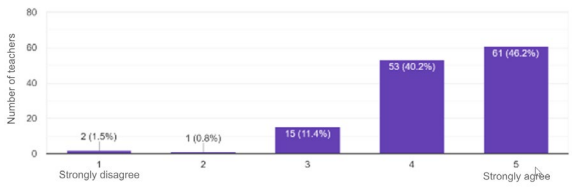


Figure 1. Digital technology makes my life easier

Answers in Fig. 2 present that 69.6% of teachers often use digital technology for leisure (46 agree, 46 strongly agree), with only 10 teachers disagreeing. This suggests that technology is also a form of personal enjoyment and relaxation. This incorporation into everyday life could also make teachers feel at ease with digital platforms and thus indirectly support their use in education contexts.

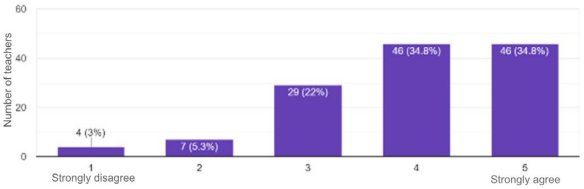


Figure 2. I often use digital technology for relaxation and entertainment

From answers presented in Fig 3. it is evident that 87.9% of teachers have high confidence in their digital skills, with 59 agreeing and 57 strongly agreeing while only 5 teachers disagreed (3.8%). A high level of confidence is essential for the successful integration of technology in the classroom because competent teachers are more inclined to try out and use new technologies, which improves their instruction and increases student engagement.

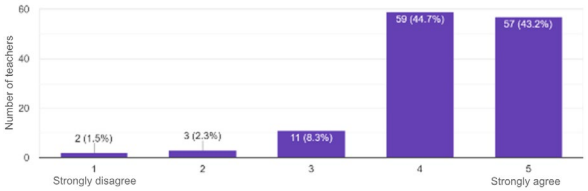


Figure 3. I think I am good at handling new digital technologies for both private and professional purposes

Responses presented in Fig. 4 to a statement "I am pleased with the progress of artificial intelligence" are mixed: 43 disagree or strongly disagree (32.6%), 46 are neutral (34.8%) and 43 agree or strongly agree (32.6%). This ambivalence may reflect uncertainty about AI's educational applications or scepticism about its broader societal impact. Some teachers may feel unprepared for rapid changes or lack information about AI's benefits and risks.

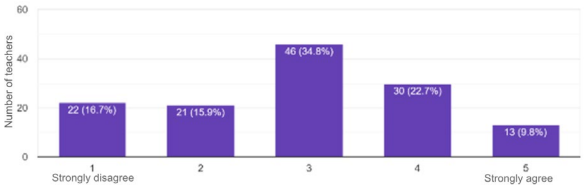


Figure 4. I am pleased with the progress of artificial intelligence

A significant portion 44.7% (36 strongly disagree, 23 disagree) do not feel threatened by AI, but 34.1% (45 teachers) are neutral and 21.2% (28 teachers) express concerns. This indicates underlying anxiety about automation and job security, even if most teachers do not see AI as an immediate threat. Ongoing professional development and clear communication about the role of AI in education could help address these concerns.

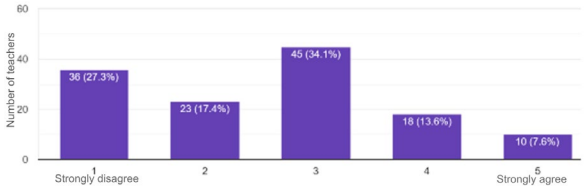


Figure 5. I think that the development of artificial intelligence could threaten my job

Frequent use of digital technology in the teaching process is reported by 75.0% of teachers (51 agreeing and 48 strongly agreeing), with only 5.3% disagreeing. This shows that digital tools are well integrated into teaching routines, likely enhancing lesson delivery, student engagement, and access to diverse resources. However, the small number of dissenters may reflect infrastructural or attitudinal barriers in specific contexts.

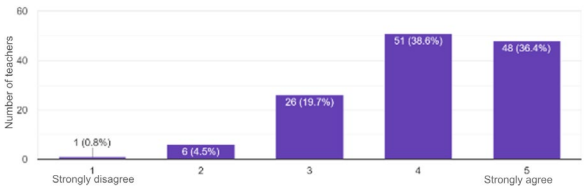


Figure 6. I often use digital technology in the teaching process

Data in Fig. 7 presents that 69.7% of teachers (53 agree, 39 strongly agree) believe technology enhances teaching, with only 8.3% disagreeing. This supports the view that digital tools contribute to more effective, dynamic and interactive instruction, potentially leading to better student outcomes and more efficient classroom management.

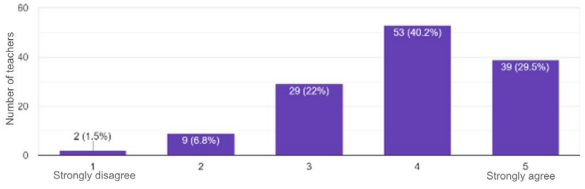


Figure 7. I believe that digital technology significantly improves the teaching process

Fig. 8 presents that 70.5% of teachers (55 agree, 38 strongly agree) think that technology makes teaching easier, while only 10.6% disagree. This suggests that digital tools help streamline lesson planning, resource sharing and classroom activities, reducing teacher workload and improving organization.

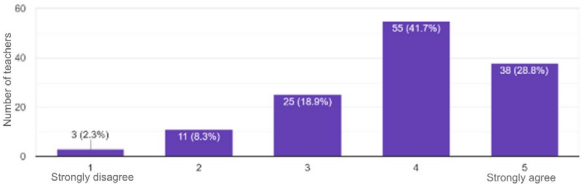


Figure 8. It is easier to conduct classes with the use of digital technology

On statement "Students learn more when the material is taught and practiced with the help of digital technology" opinions are more divided: 50.8% (47 agree, 20 strongly agree), 16.7% (22 teachers) disagree and 32.6% are neutral. While many see benefits for student learning, others are cautious, possibly due to variability in student access, digital literacy, or the quality of digital resources.

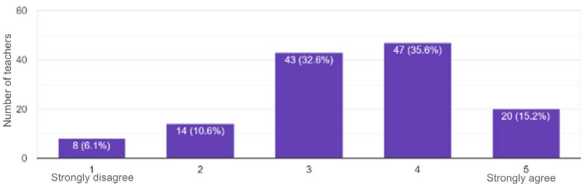


Figure 9. Students learn more when the material is taught and practiced with the help of digital technology

Data in Fig. 10 display that time constraints are evident, with 40.2% teachers disagreeing that they have enough time to design digital teaching materials for their students (24 strongly disagreeing, 29 disagreeing), 32.6% neutral and only 27.3% agreeing with that statement. This barrier may limit the creation of tailored digital resources and slow the pace of technology integration, emphasizing the need for institutional support and better workload management.

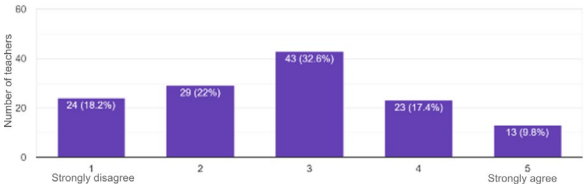


Figure 10. I have enough time to design digital teaching materials for my students

Teachers' enthusiasm for technology exploration is moderate: 51.5% agree that they like to explore and use new digital tools, but 23.5% disagree or strongly disagree and 25.0% are neutral. This split may reflect differences in personal motivation, available training, or perceived relevance of new tools. Professional development and peer support could encourage more teachers to innovate and be willing to explore the use of new digital tools in their everyday teaching.

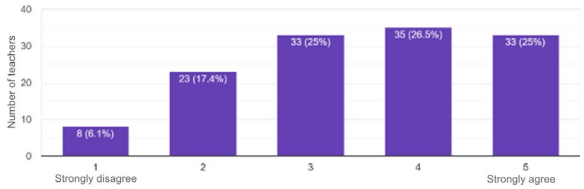


Figure 11. I like to explore and use new digital tools that can improve my teaching

A clear majority of teachers (89.3%; 65 strongly disagree, 25 disagree, 28 neutral) report unreliable student devices, with only 10.6% (14 teachers) agreeing. This is a significant infrastructural challenge, as access to functional technology is essential for effective digital teaching. Addressing this issue is critical for equitable technology integration.

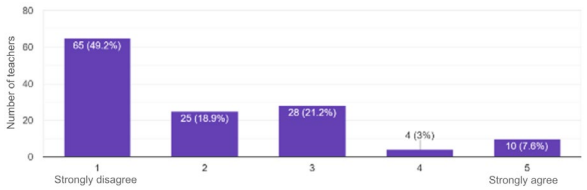


Figure 12. My students regularly bring school tablets that are functional and charged, so we can use digital technology in class without any problems

Majority of teachers use digital technology every day in the teaching process (85.6%; 38 agree, 38 strongly agree), but 19 disagree or strongly disagree (14.4%). This suggests that while technology is widely used, some teachers or schools still face obstacles, possibly related to resources, training, or attitudes.

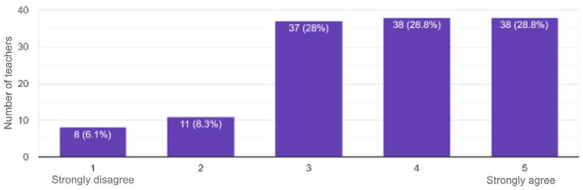


Figure 13. I use digital technology every day in the teaching process

Many teachers (56.8 %; 33 agree, 42 strongly agree) rely on publisher-provided digital content, showing the importance of accessible, high-quality resources. This reliance may be due to time constraints or a lack of confidence in creating original materials.

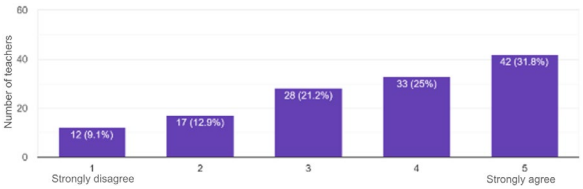


Figure 14. I often use ready-made digital materials provided by the textbook publisher I use (IZZI, e-sfera, and similar)

Responses are mixed: 33 agree, 16 strongly agree, 37 disagree or strongly disagree, and 46 neutral. There is no clear preference, reflecting varied teaching styles and possibly differences in digital teaching materials quality or subject matter suitability.

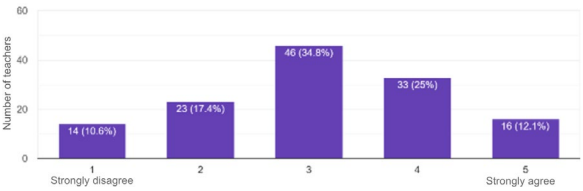


Figure 15. I use digital teaching materials (games, quizzes) more frequently than workbooks and worksheets to practice the content

Most teachers (65.2%; 58 strongly disagree, 28 disagree) do not frequently use digital assessments, with only 19 agreeing (14.4%). This may be due to technical barriers, lack of training, or preference for traditional assessment methods. Increasing support for digital assessment could enhance formative feedback and student tracking.

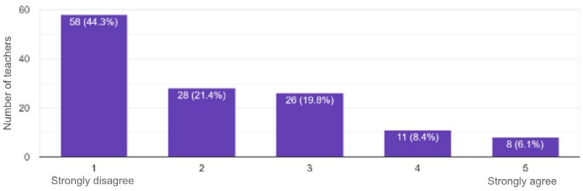


Figure 16. I often conduct digital knowledge assessments

Almost half (48.5%; 37 strongly disagree, 27 disagree) prefer traditional resources, with only 25 agreeing (18.9%). Digital content is seen as supplementary rather than a replacement, underscoring the enduring value of conventional materials, especially in primary education where tactile and interpersonal experiences are crucial.

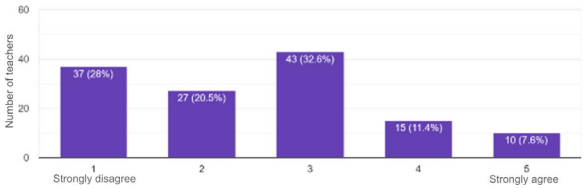


Figure 17. I give preference to digital teaching content over traditional ones

4 Discussion

The survey findings describe primary school teachers' experiences, attitudes and barriers to integrating digital technology and artificial intelligence in education. Quantitative data analysis reveals several central themes that affect educational policy, school leadership, professional development and future research.

Our results indicate a high degree of acceptance and self-efficacy among primary school teachers regarding digital technology. Specifically, teachers overwhelmingly agreed that "Digital technology makes my life easier" (Mean=4.29) and reported high confidence in their digital skills, as shown by their agreement with "I think I am good at handling new digital technologies for both private and professional purposes" (Mean=4.26). This aligns with findings by Yu et al. (2023), who identified teachers' digital competence and positive attitudes as crucial predictors of technology uptake. Furthermore, the frequent use of digital technology for personal relaxation and entertainment (Mean=3.93) suggests a comfort level that likely facilitates its professional adoption. This high level of digital self-efficacy represents a strong foundation for continued digital innovation in schools, empowering teachers to explore and adapt to new technological tools.

Teachers largely agree that digital technology improves the teaching process, makes lesson delivery easier and is regularly used in classroom practice. Many recognize the potential for digital tools to enhance student learning, although this perception is less unanimous, with a substantial minority expressing neutrality or skepticism. This suggests that while the infrastructure for digital integration is present, the perceived pedagogical value may depend on subject area, student readiness or the quality of digital resources available.

Barriers still exist despite high levels of digital involvement. The two biggest ones are infrastructure issues and timing restrictions. The great usage of publisher resources may be explained by the fact that nearly all teachers state that they don't have enough time to create their own digital teaching materials. Such reliance demonstrates the need for quality, curriculum-aligned digital content from educational publishers, but also for greater institutional support for instructors in developing and customizing resources for their context.

Despite the high levels of digital involvement reported by teachers, significant infrastructural barriers persist, notably concerning student device reliability. Our findings indicate that nearly 90% of teachers reported issues with students regularly bringing functional and charged school tablets to class (Mean = 2.01). This directly impedes the delivery of planned digital lessons and prevents the full realization of digital education's potential. Such challenges align with the broader literature on ICT adoption, which consistently identifies equitable access to devices and adequate infrastructure as critical prerequisites for successful technology implementation (Pezo, 2024). Furthermore, this structural deficit risks exacerbating existing educational inequalities between students with and without reliable access to technology, a concern widely documented in research on the digital divide, especially as it relates to socioeconomic disparities (Wang et al., 2024). When technology integration is hindered by such fundamental issues, it can lead to frustration within the educational community, undermining efforts to enhance learning through digital tools (Pezo, 2024).

While digital technology is widely used in Croatian primary schools, as evidenced by teachers' frequent reported use, the data unequivocally suggests that traditional teaching materials remain deeply valued, serving as a complement rather than being supplanted by digital content. Most teachers reported not giving preference to digital content over conventional resources (Mean = 2.50), and digital assessments are not yet routine (Mean = 2.11). This aligns with a growing body of literature that advocates for a balanced, blended pedagogical approach, recognizing that digital tools serve to enhance and augment, rather than simply replace, established educational practices (Vidal-Esteve, M. I., & Martín-Gómez, S., 2023).

The data suggests that, for many educators, digital tools serve as a supplement rather than a replacement for established pedagogical practices. This coexistence highlights the enduring importance of tactile, interpersonal, and contextually rich learning experiences, especially in primary education.

The study's findings reveal a divided stance among primary school teachers regarding artificial intelligence (AI). While a segment expresses optimism and pleasure with AI's progress (Mean = 2.93, with 32.6% agreeing or strongly agreeing), a notable portion remains neutral or expresses concerns about potential job security threats (Mean = 2.57, with 21.2% expressing concerns). This ambivalence reflects

broader societal uncertainties about AI's evolving role in the workforce and within educational contexts. Research shows that educators, despite acknowledging AI's potential, often harbor concerns about job displacement, ethical implications, and the changing dynamics of human-computer interaction in teaching (Cojean et al, 2023). As highlighted by Garcia et al. (2025) in their systematic review, addressing these concerns is critical. The observed mixed perceptions underscore the urgent need for targeted professional development and clear policy frameworks that not only educate teachers about AI's practical applications and limitations but also alleviate anxieties by demonstrating how AI can augment, rather than replace human educators, thereby fostering successful and ethical integration.

A particularly noteworthy finding, challenging common assumptions about technology adoption, is the lack of significant correlation between teachers' age or teaching experience and their positive attitudes towards ICT integration (Factor 1: Use of digital technology in the classroom; Factor 2: *Preference for use and creation of digital teaching materials*). This outcome is especially interesting as it runs counter to a prevalent perception that older, more experienced individuals, including educators, might be less knowledgeable or more sceptical about new technologies like ICT and AI. Instead, our data from primary school teachers in Croatia suggest that engagement with and positive perceptions of digital tools are not solely confined to younger, less experienced educators.

This finding aligns with more nuanced views in the literature, including research by Šabić, Baranović, & Rogošić (2022), which, while identifying gender and age interaction effects, also points to the high self-efficacy of teachers in using ICT, even among older cohorts. Their work suggests that experience, though often correlated with age, does not inherently dictate lower digital self-efficacy or less positive attitudes. This implies that, regardless of their career stage, primary school teachers in Croatia are actively embracing digital tools. This is a positive indicator for sustainable digital transformation, as it suggests that well-designed training and supportive environments can empower all teachers, rather than just a specific demographic cohort, to effectively leverage technology in their classrooms. Future research could explore the specific mechanisms or types of support that contribute to this age-independent integration of ICT.

Despite its strengths, this study has several limitations that warrant consideration. The most significant is the sampling methodology. The use of a convenience sample, primarily distributed through teacher groups in specific primary schools and via Facebook, introduces the potential for self-selection bias. It is plausible that teachers who are more digitally engaged, comfortable with online surveys, or already possess a more positive predisposition towards technology were more inclined to participate. This could lead to an overrepresentation of technologically enthusiastic teachers in our sample, potentially inflating the reported levels of digital competence, comfort, and positive attitudes towards ICT integration. Consequently, the findings, particularly those related to the prevalence of technology use and positive perceptions, may not be fully generalizable to the broader population of primary school teachers in Croatia, especially those who are less digitally inclined or have limited access to online teacher communities.

Another limitation pertains to the cross-sectional design of the study. While it provides a snapshot of attitudes at a specific point in time (April-May 2023), it cannot capture the evolution of teachers' attitudes or the long-term impact of ICT integration. Longitudinal studies would offer deeper insights into how these perceptions change over time, especially with ongoing technological advancements and policy shifts.

Finally, while the survey instrument covered several domains, the reliance on self-reported data means that the findings reflect teachers' perceptions rather than direct observations of their classroom practices. There might be discrepancies between reported and actual behaviours. For instance, while most teachers report frequent use of digital technology, the depth and pedagogical effectiveness of this use are not captured by the survey. Additionally, the quantitative nature of the survey, while efficient for broad data collection, does not delve into the qualitative nuances behind teachers' responses, particularly regarding the mixed attitudes towards AI or the reasons behind preferring traditional resources.

5 Conclusion

This study provides timely insights into primary school teachers' perceptions and experiences with ICT and AI integration in Croatia. Our findings indicate a predominantly digitally confident and active teaching workforce that recognizes the positive impact of technology on their professional lives and teaching effectiveness. Key challenges remain, particularly the widespread issue of unreliable student devices and significant time constraints for digital material development. Teachers also exhibit mixed attitudes toward AI, balancing optimism with concerns about job security.

To foster more effective and equitable digital integration, our results underscore the need for targeted institutional support. This includes direct interventions to ensure student device functionality, enhanced access to efficient digital content creation tools and specialized professional development addressing AI literacy, ethical considerations and practical applications. Such initiatives should emphasize pedagogical approaches that thoughtfully blend digital and traditional methods.

While offering a valuable snapshot of these dynamics, the study's reliance on a convenience sample necessitates caution in generalizing findings. Future research should employ more representative sampling strategies and consider mixed-methods approaches to delve deeper into the nuanced factors influencing teachers' technology adoption. Ultimately, this research highlights a critical dual reality in Croatian primary education: a capable teaching force ready for digital transformation yet hindered by persistent infrastructural and systemic barriers. Addressing these will be crucial for sustained digital progress.

References

Bautista-Vallejo, J. M., Hernández-Carrera, R. M., Moreno-Rodriguez, R., & Lopez-Bastias, J. L. (2020). Improvement of memory and motivation in language learning in primary education through the interactive digital whiteboard (idw): The future in a post-pandemic period. *Sustainability*, *12*(19), 8109.

Cojean, S., Brun, L., Amadieu, F., & Dessus, P. (2023). Teachers' attitudes towards AI: what is the difference with non-AI technologies?. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Garcia, O. A., García, M. D. C. R., & Alonso-Secades, V. (2025). Perceptions, Strategies, and Challenges of Teachers in the Integration of Artificial Intelligence in Primary Education: A Systematic Review. Journal of Information Technology Education: Research, 24, 006.

Ghomi, M., Redecker, C. (2019). Digital Competence of Educators (DigCompEdu): Development and Evaluation of a Self-assessment Instrument for Teachers' Digital Competence. In *CSEDU (1)* (pp. 541-548).

Maeneja, R., Rato, J., & Ferreira, I. S. (2025). How Is the Digital Age Shaping Young Minds? A Rapid Systematic Review of Executive Functions in Children and Adolescents with Exposure to ICT. *Children*, *12*(5), 555.

Pezo, E. (2024). Adoption of ICT by Basic Education Teachers: A systematic review. Journal of Namibian Studies : History Politics Culture, 40, 314–331. https://doi.org/10.59670/3tn2an17

Røkenes, F. M., Grüters, R., Skaalvik, C., Lie, T. G., Østerlie, O., Järnerot, A., Humphrey, K., Gjøvik, Ø., & Letnes, M.-A. (2022). Teacher educators' professional digital competence in primary and lower secondary school teacher education. Nordic Journal of Digital Literacy, 17(1), 46–60. https://doi.org/10.18261/njdl.17.1.4

Soriano-Sánchez, J. G. (2025). The Impact of ICT on Primary School Students' Natural Science Learning in Support of Diversity: A Meta-Analysis. *Education Sciences*, *15*(6), 690.

Šabić, J., Baranović, B., & Rogošić, S. (2022). Teachers' self-efficacy for using information and communication technology: The interaction effect of gender and age. *Informatics in education*, *21*(2), 353-373.

Vidal-Esteve, M. I., & Martín-Gómez, S. (2023). Digitalization of Classrooms: A Comparative Study on Teachers' Perceptions about the Use of Digital Teaching Materials in Early Childhood and Primary Education. *Education Sciences*, *13*(11), 1156.

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. International Journal of Educational Technology in Higher Education, 21(1), 15.

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. Expert Systems with Applications, 252, 124167. https://doi.org/10.1016/j.eswa.2024.124167

Yu, J., Sukumaran, S., Li, H., & Zeng, J. (2023). Examine Primary School Teachers' Attitudes Towards Technology Integration in Primary Education. Environment-Behaviour Proceedings Journal, 8(26), 383-388.

Can We Trust AI in the Classroom? Examining Ethical and Privacy Challenges of LLMs – A Literature Review

Ivana Ružić, Igor Balaban
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42 000 Varaždin, Croatia
{ivruzic, igor.balaban}@foi.unizg.hr

***Abstract**. This review paper examines the ethical and data privacy challenges associated with using Large Language Models (LLMs) in education. While LLMs like ChatGPT offer personalized support and expanded access to learning, their implementation raises concerns about bias, academic integrity, and student data protection. Key risks include over-reliance on AI, insufficient transparency, and unclear accountability. The paper highlights the lack of research on teacher perspectives and use in education. It concludes by calling for ethical frameworks, robust privacy policies, and inclusive design to ensure that LLMs enhance educational equity and uphold trust in digital learning environments.*

Keywords. Large Language Models, education, data privacy, ethical consideration

1 Introduction

Large Language Models (LLMs) have rapidly emerged as transformative tools in education, offering new opportunities to enhance learning and teaching processes. Their ability to generate coherent, human-like responses makes them valuable for supporting students across a range of tasks, from studying and brainstorming to complex problem-solving, including coding (Due et al., 2024). Tools such as ChatGPT are increasingly being integrated into classrooms and educational platforms, promoting greater engagement and personalized learning experiences. These capabilities suggest a growing potential to democratize education and reduce disparities by providing accessible, real-time support to learners.

A significant advantage of LLMs is their ability to deliver immediate, context-sensitive assistance. Students encountering challenges with homework or complex concepts can receive clear and simplified explanations, much like having access to a personal tutor (Due et al., 2024). Moreover, LLMs accommodate various learning preferences, some students benefit from reading detailed content, while others prefer interactive, dialog-based support. This versatility makes LLMs particularly useful in under-resourced settings where individual teacher support may be limited, ensuring that learners still receive timely and customized help.

Despite their promise, the integration of LLMs into educational contexts raises serious ethical and data privacy concerns. A central ethical issue involves algorithmic bias and lack of transparency. These models may unintentionally sustain social stereotypes or produce unfair outcomes, especially if not properly monitored and evaluated (Fenu et al., 2022; Wambsganss et al., 2023). Furthermore, the use of LLMs involves processing enormous amounts of data, which raises critical questions about student privacy and data protection (Abbo et al. (2025); Zhang et al. (2022); Andries & Robertson (2023)). Ensuring anonymity, implementing secure data storage, and maintaining transparency about how data is collected and used are essential steps toward ethical deployment (Fenu et al., (2022); McDonald & Pan (2020)).

Wambsganss et al. (2023) note that while LLMs are increasingly used as writing support tools in educational environments, this trend requires ethical oversight to avoid unintended data exposure or improper use. Their study confirms that even when LLMs do not transfer gender bias to students' outputs, robust privacy measures are still necessary. As student writing and behaviour are increasingly subject to algorithmic analysis, educational stakeholders must establish clear ethical guidelines and implement fairness-oriented evaluation frameworks to maintain trust and accountability (Fenu et al., 2022).

The integration of LLMs into education offers new opportunities for personalized learning and data analysis, but it also raises critical ethical and privacy concerns. Compliance with the General Data Protection Regulation (GDPR) is essential, as it mandates strict safeguards for handling personal data within the EU. Educational use of LLMs must ensure data minimization, anonymization, and informed consent (März et al., 2024; Mamalis et al., 2024). GDPR also limits automated decision-making, requiring human oversight in any impactful decisions regarding students (März et al., 2024). Privacy risks, such as unintentional memorization of sensitive data, demand technical solutions like differential privacy and data anonymization (Xiao et al., 2023; Miranda et

al., 2024). Ethical concerns include potential algorithmic bias and misinformation, highlighting the need for transparency and fairness (Dungca, 2023). Techniques such as Privacy Protection Language Models (PPLM) and instruction-based tuning help balance effectiveness and privacy (Xiao et al., 2023). To responsibly utilize LLMs in education, institutions must prioritize privacy, legality, and ethical oversight while maximizing their educational benefits.

2 General Objective and Research Questions

The aim of this paper is to explore and understand the primary ethical concerns associated with the use of LLMs in education. The study focuses on identifying key ethical issues resulting from the integration of these technologies into educational settings, as well as analysing the perceived risks and challenges related to data privacy for students. Additionally, this literature review attempts to provide insights that can guide the responsible and secure implementation of LLMs in education.

The following research questions are at the centre of interest:

RQ1: What are the primary ethical concerns associated with the use of Large Language Models in education?

RQ2: What are the perceived risks and challenges related to data privacy when using Large Language Models in educational settings?

3 Material and Method

The literature review process was divided into four stages: identification, screening, eligibility, and inclusion, according to Boland et al. (2017) and the complete process is illustrated on Fig. 1.

In the first identification stage, a structured search strategy was created for use on the scientific databases Scopus and Web of Science. A search string: *Large Language Model** AND *ethic** AND *data privacy* AND *education*, was used for database search.

For the second stage, screening, 206 results from Web of Science and 107 results from Scopus were identified. The following additional selection criteria were used:

1. Published in English,
2. Published within the time frame 2019-2025,
3. Document type were article and conference paper,
4. Full text was available,
5. The subject areas were Computer and Social Science (Scopus) and Educational Research in Education (Web of Science).

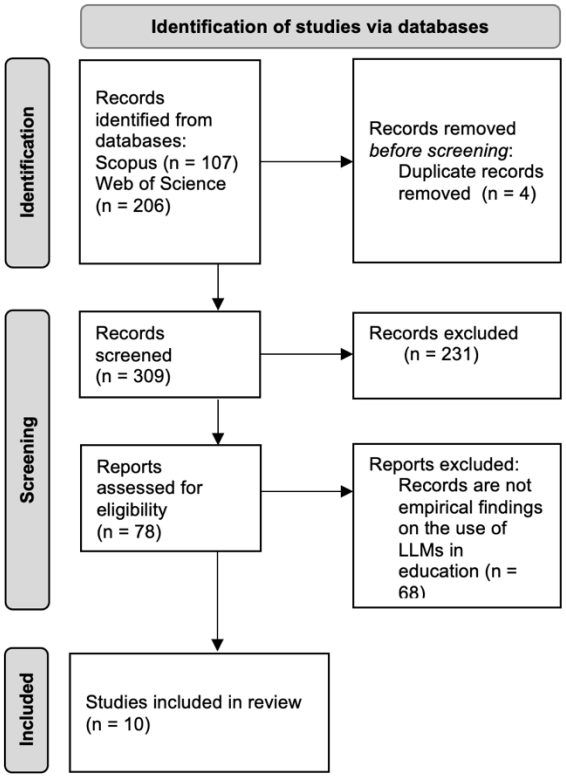


Figure 1. PRISMA flow diagram of the literature review process

As a result, 31 studies were obtained from Scopus and 51 from Web of Science. 4 duplicates were detected and excluded, and 78 studies were reviewed by titles and abstracts.

In the next stage, the final eligibility criterion was applied – studies should have included empirical findings on the use of LLMs in education. From the focus are excluded all studies that examine LLMs in education but are not related to ethical and data privacy issues, as well as studies that explore LLMs without presenting empirical research findings, offering instead only guidelines or conceptual frameworks for their use in education. A total of 68 studies were removed because they did not meet the required criteria. A total of 10 studies were included in the fourth stage of the literature review. The three studies are available in Scopus under the preprints category, reflecting their relevance and emerging importance within the rapidly evolving field of AI in education.

4 Results

The results of the literature review are presented in Table 1, showing the type of AI tool used, the number and role of participants, the educational level, and the duration of the educational activity.

Table 1. Presentation of the research results

Authors	Participants			Duration	Tool
	Role	Level	No		
Abbo et al. (2025)	students	Higher education	21	-	LLM, Robot Misty II
Zhang et al. (2022)	students	Middle education	25	30 hours	Teachable Machine, GANs
Shalevska & Kostadinovska-Stojchevska (2024)	students	Higher education	114	-	ChatGPT
Wambsganss et al. (2023)	students	Middle education	231	min. 15 min	ChatGPT
Sublime & Renna (2024)	students	Middle and Higher education	395	-	ChatGPT
Prather et al. (2023)	students	Higher education	171	-	ChatGPT
Andries & Robertson (2023)	students	Primary education	194	-	Alexa
Vrågård et al. (2024)	educators	Higher education	-	-	ChatGPT
Knowles (2021)	educators	Higher education	35	-	LLM
McDonald & Pan (2020)	students	Higher education	20	-	LLM

The results show that most studies focused on students as the primary participants, reflecting a strong interest in understanding learners' experiences and outcomes when interacting with AI tools. Only two studies (Vrågård et al., 2024; Knowles, 2021) included educators, indicating that research on teachers' perspectives and experiences remains relatively underrepresented despite its importance for effective integration of AI in educational contexts.

The majority of studies were conducted in the context of higher education, followed by middle (secondary) education. Only one study (Andries & Robertson, 2023) targeted primary education, while one study (Sublime & Renna, 2024) included participants from both secondary and tertiary levels. This distribution suggests that AI interventions are primarily explored within more advanced educational settings, potentially due to students' higher digital literacy and autonomy.

The number of participants varied significantly, ranging from small-scale studies with fewer than 30 participants (e.g., Abbo et al., 2025; Zhang et al., 2022; McDonald & Pan, 2020) to large-scale studies involving over 200 individuals (e.g., Sublime & Renna, 2024; Wambsganss et al., 2023; Andries & Robertson, 2023).

Most studies did not report the duration of the intervention, which limits our understanding of the temporal dynamics of AI tool use. Notable exceptions include Zhang et al. (2022), who conducted a 30-hour intervention, and Wambsganss et al. (2023), who reported a minimum usage duration of 15 minutes.

The most commonly used AI tool across studies was ChatGPT, featured in five studies, particularly within higher and secondary education. LLMs more broadly were employed in three studies, while other AI technologies such as Teachable Machine, GANs, Alexa, and the Misty II robot were used more sporadically. This indicates a dominant reliance on text-based conversational agents, with less frequent use of embodied or multimodal AI systems.

Some studies explored more innovative or less conventional tools such as Generative Adversarial Networks (GANs) and Teachable Machine (Zhang et al., 2022) suggesting a growing interest in creative applications of AI, particularly in STEM education. The inclusion of a physical robot (Misty II) in Abbo et al. (2025) represents a promising, though still rare, integration of social robotics into educational contexts.

4.1 Ethical Concerns Associated with the Use of Large Language Models in Education

The integration of LLMs into educational settings has caused significant ethical debate. Recent studies highlight a range of concerns that span data privacy, academic integrity, cognitive development, and equity, underscoring the need for critical oversight and responsible deployment of these technologies.

A dominant ethical concern across the literature is the handling of sensitive data. Abbo et al. (2025), Zhang et al. (2022), and Andries & Robertson (2023) emphasize that the processing of personal information, particularly in primary education, requires strict protections to avoid abuse and data breaches. Wambsganss et al. (2023) further advocate for anonymization practices and transparent data governance. McDonald & Pan (2020) also highlights risks to student privacy and autonomy, highlighting concerns about the extensive data collection by AI systems and possible lack of consent for its use.

The potential of LLMs to spread and increase biases inherent in their training data is well-documented (Zhang et al., 2022; Wambsganss et al., 2023). These biases may manipulated learning materials, reinforce stereotypes, and negatively affect learners' perceptions, especially in formative years. Abbo et al. (2025) and Prather et al. (2023) caution that this can erode fairness and inclusivity in educational outcomes. McDonald & Pan (2020) further differentiate fairness from equality, noting that AI must account for students' diverse backgrounds to avoid discriminatory impacts. Similarly, Knowles (2021) warns that AI systems built on historical data may

reinforce systemic inequities, disadvantaging marginalized groups.

LLMs raise concerns about academic dishonesty. Shalevska & Kostadinovska-Stojchevska (2024) report widespread student use of AI for assessments, potentially undermining educational standards. Prather et al. (2023) and Sublime & Renna (2024) highlight these concerns, noting risks of plagiarism and cheating, which complicate authorship attribution and damage the credibility of academic achievements. Vrågård et al. (2024) similarly emphasize the critical risk that over-reliance on AI-generated content may compromise authenticity, especially in younger students whose understanding must be genuinely reflected in their work.

Several studies (Sublime & Renna, 2024; Andries & Robertson, 2023; Prather et al., 2023) warn against over-reliance on AI-generated responses. This dependency may delay the development of critical thinking, problem-solving, and expressive writing skills. Moreover, Andries & Robertson (2023) raise the issue of anthropomorphism, where young users attribute human-like traits to LLMs, possibly distorting their understanding of AI capabilities and affecting emotional development. Vrågård et al. (2024) further highlight concerns that reliance on AI may reduce creativity and original thinking, potentially creating a dependency that limits natural development in young learners. Knowles (2021) adds that overdue postponement to AI decisions may reduce opportunities for students to engage in moral and reflective reasoning, impacting their capacity for critical ethical judgment.

LLMs can occasionally generate inaccurate or misleading content. Abbo et al. (2025) and Prather et al. (2023) highlight that without critical evaluation, students may accept false information as valid, posing long-term risks to knowledge construction and trust in educational resources. Vrågård et al. (2024) point out that this is especially problematic in primary education, where learners' critical evaluation skills are still developing.

Wambsganss et al. (2023) and Andries & Robertson (2023) underscore the lack of transparency in model training and operation. Clear documentation of data sources and decision-making processes is essential to ensure alignment with pedagogical standards. Accountability mechanisms must be established to oversee ethical deployment and usage within schools. Knowles (2021) and McDonald & Pan (2020) emphasize that unclear responsibility for AI-driven decisions poses risks, as educators and students might be misled to trust AI recommendations without critical oversight, undermining trust and redress possibilities.

Zhang et al. (2022) and Prather et al. (2023) express concern that unequal access to high-quality AI tools may compound existing educational inequalities, disadvantaging students with fewer technological resources. Knowles (2021) and McDonald & Pan (2020) further highlight how AI can sustain systemic inequities if ethical frameworks and equitable policies are not firmly in place.

McDonald & Pan (2020) argue for the integration of ethics training and empathy within AI development and educational curricula to foster systems mindful of users' diverse social contexts and needs. Knowles (2021) stresses the importance of preserving the human aspects of education: empathy, moral reasoning, and judgment, that may be reduced if AI is deployed without strong ethical oversight.

4.2 Perceived Data Privacy Risks and Challenges in Educational Settings

The integration of LLMs into educational settings introduces numerous perceived risks and challenges concerning data privacy, which span technical, ethical, and institutional dimensions.

One of the primary concerns is the potential for unintentional disclosure of sensitive personal data. As Abbo et al. (2025) highlight, LLMs often operate by generating responses based on user-provided input, which can include identifiable or sensitive student information such as learning difficulties, performance records, or behavioural notes. Since these models lack an understanding of contextual confidentiality, they may unintentionally reproduce or process private data inappropriately, particularly when embedded in educational environments that involve dynamic, real-time, or multimodal interactions (Abbo et al., 2025).

The lack of transparency of LLM data processing further complicates privacy assurance. Prather et al. (2023) emphasize the lack of transparency regarding how LLMs collect, store, and utilize user data, leading to uncertainty among users about what information is retained and who has access to it. This concern is increased by the centralized architecture of many LLM platforms, where data is stored on external servers, increasing the risk of large-scale data breaches and unauthorized third-party access. Similar concerns are highlighted by Andries and Robertson (2023), who note that users, particularly children, often lack awareness of how their data is handled, potentially leading to privacy violations, profiling, or the repurposing of educational data for commercial analytics without informed consent.

The legal and regulatory challenges are also prominent. Prather et al. (2023) point out that the use of LLMs hosted on third-party servers may not comply with regulations such as GDPR in the EU or FERPA in the United States, thereby putting institutions at risk of legal breaches. The lack of institutional control over how data is shared or retained on commercial platforms introduces governance concerns, particularly when educators and administrators cannot easily ascertain the limits of data access or deletion protocols (McDonald & Pan, 2020).

Accountability and responsibility for data protection remain unclear. McDonald and Pan (2020)

argue that the integration of LLMs into educational systems blurs the lines of responsibility between technology providers, institutions, and users. This obscurity increases the likelihood of privacy mismanagement, especially in cases where data anonymization is inadequate or when systems permit inference of additional private details from seemingly safe inputs.

While some studies acknowledge privacy concerns implicitly, they do not explore them in depth. For instance, Wambsganss et al. (2023) primarily focus on gender bias in LLM outputs and only briefly address privacy by noting that student data was anonymized. This suggests an awareness of ethical data practices, but the study does not provide a substantive discussion of privacy risks such as data misuse or long-term retention vulnerabilities.

Mitigation strategies are needed but currently underdeveloped. Abbo et al. (2025) highlights the importance of implementing strict technical and regulatory frameworks to monitor and control how LLMs access, store, and disseminate data. Additionally, they promote for user education on safe AI usage and stronger encryption and consent mechanisms. Similarly, Prather et al. (2023) and Andries and Robertson (2023) recommend that robust data governance policies and clearer disclosures about data practices are essential to restoring trust and ensuring safe educational use of LLMs.

Although LLMs offer significant potential for enhancing learning experiences, their deployment in educational contexts must be accompanied by rigorous safeguards to address the multifaceted risks to data privacy. These include the risks of accidental data exposure, insufficient user consent, non-transparent data handling practices, unclear accountability, and the potential misuse of personal information. As the use of LLMs continues to expand, establishing transparent, enforceable, and ethically grounded privacy policies will be critical to protecting the rights and safety of all educational stakeholders.

5 Discussion

The integration of LLMs into education represents a significant technological and pedagogical challenge, offering new ways to enhance learning and teaching processes. The reviewed studies confirm that LLMs, particularly tools such as ChatGPT, are being increasingly adopted in education, primarily due to their capacity to provide immediate, context-sensitive support and personalized learning experiences (Due et al., 2024). However, this enthusiasm is limited by a range of ethical concerns and data privacy challenges that must be addressed to ensure responsible use.

One of the most prominent findings across the literature is obvious ethical conflict between the advantages of LLMs and the dangers they bring. On one hand, their ability to support diverse learning preferences, increase engagement, and provide timely academic assistance, especially in under-resourced educational settings, is widely acknowledged (Due et al., 2024). On the other hand, ethical issues related to algorithmic bias, academic integrity, and the developmental risks of over-reliance on AI are consistently emphasized. Several studies (e.g., Abbo et al., 2025; Prather et al., 2023; Wambsganss et al., 2023) reveal that LLMs may unintentionally maintain social stereotypes, distort learners' perceptions, or slow down the development of critical cognitive and ethical reasoning skills. This duality reinforces the urgent need for ethical oversight frameworks and inclusion of educators in the co-design and governance of AI tools, an area that remains underexplored, as only a minority of studies included teacher perspectives (Vrågård et al., 2024; Knowles, 2021).

In addition to ethical concerns, data privacy risks represent a substantial challenge. Many of the reviewed studies highlight the potential for LLMs to unintentionally process or reveal sensitive student information (Abbo et al., 2025; McDonald & Pan, 2020). This is particularly concerning in educational contexts, where children and adolescents may not fully comprehend the implications of data sharing. Furthermore, the lack of transparency regarding data storage and usage practices, especially on commercial platforms, raises serious questions about compliance with legal and institutional regulations, including data ownership, user consent, and long-term retention. Several authors (Prather et al., 2023; Andries & Robertson, 2023) stress that these issues are added by the centralized nature of most LLM platforms, which limits institutional control and complicates efforts to ensure accountability.

Another significant theme in the literature concerns fairness in access and outcomes. Studies by Knowles (2021) and Zhang et al. (2022) raise important concerns about the potential for AI to reinforce existing educational inequalities. Students with limited access to high-quality digital infrastructure or AI-illiterate educators may be disadvantaged, which runs counter to the democratic promise of LLMs to support inclusive and accessible education. Additionally, the risk of AI tools increasing historical biases, if not carefully monitored and corrected, may further marginalize underrepresented student groups.

Across studies, there is a consensus that the implementation of LLMs in education must be accompanied by clear ethical guidelines, robust data protection policies, and transparency mechanisms. However, existing mitigation strategies remain underdeveloped or inconsistently applied. While some studies propose anonymization techniques, secure data storage, and user education (Abbo et al., 2025; Andries & Robertson, 2023), few offer concrete frameworks for institutional accountability or cross-stakeholder collaboration. The limited duration and scope of many interventions, often with small sample sizes and a focus on higher education, also limit the generalizability of

findings, particularly for younger learners or those in diverse educational contexts.

While LLMs have demonstrated considerable potential to enrich educational practice, their integration must be critically examined through both ethical and technical focus. Future research should prioritize longitudinal studies, include diverse learner and educator perspectives, and explore strategies for safeguarding data privacy and promoting fairness. Only by addressing these challenges comprehensively can we ensure that LLMs serve as responsible partners in the educational process, rather than sources of new forms of unfairness or negative impact.

6 Conclusion

The integration of LLMs in educational settings presents both significant opportunities and risks. On the one hand, LLMs have shown great potential to support personalized learning, foster student engagement, and bridge gaps in educational access. On the other hand, their deployment raises critical ethical concerns related to data privacy, algorithmic bias, and the risk of reducing human agency in the learning process.

This paper highlights the need for a balanced approach, one that embraces the pedagogical benefits of LLMs while rigorously addressing their ethical and technical challenges. Despite the growing interest in LLM applications, the current research remains fragmented, with limited focus on younger learners, teacher involvement, and long-term effects. Future research should adopt a multidisciplinary approach, actively involving educators, learners, policymakers, and developers in the co-design of AI-supported educational tools. In parallel, educational institutions must establish clear ethical guidelines, ensure transparency in data handling, and prioritize equity in access and outcomes.

Only through responsible, inclusive, and ethically informed implementation can LLMs fulfil their promise as transformative tools for education, empowering rather than replacing the human elements of teaching and learning.

References

Abbo, G. A., Desideri, G., Belpaeme, T., & Spitale, M. (2025). " Can you be my mum?": Manipulating Social Robots in the Large Language Models Era. arXiv preprint arXiv:2501.04633.

Andries, V., & Robertson, J. (2023). Alexa doesn't have that many feelings: Children's understanding of AI through interactions with smart speakers in their homes. Computers and Education: Artificial Intelligence, 5, 100176.

Boland, A., Cherry, G., & Dickson, R. (2017). Doing a Systematic Review: A Student's Guide. Sage.

Due, S., Das, S., Andersen, M., López, B. P., Nexø, S. A., & Clemmensen, L. (2024). Evaluation of Large Language Models: STEM education and Gender Stereotypes. arXiv preprint arXiv:2406.10133.

Dungca, P. A. P. (2023). The incorporation of large language models (llms) in the field of education: Ethical possibilities, threats, and opportunities. In Philosophy of Artificial Intelligence and Its Place in Society (pp. 78-97). IGI Global.

Fenu, G., Galici, R., & Marras, M. (2022, July). Experts' view on challenges and needs for fairness in artificial intelligence for education. In International Conference on Artificial Intelligence in Education (pp. 243-255). Cham: Springer International Publishing.

Knowles, M. A. (2021). Five motivating concerns for AI ethics instruction. Proceedings of the Association for Information Science and Technology, 58(1), 472-476.

Mamalis, M., Kalampokis, E., Fitsilis, F., Theodorakopoulosand, G., & Tarabanis, K. (2024). A Large Language Model based legal assistant for governance applications.

März, M., Himmelbauer, M., Boldt, K., & Oksche, A. (2024). Legal aspects of generative artificial intelligence and large language models in examinations and theses. GMS Journal for Medical Education, 41(4), Doc47.

McDonald, N., & Pan, S. (2020). Intersectional AI: A study of how information science students think about ethics and their impact. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2), 1-19.

Miranda, M., Ruzzetti, E. S., Santilli, A., Zanzotto, F. M., Bratières, S., & Rodolà, E. (2024). Preserving privacy in large language models: A survey on current threats and solutions. arXiv preprint arXiv:2408.05212.

Prather, J., Denny, P., Leinonen, J., Becker, B. A., Albluwi, I., Craig, M., ... & Savelka, J. (2023). The robots are here: Navigating the generative ai revolution in computing education. In Proceedings of the 2023 working group reports on innovation and technology in computer science education (pp. 108-159).

Shalevska, E., & Kostadinovska-Stojchevska, B. (2024). Ethics in times of advanced ai: investigating students' attitudes towards chatgpt and academic integrity. International journal of Education Teacher, 27, 72-78.

Sublime, J., & Renna, I. (2024). Is ChatGPT Massively Used by Students Nowadays? A Survey on the Use of Large Language Models such as ChatGPT in

Educational Settings. arXiv preprint arXiv:2412.17486.

Vrågård, J., Brorsson, F., & Aghaee, N. (2024, October). Generative AI in Higher Education: Educators' Perspectives on Academic Learning and Integrity. In Proceedings of The 23rd European Conference on e-Learning. Academic Conferences International.

Wambsganss, T., Su, X., Swamy, V., Neshaei, S. P., Rietsche, R., & Käser, T. (2023). Unraveling downstream gender bias from large language models: A study on AI educational writing assistance. arXiv preprint arXiv:2311.03311.

Xiao, Z., Li, T. W., Karahalios, K., & Sundaram, H. (2023, April). Inform the uninformed: improving online informed consent reading with an AI-powered Chatbot. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-17).

Zhang, H., Lee, I., Ali, S., Dipaola, D., Cheng, Y., & Breazeal, C. (2022). Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study. International Journal of Artificial Intelligence in Education. https://doi.org/10.1007/s40593-022-00293-3

The Dual-Edged Sword: Exploring Student Perceptions and Emotional Responses to AI in Academic Writing

Goran Hajdin, Igor Balaban, Marko Matus
University of Zagreb Faculty of Organization and Informatics
Laboratory for Advanced Technologies in Education
Pavlinska 2, 42000 Varaždin, Croatia
{goran.hajdin, igor.balaban, marko.matus}@foi.unizg.hr

Abstract. *This study investigates Croatian university graduate students' experiences, perceptions, and responses regarding the use of artificial intelligence (AI) tools in academic writing across three courses. Utilizing a qualitative text analysis approach, the research aimed to answer questions regarding students' reasons for AI tool selection, biggest gains, and challenges. The research examined 64 individual student perspectives derived from essays. Initial findings reveal that AI tools are primarily valued for their capacity to enhance efficiency, accelerate content generation, and stimulate idea formation, reducing the initial burden of academic tasks. However, students consistently articulated concerns regarding AI's accuracy, the reliability of its generated sources, and the generic nature of its output. The study underscores the need for explicit AI literacy education within higher education curricula and the development of pedagogical strategies that effectively integrate AI while simultaneously fostering human-centric skills such as critical analysis, synthesis, and authentic academic voice.*

Keywords. Artificial Intelligence, ICT Integration, Academic Writing, AI Literacy, Students' perceptions, Higher Education

1 Introduction

The rapid proliferation of artificial intelligence (AI) tools, particularly generative AI systems such as ChatGPT, Copilot, Perplexity AI, and NotebookLM, has significantly reshaped traditional academic writing practices in higher education. These tools offer diverse functionalities including idea generation, drafting, editing, and organizing written work, presenting students with powerful means to enhance their productivity and efficiency (Ozfidan et al., 2024). At the same time, their widespread adoption has ignited debates about academic integrity, originality, and the potential erosion of critical thinking and authentic student voice (Rabbianty et al., 2023); (Gallagher & Wagner, 2024).

Existing literature reflects a growing interest in how students perceive and emotionally respond to these technologies. Studies across different cultural contexts reveal a mixture of optimism and apprehension: while students appreciate the support AI provides for grammar correction, idea structuring, and time efficiency, they also express concerns about ethical misuse, dependency, and the impact on their learning and identity as writers (Hysaj et al., 2024); (Rashid et al., 2024); (Subaveerapandiyan et al., 2025). Emotional reactions also vary widely, with some students experiencing motivation and confidence from AI-generated feedback, while others report anxiety or confusion related to authenticity and academic expectations (Alsaiari et al., 2024); (Guo & Wang, 2024).

Despite these developments, there remains a notable gap in research focusing on student experiences with AI. Most current findings derive from large-scale studies in Anglophone or Middle Eastern educational systems (Bista & Bista, 2025); (Bui & Tong, 2025). There is a need for culturally and pedagogically contextualized studies that provide direct student insights into how they use AI in coursework, the perceived benefits and drawbacks, and how their emotional and ethical outlooks evolve through engagement with these tools.

This paper addresses this research gap by focusing on the Croatian higher education context. Specifically, it investigates practice in three courses at a single Croatian university where it focuses on students' experiences, perceptions, and emotional responses regarding the use of AI tools in academic writing across three distinct courses. By grounding the study in students' lived experiences and linking these to broader educational trends, this research aims to contribute insights to current discussions on AI in academia, offering guidance for pedagogical strategies, curriculum design, and institutional policy.

2 Literature Review

The pervasive influence of AI in higher education has substantially impacted academic writing practices,

particularly among students. The emergence of generative tools like ChatGPT, Grammarly, and QuillBot has led to an increasing reliance on AI technologies for assistance with brainstorming, grammar correction, paraphrasing, and the structural organization of written work. This shift has spurred a growing body of academic literature dedicated to understanding not only the mechanisms of AI use by students but also their sentiments towards these tools and their perceptions of the educational and ethical ramifications.

A prominent theme in the literature is the evolving conceptualization of AI as a collaborative partner in writing rather than merely a functional utility. Lin (2023) proposed a two-stage human-AI collaborative framework, positing AI as a co-authoring entity capable of assisting across all stages of academic writing, from initial planning to final revision. This work highlighted cognitive benefits such as imaginative stimulation, enhanced fluency, and reduced workload, suggesting that judiciously used AI tools can augment scholarly communication and potentially democratize access to academic publishing. This perspective finds resonance in applied research, such as Douglas (2024), who implemented a structured AI-integrated writing model across five literacy stages (questioning, locating, evaluating, synthesizing, and communicating). Graduate students in this study reported notable improvements in writing confidence, digital research skills, and academic self-efficacy following structured exposure to tools like ScholarAI and Humata.

The affective and cognitive dimensions of AI use are also a significant focus, with studies consistently reporting a spectrum of emotional and psychological outcomes from AI-assisted writing. Kim et al. (2024) conducted in-depth interviews with twenty Chinese university students using a ChatGPT-embedded writing platform, observing that students perceived AI not just as a technical assistant but as a "virtual peer" or "tutor." The tool was particularly valued for alleviating writer's block and providing language support, though students also voiced reservations about the impersonal and sometimes generic nature of the feedback. Similarly, a quantitative study by Dingal et al. (2024) in the Philippines, involving 335 high school students, demonstrated a statistically significant correlation between AI usage and improved academic writing performance. Variables such as AI relevance, user confidence, and AI literacy were identified as particularly predictive of enhanced outcomes, suggesting that technical familiarity with AI amplifies its educational efficacy.

Despite these cognitive and practical advantages, students' comprehension of AI's ethical boundaries appears to be limited. Gabriel (2024), in a study of teacher college students in Austria, noted that while participants generally found AI useful for idea generation and drafting, they lacked a nuanced understanding of its ethical implications. Many were unaware of academic policies governing AI use, which posed challenges in distinguishing between acceptable assistance and academic misconduct. This concern was further corroborated by Lund et al. (2024), whose survey of students' perceptions of academic dishonesty in the context of AI-generated content revealed considerable confusion regarding permissible versus unethical AI use, especially when outputs from tools like Grammarly closely resembled those of more advanced generative models. This ambiguity is exacerbated by broader discussions surrounding authorship and originality. Rabbianty et al. (2023) found that while faculty in Indonesian institutions appreciated the time-saving benefits of AI tools, they were concerned about preserving the authenticity of student work. Students similarly expressed apprehension that AI-generated outputs might dilute their unique academic voice. John et al. (2024) cautioned that overreliance on AI could stifle originality and critical thought, advocating for revised assessment practices, including oral defenses and reflective tasks, to holistically evaluate student understanding.

Another crucial aspect highlighted in the literature is the role of AI literacy. Hossain et al. (2025), in their study of Turkish EFL learners, observed that while students frequently employed AI tools for grammar checking and translation, their understanding of the functionality and limitations of these tools was minimal. Students often treated AI as a black-box utility, underestimating the risks associated with overdependence or misuse. The imperative for structured AI literacy programs was similarly emphasized by Taiye et al. (2024), who involved students in the co-design of a chatbot (CHAT4ISP-AI) to support academic writing in undergraduate social science education. The study concluded that stakeholder-informed AI tools could be more effectively tailored to meet both pedagogical and ethical expectations.

The literature also delineates varying emotional responses to AI use. Selim (2024) found that Saudi university students in EFL contexts valued AI tools like ChatGPT and Grammarly for enhancing the clarity and coherence of their essays. However, these students expressed ambivalence regarding whether AI use had genuinely boosted their self-confidence or writing independence. Wahyuningsih (2024) echoed this sentiment in a qualitative study of Indonesian undergraduates, where students reported using various AI tools for translation, grammar correction, and plagiarism checks. While these tools were deemed helpful for structure and efficiency, students consistently noted that AI could not foster critical thinking, analytical reasoning, or deeper conceptual engagement.

Students' preferences for AI tools also reflect a pragmatic orientation towards time efficiency and task

automation. Malik et al. (2023) reported that Indonesian students widely adopted AI for essay outlining and plagiarism detection, though some worried that these conveniences might erode creativity and lead to a mechanistic approach to writing.

Bista and Bista (2025) observed similar ambivalence among doctoral students in the United States. While these students lauded the clarity and structure provided by AI, they also highlighted inaccuracies and the necessity for faculty guidance to critically interpret AI suggestions.

Ethical concerns related to inclusivity and equity also emerge in the discourse. Omodan and Marongwe (2024) posited that AI could play a decolonizing role in academic writing by assisting non-native English speakers in overcoming linguistic barriers to publication. Their conceptual framework links AI integration with broader efforts to democratize academic knowledge production, though the authors caution against uncritical adoption without considering cultural and epistemological diversity.

Finally, the role of institutional support and instructional design is emphasized in several studies. Lee-Price (2024) explored academic writing workshops designed to educate students on ethical AI use through activities like paraphrasing and co-creation with ChatGPT. These workshops, grounded in principles of academic literacies and critical pedagogy, helped students reflect on their responsibilities and choices when employing AI in academic settings. In distance learning contexts, Maphoto et al. (2024) found that generative AI increased student engagement and writing motivation. However, educators stressed the importance of maintaining a balance between technological support and independent thinking. Collectively, these studies offer a comprehensive depiction of how students perceive, utilize, and are affected by AI in academic writing. The literature reveals a complex interplay of benefits and challenges: AI can enhance efficiency, confidence, and accessibility, but its use also raises unresolved questions about ethics, creativity, and educational equity. Addressing these tensions necessitates not only policy clarity but also pedagogical innovation and robust digital literacy development.

3 Research Methodology

This study employs a qualitative, descriptive research design to explore the multifaceted experiences, perceptions, and emotional responses of university students regarding the use of AI tools in academic writing. This approach is particularly suited for in-depth understanding of complex phenomena from the participants' perspectives. The chosen approach is inductive, allowing themes and patterns to emerge directly from the rich textual data provided by student reflections. Simultaneously, the analysis is guided by specific categories outlined in the research query, ensuring comprehensive coverage of the requested aspects of AI use (Cohen et al., 2018).

Furthermore, this research was carried out as an action research study. Action research is a qualitative methodology focused on solving practical problems within a specific context, often involving practitioners as researchers. It is cyclical, involving planning, acting, observing, and reflecting, aiming for both practical change and knowledge generation (Bradbury, 2015; Coghlan & Brydon-Miller, 2014). This approach is particularly relevant in educational settings for improving teaching practices and student learning, as it directly engages participants in the inquiry process to foster immediate and relevant improvements.

The aim of this study is to explore Croatian university graduate students' experiences, perceptions, and emotional responses regarding the use of AI tools in academic writing across three courses. To achieve this aim, the following research questions guided the investigation:

- RQ1: What are the primary reasons students choose to use AI tools in academic writing?
- RQ2: What are the biggest perceived gains for students from the application of AI in academic writing?
- RQ3: What are the biggest challenges students encounter when writing with AI?

The study gathered a total of 64 individual student perspectives, derived from essays. These participants were enrolled in three different university courses: COR#1, contributing 39 individual perspectives; COR#2, providing 23 individual perspectives; and COR#3, with 2 individual perspectives from 1 essay/group.The inclusion of multiple courses and varied student counts per essay offers a diverse, albeit self-selected, sample of experiences with AI in academic writing. The data originates from student reflections submitted as part of their academic assignments, where the application of AI tools was an integral component of the academic writing process. This context is crucial as it ensures that the reflections are directly tied to practical, real-world application of AI in their coursework. Students reported using a variety of AI tools, including prominent generative AI models such as ChatGPT (e.g., versions 3.5, 4o, ScholarAI, Plus), Microsoft Copilot, Perplexity AI, NotebookLM, and the Aria AI chatbot. This diversity of tools used offers a broader understanding of the perceived impacts and functionalities across different AI platforms.

4 Data Collection and Analysis

Student reflections were collected as written responses to assignments that required them to use and comment on their experience with AI tools for academic essay writing. This method yielded rich, first-person qualitative data detailing their interactions, perceived benefits, challenges, and emotional states. The original responses were provided in Croatian. Further, each essay was presented and followed by group discussion. To provide a comprehensive understanding of student experiences, interview responses were collected from 5 students from each course group (COR#1, COR#3, COR#2). The whole data collection process spanned two semesters, encompassing three graduate university courses. Each interview addressed five specific questions: 1) Why did you choose this tool? 2) What did you learn about academic essay writing with the use of AI? 3) How did you feel after the writing? 4) What was your biggest gain from an AI application? 5) What was your biggest challenge when writing with AI? This approach allowed for a structured presentation of emergent themes and common experiences, providing a human-centered perspective on the findings (Cohen et al., 2018).

A systematic six-phase thematic analysis approach, as outlined by Braun and Clarke (2006, 2021), was employed to identify, analyze, and report patterns within the data. I) Familiarization with the data: This initial phase involved repeated reading of the translated student responses to immerse the researchers in the content and identify initial points of interest. II) Generating initial codes: Specific phrases, sentences, or paragraphs that captured key ideas related to AI use, feelings, gains, and challenges were identified and assigned codes across the entire dataset. III) Searching for themes: Related codes were then grouped into broader, overarching themes that captured significant patterns of meaning within the data. IV) Reviewing themes: This iterative phase involved two levels of assessment: first, evaluating if the themes accurately represented the coded extracts; and second, assessing if the themes captured the essence of the entire dataset. Themes were refined, merged, or discarded as necessary to ensure coherence and representativeness. V) Defining and naming themes: A detailed narrative was developed for each theme, outlining its scope, what it represented, and its significance in relation to the research questions. Clear and concise names were assigned to each theme. VI) Producing the report: The final stage involved weaving the thematic analysis into a coherent narrative for the results section, supported by illustrative quotes from the translated student responses.

5 Limitations

This study, while providing valuable insights, is subject to several limitations. First, the reliance on student self-reflections means the data may be subject to recall bias, social desirability bias, or incomplete reporting of their experiences. Students might consciously or unconsciously present their experiences in a more favorable light or omit certain details. Second, the provided data consists of "partial responses," which inherently limits the depth and breadth of individual student narratives compared to full, in-depth interview transcripts. This constrains the ability to explore complex individual motivations or emotional states in full detail. Third, the findings are derived from a specific university context in Croatia. While providing valuable insights into this particular environment, they may not be fully generalizable to all educational systems or cultural contexts, as AI adoption and academic norms can vary significantly. Finally, the field of AI is rapidly evolving. The capabilities and common uses of AI tools are constantly changing, meaning the findings of this study might become outdated as new models emerge and students adapt their practices.

6 Results & Discussion

The thematic analysis revealed three primary themes characterizing students' experiences with AI in academic writing:

6.1 Theme 1: Efficiency and Speed as Primary Motivators (Addressing RQ1 & RQ2)

Students consistently reported that the paramount reason for adopting AI tools was their ability to rapidly generate content, reduce research and writing time, and streamline the overall academic process. This efficiency was perceived as a enabler for managing workload and meeting deadlines. For example, a student highlighted ChatGPT's provision of "quick access to a large amount of information". Another noted AI's facilitation of "quick content generation" and how it "accelerated the process". One student explicitly stated that AI "extracted information quickly" and made the "research and writing process significantly faster". Similarly, a group of students praised AI for its "great speed" in providing information that would otherwise take "much more time" to find manually. The sentiment was encapsulated by one student who described ChatGPT as "super fast, useful, and incredibly amazing" for its ability to "write chapters in seconds". This consistent emphasis on "efficiency" and "speed" aligns strongly with prior research,

supporting findings by Malik et al. (2023) regarding essay outlining and Dingal et al. (2024) on improved academic writing performance. Lin (2023) also noted reduced workload as a cognitive benefit of AI. This indicates a shift in student strategy, where AI serves as a "kick-starter" or scaffolding mechanism, particularly for overcoming writer's block or structuring initial drafts. The high value placed on speed and efficiency points to the practical challenges students face in managing academic workloads. If AI reduces the "activation energy" required to begin writing, it could lead to more frequent engagement with academic tasks, even if the subsequent human effort for refinement remains substantial. This transforms AI from a mere tool into a strategic resource for managing academic workload and initiating complex assignments.

6.2 Theme 2: AI as a Source of Ideas and Structure (Addressing RQ1 & RQ2)

Beyond mere efficiency, students valued AI for its capacity to generate diverse ideas, propose structural outlines, and offer alternative perspectives on topics. This functionality was seen as highly beneficial for organizing thoughts, enriching content, and exploring different angles of a subject. One student found AI helpful for "content proposals". Another noted AI's ability to "give explanations and more different possible solutions". A student described AI as an "excellent source of ideas" and "explanations for specific problems". The utility of AI in providing "structured and quick shaping of thoughts" and helping to "connect key contemporary themes" was also appreciated. Furthermore, AI's provision of "structured ideas" and "enrichment with new perspectives" was highlighted. Several students simply listed AI as a "source of ideas" or recommended it for "basic structure and guidelines" and "quickly gathering basic information". A particularly insightful comment mentioned AI's utility in "generating ideas, different perspectives" when one is "thinking in only one way". This aligns with Kim et al. (2024) who found students valued AI for alleviating writer's block and Gabriel (2024) who noted AI's utility for idea generation and drafting support. AI's role in idea generation and structuring acts as a significant cognitive offload, particularly in the initial ideation and organizational phases of writing. This frees up students' mental resources, potentially allowing them to focus on higher-order cognitive tasks like critical analysis, synthesis, and refinement, provided they engage critically with the AI's output rather than passively accepting it. By providing initial ideas and structures, AI reduces the burden of starting from scratch. This "offload" can be beneficial if students then use their freed-up cognitive capacity for deeper engagement, such as evaluating the AI's ideas, synthesizing them with their own thoughts, and refining the overall structure. The effectiveness of AI in this context is contingent upon the student's AI literacy and critical engagement. If used passively, it risks producing generic, unoriginal work. If used actively, it can augment human creativity and analytical depth. This points to a critical pedagogical challenge: how to teach students to leverage AI for ideation without sacrificing their own intellectual contribution.

6.3 Theme 3: Concerns about Accuracy, Reliability, and Depth (Addressing RQ3)

A pervasive and significant challenge reported by students was the questionable accuracy and reliability of AI-generated information, particularly concerning citations and factual data. This often resulted in superficial, generic, or repetitive content that lacked the depth, nuance, or human-like coherence expected in academic writing. For example, one student stated that "literature was not correctly cited, and all data needed additional verification, as the data was not accurate". Another noted that the "literature was not entirely reliable" and the tool "will not every time offer existing and correct" information. A student observed that the "text... sounds like it wasn't written by a human" and lacked "examples or reasons for certain points". Concerns were raised about AI's "misunderstanding of certain topics" and its tendency to output "pre-written information... including unverified sources". Difficulties with "proper referencing and finding reliable references" were reported, with text often turning out "unnatural". One student found that "cited literature did not exist in the format ChatGPT provided". It was also found "very difficult to get expert and accurate definitions" from AI. Students noted that "content sometimes retains a generic character" and "lacked emotional tone and personal dimension". AI sometimes "deviated from the topic" and had a "lack of sources and inability to provide them". One group found the tool used "words the student wouldn't personally use" and "repeated the same point twice". Answers were often "not detailed enough" and showed an "apparent blockage in expressing opinions". AI's "tendency to list things... lacking coherent text writing ability" was criticized. Sources were sometimes "somehow strange" and "not among the top results on Google Scholar". A student reported that AI "refused to cite sources" despite insistence. Finally, one student noted that AI "often gives partially or completely inaccurate answers, and sometimes even completely incorrect ones," and "relies too much on its mechanical learning process... sometimes neglecting online sources". These pervasive issues with accuracy, referencing, and generic output directly reflect the substantial confusion around what constituted permissible versus unethical use

of AI highlighted by Lund et al. (2024) and Gabriel (2024)'s observation of students' limited nuanced awareness of ethical dimensions. This also directly ties into Rabbianty et al. (2023) and John et al. (2024)'s concerns about maintaining authorship and originality, as students themselves voiced fears about AI diluting their academic voice and the need for verification. The frequent reports of incorrect or non-existent sources underscore a critical gap in AI's current capabilities and the urgent need for students to develop robust verification skills. This aligns with Bista and Bista (2025) who highlighted inaccuracies and the need for faculty guidance to critically interpret AI suggestions.

Beyond these primary themes, students also expressed concerns regarding AI's potential to impede their critical thinking and originality, suggesting a risk of "cognitive atrophy" if used passively. This highlights the imperative for students to critically engage with and verify AI-generated content, recognizing AI as an aid rather than a definitive authority. Furthermore, students reported mixed emotional responses, balancing initial satisfaction with frustration over AI's limitations and a diminished sense of personal contribution. These aspects, while noted, fall outside the direct scope of the research questions addressed in this paper.

7 Conclusion

This study provided empirical insights into Croatian university graduate students' experiences with AI in academic writing, highlighting its dual role as an efficiency enhancer and a source of challenges regarding accuracy and depth. The findings underscore a complex interplay between the convenience offered by AI tools and the imperative for students to maintain critical engagement, verify information, and preserve their unique academic voice. Students value AI for its speed and ability to generate ideas and structure, recognizing its utility in managing academic workload and overcoming initial writing hurdles. However, they consistently express significant concerns about AI's accuracy, the reliability of its sources, and its tendency to produce generic content, which they perceive as undermining their intellectual contribution. In summary, students primarily choose AI for efficiency and idea generation (RQ1, RQ2). While AI offers gains in speed and content generation, its biggest challenges lie in accuracy, reliability, and its potential to hinder critical thinking (RQ2, RQ3). The study reiterates the need for comprehensive AI literacy education within higher education. This education should go beyond tool instruction, focusing instead on fostering critical evaluation, ethical considerations, and the integration of AI as an augmentation to, rather than a replacement for, human intellect. Pedagogical strategies should adapt to this evolving landscape, designing assessments that prioritize process, critical engagement, and authentic student voice. Future research could explore several avenues to deepen understanding of AI's evolving role in education. Longitudinal studies could track the long-term impact of AI on student learning outcomes and cognitive development, providing insights into how consistent AI use shapes intellectual growth. Comparative studies across different disciplines or educational systems would help understand contextual variations in AI adoption and its perceived effects. Finally, action research into the effectiveness of specific pedagogical interventions designed to foster responsible and critical AI use would be invaluable, providing empirical evidence for best practices in an AI-augmented academic environment.

Acknowledgements

This work was supported by the Erasmus+ CUTIE – Co-Creation at Universities Through Digital Education project (Grant No. 2023-1-HR01-KA220-HED-000137890). The authors thank the project consortium partners for their invaluable feedback during data collection.

References

Alsaiari, O., Baghaei, N., Lahza, H., Lodge, J., Boden, M., & Khosravi, H. (2024). Emotionally enriched feedback via generative AI. [Preprint].

Bista, K., & Bista, R. (2025). Leveraging AI tools in academic writing. American Journal of STEM Education. https://doi.org/10.32674/9m8dq081

Bradbury, H. (Ed.). (2015). The SAGE handbook of action research (3rd ed.). Sage.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3, 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2021). Thematic analysis: A practical guide. Sage.

Bui, T. T. U., & Tong, T. V. A. (2025). The impact of AI writing tools on academic integrity: Unveiling English-majored students' perceptions and practical solutions. AsiaCALL Online Journal. https://doi.org/10.54855/acoj.251615

Coghlan, D., & Brydon-Miller, M. (Eds.). (2014). The SAGE encyclopedia of action research. Sage.

Cohen, L., Manion, L., & Morrison, K. (2018). Research Methods in Education (8th ed.). Routledge. https://doi.org/10.4324/9781315456539

Dingal, S. M. L., Balbin, C. J. P., Buenafe, S. G. L., Buenavista, P. E. M., Cabrera, C. A., Lorica, V. M. T., Olalo, A. M., Pataganao, L. A. A., Wahing, M. R. A., Clamares, K. J. M., & Pelandas, A. M. O. (2024). Artificial intelligence (AI) usage and its influence on the students' academic writing: A quantitative–correlation investigation. International Journal of Research and Innovation in Social Science, 8(4). https://doi.org/10.47772/ijriss.2024.804216

Douglas, K. H. (2024). On the effective and ethical use of AI in academic writing. Журнал серии «Филологические науки», 4(75). https://doi.org/10.48371/phils.2024.4.75.026

Gabriel, S. (2024). Generative AI in writing workshops: A path to AI literacy. International Conference on AI Research. https://doi.org/10.34190/icair.4.1.3022

Gallagher, J. R., & Wagner, K. (2024). Comparing student and writing instructor perceptions of academic dishonesty when collaborators are artificial intelligence or human. Journal of Business and Technical Communication, 38(3), 266–288. https://doi.org/10.1177/10506519241239937

Hossain, Z., Çelik, Ö., & Hınız, G. (2025). Exploring EFL students' AI literacy in academic writing: Insights into familiarity, knowledge and ethical perceptions. Kuramsal Eğitimbilim, 18(2). https://doi.org/10.30831/akukeg.1538011

Hysaj, A., Freeman, M., & Hamam, D. (2024). Using AI tools to enhance academic writing and maintain academic integrity. In Proceedings from AI and Multicultural Learning Conference (pp. 57–66). https://doi.org/10.1007/978-3-031-61305-0_4

John, K., Hussin, S. N., Yusoff, R. M., Rocky, N. A. R. M., Nawi, M. N. M., Hussain, S. S. S., & Mutalip, S. S. M. (2024). Mini review on the ethical use of AI bots in academic writing. International Journal on e-Learning and Higher Education, 19(2). https://doi.org/10.24191/ijelhe.v19n2.1924

Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on generative AI-assisted academic writing. Education and Information Technologies, 30, 1265–1300. https://doi.org/10.1007/s10639-024-12878-7

Lee-Price, S. (2024). 'The exciting AI adventure': Reflections on the ethical use of generative AI in academic writing. Journal of Learning Development in Higher Education, 32. https://doi.org/10.47408/jldhe.vi32.1399

Lin, Z. (2023). Techniques for supercharging academic writing with generative AI. arXiv. https://doi.org/10.48550/arXiv.2310.17143

Lund, B., Lee, T. H., Mannuru, N. R., & Arutla, N. (2024). Student perceptions of academic misconduct in the age of generative AI. Proceedings of the ALISE Annual Conference. https://doi.org/10.21900/j.alise.2024.1753

Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., Darwis, A., & Marzuki. (2023). Exploring artificial intelligence in academic essay: Higher education student's perspective. International Journal of Educational Research Open, 5, 100296. https://doi.org/10.1016/j.ijedro.2023.100296

Maphoto, K., Sevnarayan, K., Mohale, N. E., Suliman, Z., Ntsopi, T. J., & Mokoena, D. (2024). Advancing students' academic excellence in distance education: Exploring the potential of generative AI integration to improve academic writing skills. Open Praxis, 16(2). https://doi.org/10.55982/openpraxis.16.2.649

Omodan, B., & Marongwe, N. (2024). The role of artificial intelligence in decolonising academic writing for inclusive knowledge production. Interdisciplinary Journal of Education Research, 6(S1), Article 06. https://doi.org/10.38140/ijer-2024.vol6.s1.06

Ozfidan, B., El-Dakhs, D., & Alsalim, L. (2024). The use of AI tools in English academic writing by Saudi undergraduates. Contemporary Educational Technology. https://doi.org/10.30935/cedtech/15013

Papakonstantinidis, S., Kwiatek, P., & Spathopoulou, F. (2024). Embrace or resist? Drivers of artificial intelligence writing software adoption in academic and non-academic contexts. Contemporary Educational Technology, 16(1). https://doi.org/10.30935/cedtech/14250

Rabbianty, E. N., Azizah, S., & Virdyna, N. K. (2023). AI in academic writing: Assessing current usage and future implications. INSANIA: Jurnal Pemikiran Alternatif Kependidikan, 28(1a), 38–51. https://doi.org/10.24090/insania.v28i1a.9278

Rashid, S., Malik, S., Abbas, F., & Khan, J. A. (2024). Pakistani students' perceptions about knowledge, use and impact of artificial intelligence (AI) on academic writing: A case study. Journal of Computers in Education. https://doi.org/10.1007/s40692-024-00338-7

Selim, A. S. M. (2024). The transformative impact of AI-powered tools on academic writing: Perspectives of EFL university students. International Journal of

English Linguistics, 14(1), 14–29. https://doi.org/10.5539/ijel.v14n1p14

Silverman, D. (2010). Doing qualitative research (3rd ed.). Sage.

Subaveerapandiyan, A., Kalbande, D., & Ahmad, N. (2025). Perceptions of effectiveness and ethical use of AI tools in academic writing: A study among PhD scholars in India. Information Development. https://doi.org/10.1177/02666669251314840

Subedi, R., & Nyamasvisva, T. E. (2024). A review of the influence of artificial intelligence in academic writing. Journal of Computer Science and Information Technology, 2(1), 19–30. https://doi.org/10.70248/jcsit.v2i1.1134

Taiye, M. A., High, C., Velander, J., Matar, K., Okmanis, R., & Milrad, M. (2024). Generative AI-enhanced academic writing: A stakeholder-centric approach for the design and development of CHAT4ISP-AI. Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing. https://doi.org/10.1145/3605098.3636055

Wahyuningsih, S. (2024). How artificial intelligence (AI) supports undergraduate students' academic writing: Evidence from Indonesia. Prominent, 7(2), 123–135. https://doi.org/10.24176/pro.v7i2.11444

Software Standards in Action: A Complaint-Driven Game Inspired by Real Project Failures

Sevgi Koyuncu Tunç
Çankaya University
Software Engineering Dept., Faculty of Engineering Yukarıyurtçu Mahallesi
Mimar Sinan Caddesi No:4 06790 Etimesgut/Ankara, Türkiye
sevgik@cankaya.edu.tr

***Abstract**. Complaint to Compliance is a classroom-based game teaching ISO/IEC/IEEE 12207 standards through real-world complaint-driven scenarios. Grounded in experiential learning, its low-tech, card-based design fosters active learning via guided discussion. A qualitative case study (n=21) analyzed student perceptions, revealing: (1) conceptual bridging of theory and practice, (2) professional identity development, and (3) enhanced engagement through gamification. The accessible design offers a scalable model for standards education. Preliminary findings suggest pedagogical value for information systems education, with further cross-cultural validation needed.*

Keywords. ISO/IEC/IEEE 12207, software engineering education, gamification, serious games, process standards, classroom activity, empirical study, curriculum innovation.

1 Introduction

International standards are essential to ensuring the quality, reliability, and maturity of software systems. Among them, ISO/IEC/IEEE 12207 serves as a comprehensive framework for software lifecycle processes, encompassing acquisition, development, operation, support, and maintenance. The standard defines a unified structure for software engineering activities, applicable across diverse domains and organizational contexts.

Despite its foundational role, ISO/IEC/IEEE 12207 is often taught through traditional lectures and passive methods, limiting students' ability to apply its principles in practice. This pedagogical gap is well-documented: Calderón and Ruiz (2016) stress the need to integrate process-oriented frameworks into curricula, while Sánchez-Gordón (2016) highlights the value of exposing students to real-world standards. Aydan et al. (2017) and Calderón et al. (2018b) further argue that abstract instruction fails to prepare students for industry roles, advocating instead for authentic, experiential learning environments. Recent work by Koyuncu Tunç et al. (2024) underscores these challenges in Türkiye, where standards education remains disconnected from practical contexts.

While gamified learning (e.g., AI-supported platforms [Zhang et al., 2023]) shows promise, low-tech, scalable methods for teaching ISO/IEC/IEEE 12207 are underexplored. This study addresses that gap by presenting Complaint to Compliance, a case study of an innovative, complaint-driven game that bridges theory and practice. Grounded in experiential learning theory, the game uses real-world project failures to contextualize standard processes, fostering active engagement through collaborative problem-solving.

Through qualitative analysis of student perceptions (*n* = 21), this research examines the game's design and pedagogical utility, offering educators a replicable model for standards education. By reframing ISO/IEC/IEEE 12207 as a tool to diagnose and mitigate project risks, the intervention aims to shift student perspectives—from viewing standards as bureaucratic constraints to valuing their practical role in software engineering.

2 Background

2.1 ISO/IEC/IEEE 12207

The 2017 revision of ISO/IEC/IEEE 12207 reorganized its process model to align with ISO/IEC/IEEE 15288, enhancing consistency across systems and software engineering standards. It defines four main process categories:

1. Agreement Processes
2. Organizational Project-Enabling Processes
3. Technical Management Processes
4. Technical Processes

This structured categorization promotes both flexibility and rigor, allowing for implementation across various domains (see Fig. 1). However, its complexity poses significant instructional challenges in academic settings. Students often struggle to internalize the standard's abstract terminology and multifaceted process structure, highlighting the need for more engaging and context-aware teaching methods.

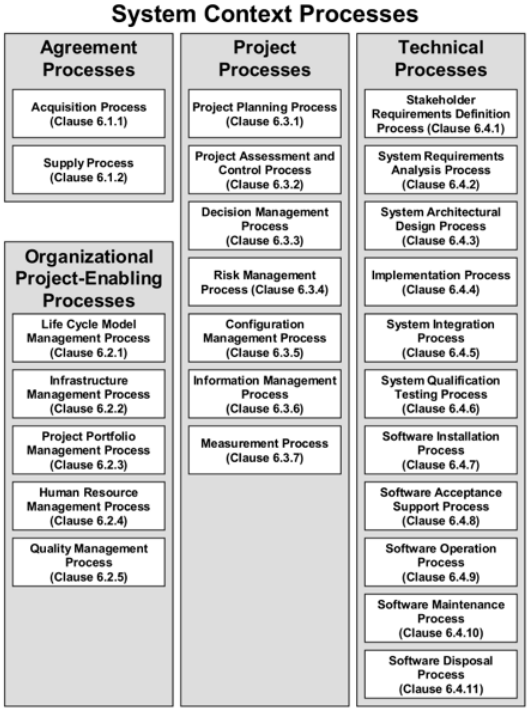


Figure 1. ISO/IEC/IEEE 12207 Processes

2.2 Serious Games in Software Engineering Education

Serious games—defined as games designed primarily for educational or training purposes—have gained traction in software engineering education. They offer immersive experiences where learners engage in simulated scenarios, interactive problem-solving, and context-driven decision-making.

Calderón et al. (2018a) conducted a multivocal literature review, concluding that serious games significantly enhance student engagement, motivation, and learning outcomes. Other studies support these findings: Garcia et al. (2020) and Calderón et al. (2017a) successfully applied game-based learning to teach ISO/IEC/IEEE 29148 and ISO/IEC 29110, respectively. Herranz et al. (2016) reported improved comprehension of software process concepts when gamification was introduced alongside traditional methods.

Several games have been developed to support standards instruction. Mesquida and Mas (2018) created MiProJOC, simulating ISO/IEC/IEEE 12207 project planning and task management. Garcia et al. (2025) introduced the Metric Cake Shop game to teach ISO/IEC/IEEE 15939, focusing on metrics and measurement. Aydan et al. (2015, 2017) proposed early-stage games aimed at making ISO/IEC/IEEE 12207 more accessible to students. Calderón et al. (2017b, 2018b) extended the concept to ISO 21500, emphasizing project management competencies.

While these efforts illustrate the potential of gamification in standards education, few games explicitly target ISO/IEC/IEEE 12207, and even fewer utilize complaint-driven, case-based analysis rooted in real industry issues. This study aims to fill that gap by designing and evaluating a game that builds students' process literacy through the analysis of authentic project complaints.

3 Research Methodology

This study aims to evaluate the pedagogical effectiveness of a complaint-driven game designed to teach the ISO/IEC/IEEE 12207 standard. Two research questions guide the investigation:

RQ1: To what extent does the complaint-matching game enhance students' understanding of ISO/IEC/IEEE 12207 lifecycle processes?

RQ2: How do students perceive the relevance, engagement, and educational value of the game-based learning approach?

These questions address both cognitive outcomes and learner perceptions, enabling a comprehensive evaluation of the instructional intervention.

3.1 Research Design

This study adopts a qualitative case study design, examining the implementation of the Complaint to Compliance game in a naturalistic classroom setting. Thematic analysis (Braun & Clarke, 2006) of student reflections and instructor observations serves as the primary method to evaluate pedagogical utility, aligning with the study's goal of exploring how gamification bridges standards theory and practice.

3.2 Participants and Intervention Description

Twenty-nine undergraduate students enrolled in the Software Process Management course at XXX University, participated in this study as an integrated component of their formal coursework. Their participation reflected authentic engagement with the course content, providing a naturalistic setting for evaluating the intervention's impact on learning outcomes and professional awareness.

The game was implemented in a Software Process Management course at [University xxx], where students (*n* = 21) had no prior formal instruction on ISO/IEC/IEEE 12207. Due to the real-world constraints of classroom settings, participation was non-randomized. While absent students (*n* = 8) provided incidental comparison data, their exclusion from the intervention reflects typical attendance dynamics rather than experimental design. As such, quantitative differences in exam scores are interpreted as supplementary descriptive trends, not causal evidence—a limitation explicitly acknowledged in Section 6.4.

3.4 Data Collection

Triangulated data sources given in Table 1 ensured methodological rigor:

Table 1. Data collection methods

Data Type	Data Source	Collection Phase
Observational	Facilitator field notes	During gameplay session
Qualitative	Structured reflections (Google Forms)	Post-game Immediately
Quantitative	Scenario-based exam questions	1-week follow-up

1. The instructor documented observational notes on participation, group dynamics, and problem-solving strategies.
2. Following the game, students were prompted to reflect on the activity through questions. With these prompts and guidance, students were invited to complete a brief post-session survey designed to evaluate the perceived educational value of the activity. The survey consisted of three Likert-scale items and one open-ended reflection question. Students rated their agreement with statements on a five-point scale (1 = Strongly Disagree to 5 = Strongly Agree), reflecting on: (1) whether real-life project complaints helped them understand the importance of clearly defined processes, (2) whether the game format supported their focus and engagement, and (3) the extent to which the activity improved their understanding of team dynamics in software development. The open-ended question invited students to share additional feedback yielding diverse and authentic perspectives.
3. To evaluate learning outcomes, all students completed an exam one week after the intervention. The exam included scenario-based and multiple-choice questions targeting core ISO/IEC/IEEE 12207 concepts, including lifecycle processes, terminology, and process roles. This served as a post-test for both groups. Students' exam scores were recorded alongside their attendance status. Based on this, participants were categorized into the intervention or non-intervention group for subsequent analysis.

3.5 Qualitative Analysis (Primary Focus)

Thematic analysis of student reflections followed Braun and Clarke's (2006) six-phase approach to ensure methodological rigor:

1. Familiarization: All responses (n=21) were read multiple times while noting initial patterns.
2. Initial Coding: Data were coded line-by-line to identify meaningful units related to learning outcomes.
3. Theme Development: Codes were clustered into potential themes through iterative comparison.
4. Theme Review: Themes were checked against the dataset for consistency and relevance.
5. Theme Definition: Clear names and definitions were established for each theme.
6. Report Production: Selected exemplar quotes were chosen to illustrate each theme.

The following steps were employed to enhance trustworthiness:

- Peer validation: Two independent coders achieved κ=0.85 inter-rater reliability on 30% of responses
- Negative case analysis: Actively seeking disconfirming evidence (e.g., Student7's hierarchy confusion)
- Reflexive journaling: Documenting analytical decisions throughout the process

This process yielded three robust themes regarding the game's educational impact, as detailed in Section 5.

3.6 Quantitative Analysis

Performance data were analysed descriptively given sample limitations:

- Effect size caution: Cohen's d reported but interpreted conservatively
- Score distributions were visualized via boxplots to highlight variability.
- Statistical tests: Independent samples t-test noted as exploratory.

4 Game Design

To enhance students' understanding of ISO/IEC/IEEE 12207 in a practical and engaging manner, a classroom activity titled "Complaint to Compliance" was developed. The game was grounded in authentic scenarios frequently encountered in real-world software projects (see Table2) and aimed to foster experiential learning through structured peer discussion and guided analysis.

The initial complaint statements were drafted based on the author's extensive industry experience, having worked on more than 25 medium- and large-scale private sector and e-government software projects since 2005. Drawing on insights from all stages of the software development lifecycle (SDLC), the author compiled a set of 74 statements reflecting common project issues, communication breakdowns, and process deficiencies.

Table 2. Examples of complaints and relevant processes in the game

Team Member Complaint	Associated Standard Process
Tom (Technical Team Lead): "I requested vacation time for March, but I don't see it reflected in the project schedule."	6.3.1 Project Planning
Mary (Developer): "Harry (Developer), how did you decide to use this third-party library for map integration? Did you consult with the Team Lead?"	6.3.3 Decision Management
Susan (Project Manager): "We've exceeded the pessimistic estimate for the first milestone and are now one week behind schedule."	6.3.2 Project Assessment. and Control

To validate and refine the game design, the activity was piloted in two settings: the author's home university and a workshop at a Belgian university during Spring 2024 (Fig. 2). Feedback from students and facilitators in both implementations informed iterative improvements to the original 74 complaint statements. Through this process, unclear, context-dependent, or overly complex prompts were eliminated, resulting in a finalized set of 57 streamlined scenarios. These carefully curated complaints reflect authentic, recurring challenges in software projects while maintaining pedagogical clarity. Each scenario was attributed to realistic team roles (e.g., project managers, developers, devops) to enhance relatability and discussion value. The refined version was subsequently implemented at the home university in 2025, with data from this session serving as the basis for the final assessment.

Each complaint was printed on a card and randomly distributed among the students. During the activity, the instructor introduced the ISO/IEC/IEEE 12207 lifecycle processes' purposes, inputs, outputs and actions one by one.



Figure 2. "Complaint to Compliance" Game as a workshop for Engineering Students- 2024

While listening to the instructor, students were asked to match their assigned complaints to the ISO/IEC/IEEE 12207 process they believed addressed the issue. When a student identified a relevant process, they read their complaint aloud and pinned it under the corresponding category on the classroom board (see Fig. 3). In cases where matches were uncertain or debatable, the instructor facilitated open-ended discussions to encourage critical thinking and collaborative reasoning.

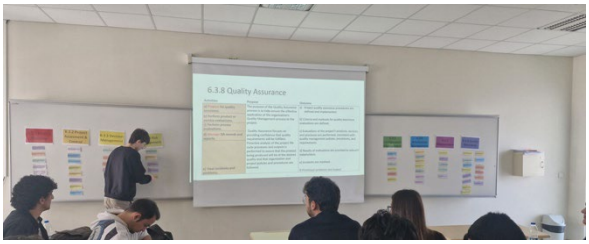


Figure 3. "Complaint to Compliance" Game as a part of Software Process Management Course - 2025

The game supported multiple valid interpretations for certain complaints—for example, a scenario could be justifiably linked to either Information Management or Configuration Management. In such cases, students were invited to justify their decisions and consider alternative perspectives, promoting flexibility in process reasoning.

The 60-minute interactive session was systematically documented through instructor field notes capturing participation dynamics and problem-solving behaviours. Post-intervention data collection employed a mixed-format survey instrument evaluating perceived utility of complaint scenarios for understanding process formalization, game format's effectiveness in maintaining engagement, and enhanced awareness of software team dynamics. Open-ended responses complemented these quantitative measures, enabling triangulation of cognitive gains and affective outcomes. This dual-method approach aligned with Kirkpatrick's (1996) framework for evaluating training interventions, assessing both immediate reactions and conceptual understanding.

5 Results

This section presents the findings from the in-class observations, post-survey and an exam for evaluating the impact of the complaint-driven instructional game on students.

5.1 Behavioral Observations

Field observations during gameplay sessions revealed striking contrasts with traditional lecture-based instruction. Participants exhibited sustained cognitive engagement, with lively peer debates emerging organically around ambiguous complaint classifications — particularly those involving cross-process dependencies. The instructor's role shifted markedly from content deliverer to facilitator, with interventions primarily focused on clarifying nuanced process boundaries in the ISO/IEC/IEEE 12207

framework. Most notably, the competitive yet collaborative dynamic fostered immediate knowledge co-construction, as students collectively reasoned through real-world scenarios while challenging each other's interpretations. This stood in sharp relief to observed passive note-taking behaviours during conventional standards instruction, suggesting the game's effectiveness in promoting active learning. Several participants spontaneously drew connections to prior project failures, indicating early development of professional diagnostic reasoning skills.

5.2 Post-Survey Results

5.2.1 Likert-scale survey results (n=21)

Survey responses revealed strong consensus regarding the intervention's pedagogical value, with near-universal endorsement of its effectiveness in bridging theoretical and practical understanding as can be seen in Fig. 4. Participants overwhelmingly agreed that the real-world complaint scenarios elucidated the necessity of formalized processes (M = 4.9/5), with qualitative reflections emphasizing how tangible examples—such as mismanaged third-party library integrations—transformed abstract clauses into actionable knowledge. The game-based format was similarly praised for sustaining engagement (M = 4.7/5), with multiple respondents contrasting its dynamic nature favourably against traditional lecture-based standards instruction. Notably, many students reported heightened awareness of role-specific challenges in software development (M = 4.5/5), suggesting the activity's utility in fostering professional mindset development.

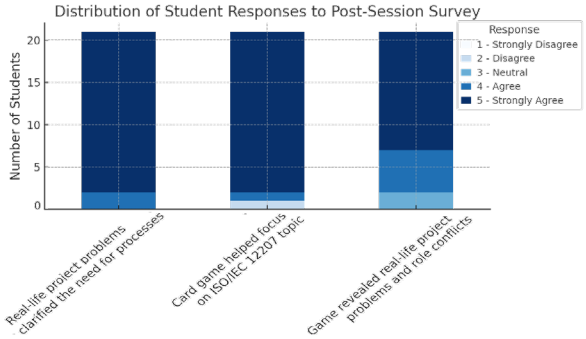


Figure 4. Distribution of student responses to post session survey

These quantitative findings were reinforced by open-ended responses, in which participants frequently described the experience as "revelatory" in connecting ISO/IEC/IEEE 12207 processes to authentic project pain points. The consistency between high Likert-scale ratings and unsolicited qualitative remarks underscores the intervention's success in making standards education both accessible and professionally relevant.

5.2.2 Thematic Analysis of Open-Ended Responses

Thematic analysis of student reflections revealed three dominant patterns, each with distinct pedagogical implications:

1. Conceptual Bridging (Theory-Practice Connections): Participants consistently emphasized how real-world complaints transformed their engagement with ISO/IEC/IEEE 12207.

This theme highlights how the game helped students bridge the gap between abstract ISO/IEC/IEEE 12207 concepts and their practical application in real-world scenarios. Students frequently mentioned how the game concretized their understanding. Here are some of the statements from students:

- "I think this game was very useful for learning the scope of the standards and experiencing their real-life application." (Student1)
- "Trying to figure out which ISO/IEC/IEEE 12207 process each card related to helped me put the theory I had learned into practice. This not only improved retention but also helped me better understand the logic behind the topic." (Student10)
- "For example, one card said the system was delivered without being tested, and I immediately thought of the verification process." (Student10)
- "I understood the concept of standards and why we need them after this game." (Student13)
- "I read my cards, figured out the problems and link them to the possible processes." (Student17)
- "The complaints on the cards made me understand how things can go wrong if people do not have written guides/processes that defines how to do a task." (Student20)
- "Made us connect ISO/IEC/IEEE 12207 standard processes with real life project cases." (Student21)

This theme underscores the game's success in reducing cognitive load (Sweller, 1988) by grounding standard clauses in concrete dilemmas.

2. Professional Identity Development: Many responses reflected growing empathy for industry roles. This theme highlights how the game led students to grasp the practical importance and value of structured processes and standards in professional software development. They moved beyond seeing standards as mere academic concepts to recognizing their crucial role in avoiding real-world project pitfalls.

- "This made me see standards as solutions." (Student14)
- "It also helps me see who is responsible for what in the team." (Student4)
- "I learned about problems in real software projects. I saw how people in different roles can have conflicts." (Student5)
- "We enjoyed the course, discussed lots of issues and started to understand difficulties that software development teams face." (Student12)

- "I understood the concept of standards and why we need them after this game." (Student13)
- "The complaints on the cards made me understand how things can go wrong if people do not have written guides/processes that defines how to do a task. And I didn't realize how many issues could arise from unclear responsibilities and poor planning. It made me see the value of having structured processes." (Student20)
- "When teacher said that we will learn about ISO/IEC/IEEE 12207 standard, I was expecting a lecture and a theoretical course that will be very hard to follow. Then she introduced the game. I was so ambitious to match my cards with the relevant processes as soon as possible correctly. So I listened the teacher and also read the slides on the projector to understand the purpose, actions, inputs, and outputs of each process." (Student21)

Such comments suggest the intervention fostered situated learning (Lave & Wenger, 1991), bridging academic and professional mindsets.

3. Engagement through Gamification: This theme encapsulates how the game's interactive, intrinsic motivation, and novel format enhanced student motivation and participation, transforming a potentially dry topic into an enjoyable learning experience. Students consistently highlighted the element of play and its impact on their engagement. Here are some examples from student reflections:

- "The fact that it was gamified was both fun and educational." (Student2)
- "It is a fun and interactive way to learn a complex standard. Instead of only reading the theory, we get to think about real problems and discuss them together." (Student4)
- "I liked the 'Complaints To Compliance' card game. The game was fun and easy." (Student5)
- "Had fun and learned a lot of best practices in this course." (Student6)
- "I think it was an interactive and fun game that really helped us learn a relatively boring topic easier." (Student8)
- "Having concrete examples from real life made the learning process enjoyable by allowing us to imagine the situations rather than overloading our minds with existing documents." (Student9)
- "Normally, classes tend to be quite monotonous, but this activity made the lesson much more interactive." (Student10)
- "It was like we were playing a game, not like we were doing a lesson. This gave me a different enthusiasm and motivation. It was nice to learn by having fun and having fun without getting bored." (Student11)
- "We enjoyed the course, discussed lots of issues..." (Student12)
- "I hope we can have more gamified courses for this type of text-based content. Otherwise, it is hard to keep my concentration while listening the professors." (Student14)
- "I think it was a very good and fun way to learn." (Student15)
- "This game motivated me about the course." (Student17)
- "The course was really interesting and fun." (Student19)
- "It was an interesting game. I did not see such an educational game before." (Student20)
- "So, it was really fun and engaging game that made us connect ISO/IEC/IEEE 12207 standard processes with real life project cases." (Student21)
- "Some of us were very competitive about matching all the cards correctly." (Student21)

Overall, the game helped bridge the gap between theoretical instruction and real-world complexity. It increased students' readiness for professional practice by promoting structured reasoning, practical awareness, and an appreciation of software process standards as actionable tools—not bureaucratic.

5.3 Quantitative Analysis

A total of 29 students completed the exam. Among them, 21 students attended the game session (Intervention Group), while 8 were absent (Non-Intervention Group). The exam, administered one week after the intervention, served as a post-test to assess learning outcomes. Although participants achieved higher exam scores, these findings remain exploratory given the study's naturalistic design and small sample size. The top 20 performers all attended the game, suggesting a strong positive impact. Students who did not participate are clustered in the lower score range, with no one exceeding 62. The quantitative patterns (see Table 3) align with qualitative findings, though the study design precludes causal claims.

Table 3. Group level statistics

Group	N	Mean	Std. Dev	Min	Max
Attended the Game	21	69.0	13.56	37	94
Did Not Attend	8	45.3	13.73	28	67

The boxplot in Fig. 5 illustrates the tighter score distribution and higher median among students who participated in the game.

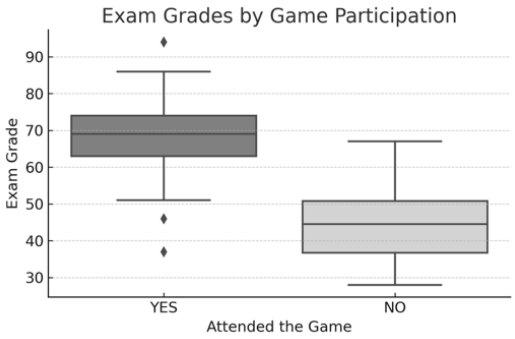


Figure 5. Distribution of Exam Grades by Game Participation

6 Discussion

6.1 Innovation in Design

The Complaint to Compliance game introduces three key innovations in standards education:

1. **Authentic Scenario Design**: Grounded in 20 years of industry experience (25+ projects), the game's 57 complaint cards operationalize *situated cognition* (Greeno, 1998) by embedding ISO/IEC/IEEE 12207 processes in tangible project failures. Unlike abstract examples (e.g., Calderón et al., 2018a), these scenarios mirror what Herranz et al. (2016) term the "simulation gap" — the disconnect between standards and their real-world consequences.
2. **Multi-Dimensional Mapping**: The game intentionally includes complaints spanning multiple standard clauses (e.g., Configuration Management *and* Quality Assurance), reflecting the interconnected nature of software processes in practice. This design choice cultivates *systems thinking* (Senge, 1990), a skill critical for professional practice but neglected in conventional standards pedagogy.
3. **Low-Tech Accessibility**: By leveraging physical cards and peer debate instead of digital interfaces, the intervention achieves scalability while addressing resource constraints noted in global engineering education (Koyuncu Tunç, 2024). This aligns with García et al.'s (2020) calls for "pedagogical minimalism" in standards instruction.

The design embodies Lave and Wenger's (1991) legitimate peripheral participation by positioning students as analysts diagnosing authentic process breakdowns — a significant departure from passive, lecture-based standards training.

6.2 Comparison with Related Work

The Complaint to Compliance game advances beyond existing serious games for standards education through three key distinctions, while maintaining alignment with their shared pedagogical goals:

1. **Diagnostic vs. Procedural Focus**: While MiProJOC (Mesquida & Mas, 2018) trains students in planned process execution (e.g., ISO/IEC 12207 project scheduling), this game emphasizes retrospective failure analysis—a higher-order cognitive skill (Bloom, 1956) critical for process improvement in industry. This addresses Sánchez-Gordón's (2016) observed gap in "abstraction-to-application" transitions.
2. **Scalability**: Unlike Metric Cake Shop (García et al., 2025), which requires specialized measurement tools, this card-based design achieves comparable engagement (M=4.7/5) with minimal resources—validating Calderón et al.'s (2018a) argument for scalable interventions in resource-constrained institutions.
3. **Professional Relevance**: Where Aydan et al. (2017) teach standard components in isolation, this complaint-driven approach demonstrates their practical utility in mitigating project risks (e.g., untested deliveries linking to Verification Processes). This directly tackles Calderón et al.'s (2018a) finding that students dismiss standards as "bureaucratic" without real-world context.

Convergence: All approaches successfully leverage gamification to overcome Sánchez-Gordón's (2016) "abstraction barrier," but this method uniquely combines:

- Cognitive authenticity (diagnosing real project failures)
- Accessibility (low-tech materials)
- Professional framing (standards as preventive tools)

6.3 Practical Implications and Transferability

The game's design offers several practical implications:

1. **Curriculum Integration:** The Complaint to Compliance game offers versatile implementation across educational and professional contexts. Its 60-minute format fits standard class periods, serving effectively as either an introductory engagement activity or capstone exercise, addressing common scheduling constraints in software engineering curricula (Calderón et al., 2018b).
2. **Adaptability:** The framework's adaptable methodology extends to other standards like ISO/IEC 29110 (for small entities) through tailored complaint sets, with a replicable process for deriving scenarios from project retrospectives (García et al., 2020).
3. **Professional Training:** Pilot workshops demonstrate its value for industry onboarding and audit training, particularly in emerging economies

where Koyuncu Tunç (2024) identified gaps between academic standards education and industry needs.

Successful implementation across diverse contexts - from Belgian university workshops to Turkish undergraduate courses - confirms its scalability. The minimal resource requirements further enhance adoption potential for both academic and corporate training environments.

6.4 Limitations

Three key limitations warrant consideration. First, while implemented at two universities, the cultural context limits generalizability to diverse educational settings. Second, the absence of longitudinal data prevents assessment of long-term knowledge retention and professional application - a critical gap given the study's emphasis on real-world relevance. Third, the small control group (n=8) and non-randomized design, though methodologically justified for this exploratory study, constrain causal claims about effectiveness. As a case study, this research prioritizes rich contextual insights over generalizability. The lack of randomized controls and small sample size prevent causal inferences, but the qualitative findings provide valuable pedagogical insights.

6.5 Future Enhancements

Future research should pursue: (1) multi-institutional validation across different cultural contexts, particularly in developing economies where standards education faces unique challenges (Koyuncu Tunç, 2024); (2) longitudinal tracking of students into industry roles; and (3) controlled studies with larger samples. Technical enhancements could include a digital platform for asynchronous play and AI-generated scenarios (Zhang et al., 2023), while maintaining the low-tech option's accessibility.

These limitations notwithstanding, the qualitative findings suggest the game's potential for making standards tangible. The game's success in making ISO/IEC/IEEE 12207 tangible suggests a promising foundation for broader adoption.

7 Conclusion

This study presents Complaint to Compliance, an innovative low-tech game that transforms ISO/IEC/IEEE 12207 standards education from abstract theory to engaged practice. Through qualitative analysis of student experiences, we identified three key pedagogical benefits: (1) strengthened connections between theoretical standards and practical applications, (2) development of professional identity through process-aware thinking, and (3) enhanced engagement via collaborative problem-solving.

The game's significance lies in its dual accessibility:

- **Conceptual**: Making complex standards tangible through real-world project failures
- **Logistical**: Requiring only printed materials for implementation

The findings suggest this approach successfully addresses two critical gaps in engineering education:

- The relevance gap identified by Koyuncu Tunç (2024), where students struggle to see standards as practical tools
- The resource gap in emerging economies, where high-tech solutions may be impractical

For educators, this work offers:

- A ready-to-implement template for standards education
- Evidence that low-tech gamification can achieve high engagement
- A model for developing professional mindset alongside technical knowledge

Future research should explore:

- Longitudinal studies to assess knowledge retention
- Multi-institutional implementations across cultural contexts
- Potential integrations with digital tools (e.g., AI-generated scenarios)

This study ultimately demonstrates that effective standards education requires moving beyond passive instruction to active, context-rich experiences. By positioning students as problem-solvers analysing real project challenges, we prepare them not just to know standards, but to use them as instruments for quality in professional practice.

Acknowledgments

I sincerely thank Prof. Dr. Hayri Sever for encouraging me to teach the Software Process Management and Quality Standards courses. His trust and support motivated me to approach these subjects creatively and explore new methods in education. I also thank my students for their active participation and valuable contributions to this research.

References

Aydan, U., Yilmaz, M., Clarke, P. M., & O'Connor, R. V. (2017). Teaching ISO/IEC 12207 software lifecycle processes: A serious game approach. Computer Standards and Interfaces, 54, 129–138. Scopus. https://doi.org/10.1016/j.csi.2016.11.015

Aydan, U., Yilmaz, M., O'Connor, R.V. (2015). Towards a Serious Game to Teach ISO/IEC 12207

Software Lifecycle Process: An Interactive Learning Approach. In: Rout, T., O'Connor, R., Dorling, A. (eds) Software Process Improvement and Capability Determination. SPICE 2015. Communications in Computer and Information Science, vol 526. Springer, Cham. https://doi.org/10.1007/978-3-319-19860-6_17

Bloom, B.S. (1956). Taxonomy of Educational Objectives. David McKay.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Calderón, A., M. Ruiz, and R.V. O'Connor. (2017a). "Coverage of ISO/IEC 29110 Project Management Process of Basic Profile by a Serious Game." In Commun. Comput. Info. Sci., eds. Messnarz R., Stolfa J., Stolfa S., and O'Connor R.V. Springer Verlag, 111–22. doi:10.1007/978-3-319-64218-5_9.

Calderón, A., Ruiz, M., & O'Connor, R. V. (2017b). Coverage of the ISO 21500 standard in the context of software project management by a simulation-based serious game. In O'Connor R.V., Mas A., Rout T., Mesquida A., & Dorling A. (Eds.), Commun. Comput. Info. Sci. (Vol. 770, pp. 399–412). Springer Verlag; Scopus. https://doi.org/10.1007/978-3-319-67383-7_29

Calderón, A., M. Ruiz, and R.V. O'Connor. 2018a. "A Multivocal Literature Review on Serious Games for Software Process Standards Education." Computer Standards and Interfaces 57: 36–48. doi:10.1016/j.csi.2017.11.003.

Calderón, A., M. Ruiz, and R.V. O'Connor. 2018b. "A Serious Game to Support the ISO 21500 Standard Education in the Context of Software Project Management." Computer Standards and Interfaces 60: 80–92. doi:10.1016/j.csi.2018.04.012.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Routledge.

García, I., Pacheco, C., León, A., & Calvo-Manzano, J. A. (2020). A serious game for teaching the fundamentals of ISO/IEC/IEEE 29148 systems and software engineering – Lifecycle processes – Requirements engineering at undergraduate level. Computer Standards and Interfaces, 67. Scopus. https://doi.org/10.1016/j.csi.2019.103378.

García, I., Pacheco, C., López, I., Calvo-Manzano, J. A., & Flores-Rios, B. L. (2025). Metric cake shop: A serious game for supporting education on ISO/IEC/IEEE 15939:2017 – Systems and software engineering – Measurement process in the context of an undergraduate software engineering course. Computer Standards and Interfaces, 91. Scopus. https://doi.org/10.1016/j.csi.2024.103880

Greeno, J.G. (1998). The situativity of knowing. Educational Researcher, 27(1), 5-11.

Herranz, E., Colomo-Palacios, R., de Amescua Seco, A., & Sánchez-Gordón, M.-L. (2016). Towards a gamification framework for software process improvement initiatives: Construction and validation. Journal of Universal Computer Science, 22(12), 1509–1532. Scopus.

ISO/IEC/IEEE. (2017). *ISO/IEC/IEEE 12207:2017: Systems and software engineering – Software life cycle processes*. Geneva: International Organization for Standardization.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (1996). *Evaluating training programs: The four levels*. Berrett-Koehler Publishers.

Koyuncu Tunç,S. (2024) "Investigating the Adoption of International Software Quality Standards in Turkiye: A Comprehensive Analysis," 9th International Conference on Computer Science and Engineering (UBMK), Antalya, Turkiye, 2024, pp. 1052-1057, doi: 10.1109/UBMK63289.2024.10773457.

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge University Press.

Mesquida, A.-L., & Mas, A. (2018). Experiences on the use of a game for improving learning and assessing knowledge. Computer Applications in Engineering Education, 26(6), 2058–2070. Scopus. https://doi.org/10.1002/cae.21991.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1–13. https://doi.org/10.1177/1609406917733847

Senge, P.M. (1990). The Fifth Discipline. Doubleday.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Social Robots and AI-Driven Learning: A Pilot Study on Student Engagement in Higher Education

Daniel Hari, Maja Kerneža, Dejan Zemljak, Andrej Flogie
Faculty of Science and Mathematics
Institute of Modern Technologies
Koroška cesta 160, Maribor, Slovenia
{daniel.hari, maja.kerneza1, dejan.zemljak1, andrej.flogie}@um.si

***Abstract**. This pilot study explores how the Slovenian-speaking social robot AlphaMini, powered by ChatGPT, affects student engagement in higher education. Conducted with 22 students in a knowledge management class, the research used questionnaires and classroom observation to assess behavioral, emotional, and cognitive engagement. Students reported higher levels of emotional and behavioral engagement during the session with AlphaMini. Most students viewed the robot positively and recognized its potential for inclusive education. These findings suggest that localized social robots can motivate students and foster engagement in both higher and early learning environments.*

Keywords. artificial intelligence, social robot, chatGPT, student motivation, deducation

1 Introduction

The integration of Artificial Intelligence (AI) into education is reshaping teaching practices and creating new opportunities for active, personalized, and interactive learning. Among these emerging technologies, social robots powered by generative AI are gaining traction due to their ability to communicate in learners' native languages and simulate human-like interaction. Such robots are particularly promising for enhancing student engagement and motivation, which are key factors in successful learning (Hameed et al., 2023, Cai et al., 2024). One notable example is AlphaMini, a compact, humanoid robot equipped with speech recognition, facial detection, and the capacity for emotional expression. AlphaMini has been adapted to communicate in Slovenian, making it particularly valuable in contexts where localized tools are scarce. Social robots like AlphaMini and NAO are designed to foster natural, socially intelligent interaction through voice, gestures, facial expressions, and emotional cues. Although they are not physically human, their expressive features support the experience of social presence the sense of being in the company of another socially aware being which contributes to the development of emotional bonds and peer-like relationships with students (Breazeal, 2003; Belpaeme et al., 2018). Prior research shows that social robots can function effectively as peer tutors, especially in language and STEM education (Kanda et al., 2011; Augello et al., 2020). Their consistent, non-judgmental, and emotionally responsive interactions offer advantages over traditional digital tools, which often lack a human or affective dimension (Hameed et al., 2023, Cai et al., 2024). In the Slovenian context, where few AI-based educational tools are tailored to the local language and culture, the presence of a Slovenian-speaking robot like AlphaMini represents a step toward more inclusive and equitable educational technologies (Nityashree et al., 2023). Social robots also align well with the goals of 21st-century education, supporting the development of collaboration, communication, critical thinking, and digital competence (Aberšek & Flogie, 2022). These capabilities are especially relevant when teaching abstract and cognitively demanding topics such as knowledge management, where verbal clarity, emotional engagement, and real-time interaction can aid comprehension. Despite the rising use of generative AI in classrooms particularly through tools like ChatGPT or Copilot there remains a notable lack of empirical research on the impact of such technologies when embodied in physical robots (Zhang & Aslan, 2021; Augello et al., 2020).

Most studies focus on chatbot-based tutoring systems, intelligent learning platforms, or writing assistants (Zawacki-Richter et al., 2019; Holmes et al., 2022), while only a few explore the pedagogical potential of embodied AI, especially in formal educational settings (Belpaeme et al., 2018; Mubin et al., 2013). The combination of a robot's physical presence with the linguistic capabilities of generative AI presents a unique learning experience. Unlike screen-based tools, social robots offer embodied, affective, and interactive communication, which can strengthen emotional and behavioral engagement. When paired with AI systems like ChatGPT, these robots are capable of dynamic, unscripted, and context-sensitive responses, making them highly suitable for exploratory learning and dialogue-rich environments (Hameed et al., 2023, Cai et al., 2024). Prior research

suggests that this blend of embodiment and intelligent language use contributes to deeper motivation, trust, and inclusivity in diverse learning contexts (Fridin, 2014; Kanda et al., 2011). This pilot study addresses the current research gap by examining how a Slovenian-speaking social robot integrated with generative AI affects university students' engagement when learning abstract topics such as knowledge management. The study was conducted with part-time students enrolled in management and logistics courses at the Faculty of Commercial and Business Sciences. To guide our investigation, we posed the following research questions (RQ) and hypotheses (H): RQ1: How does the presence of the AlphaMini social robot in the classroom affect student engagement across behavioral, emotional, and cognitive dimensions?

- H1a: The use of AlphaMini will significantly increase students' emotional engagement.
- H1b: The use of AlphaMini will significantly increase students' behavioral engagement.
- H1c: The use of AlphaMini will moderately increase students' cognitive engagement.

RQ2: How does prior experience with social robots influence students' engagement during the lesson?

- H2: Students with previous experience will report higher emotional engagement.
- H3: Students with prior experience will display greater behavioral engagement.
- H4: Students with prior experience will demonstrate stronger cognitive engagement.

The remainder of this paper is structured as follows. Section 2 outlines the methodology and instruments used in the pilot study. Section 3 presents the results, followed by a discussion in Section 4. The final section summarizes the main conclusions and offers implications for future educational use of AI and robotics.

2 Methodology

This study used a simple case study approach, combining a short classroom experiment with a questionnaire and basic observation of student reactions. The questionnaire comprised nine items designed to assess three dimensions of engagement: behavioral (e.g., "I participated more because of the robot"), emotional (e.g., "I felt comfortable and curious"), and cognitive (e.g., "I understood the topic better"). Each dimension included three items, rated on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). The questionnaire items were adapted from established engagement measurement tools, including the Student Engagement Instrument (Appleton et al., 2006) and similar studies on educational robotics (Belpaeme et al., 2018). An additional item inquired about prior experience with social robots (yes/no).

The goal was to find out how students respond when a social robot like AlphaMini is used in teaching. The study included 22 master's students from the Faculty of Management and Logistics. All students attended a live classroom session on knowledge management, where the robot AlphaMini presented key ideas in Slovenian, supported by ChatGPT. After the session, students filled out a short online questionnaire, where they shared how they felt during the lesson how focused they were, how much they participated, and whether they felt comfortable or motivated. We also observed the session and noted how students interacted with the robot for example, if they smiled, asked questions, or worked together in small groups.

The Table 1 below shows the basic information about the participants:

Table 1. Number and Percentage of Students by Field of Study

	N	%	M (SD)
Gender (total)	22	100	
Male	5	22.7	
Female	15	68.2	
Age			36.4 (10.3)
Field of study			
Business and Administrative Sciences, law, management	22	100	
Employment status (total)	22	100	
Employed	21	95.5	
Not (yet) employed	1	4.5	

Note: N = number of respondents, M = mean, SD = standard deviation.

Two respondents (9,1 %) opted not to answer the question regarding gender identity.

The questionnaire focused on three types of engagement: Behavioral (e.g., Did I participate more because of the robot?), Emotional (e.g., Did I feel comfortable and curious?), Cognitive (e.g., Did I understand the topic better?). Students answered on a scale from 1 (strongly disagree) to 5 (strongly agree). We also asked whether students had used a social robot before. This helped us compare if students with experience responded differently than those seeing the robot for the first time. During the lesson, we looked for visible signs of engagement, such as:

- Talking to the robot
- Paying attention
- Asking questions
- Smiling, laughing, or showing interest

These observations helped us better understand the answers from the questionnaire. For example, some

students who said they felt more motivated also showed more visible curiosity during the session.

3 Results

The results of the questionnaire show that students responded positively to the use of the social robot AlphaMini during the lesson. Due to the small sample size and exploratory nature of the pilot study, inferential statistical tests such as t-tests or ANOVA were not conducted. Instead, group means and Pearson correlations were used to identify patterns in the data. From Table 2 we see that all three types of engagement behavioral, emotional, and cognitive were rated fairly high on average. The correlation table shows how closely connected the three types of engagement were:

Table 2. Pearson Correlation Matrix Between Dimensions of Student Engagement

	1	2	3
1. Behavioral engagement	1	0.76**	0.47*
2. Emotional engagement		1	0.65**
3. Cognitive engagement			1

Note: **p < 0.01; * p < 0.05

The strongest relationship was between behavioral and emotional engagement (r = 0.76), which means that students who were more emotionally involved also tended to participate more actively in the lesson. There was also a moderate link between emotional and cognitive engagement (r = 0.65), suggesting that feeling comfortable and interested helped students focus and think more deeply. The weakest but still meaningful connection was between behavioral and cognitive engagement (r = 0.47). These results support RQ1, showing that the robot increased both emotional and behavioral involvement, with cognitive engagement also improving, though slightly less. To explore whether previous experience with social robots affected engagement (RQ2), we compared average scores between students who had already used a robot and those who had not.

Table 3. Mean Differences in Engagement Based on Prior Experience with a Social Robot

	Experience with Social Robot	
	Yes	No
Variables	M (SD)	M (SD)
Behavioral engagement	4.1 (0.5)	3.58 (0.7)
Emotional engagement	3.54 (0.3)	3.43 (0.5)
Cognitive engagement	3.42 (0.7)	3.2 (0.7)

Note: M = Mean, SD = Standard Deviation

From Table 3 we see that students who had prior experience with social robots reported higher scores across all three areas: The biggest difference was in behavioral engagement, suggesting that familiarity helped students feel more confident and active in the session. Emotional engagement was also higher, although the difference was smaller. Cognitive engagement followed a similar trend. These findings support H2, H3, and H4, showing that previous exposure to social robots may increase comfort, interest, and participation. Most students responded positively to AlphaMini. Students with robot experience were more engaged overall. The strong link between emotional and behavioral engagement suggests that students need to feel safe and interested before they are ready to participate actively. Even students with no prior experience rated their engagement relatively high, which shows that AlphaMini was well accepted even as a novelty. These results suggest an association between the presence of AlphaMini and increased emotional and behavioral engagement.

4 Discussion

The aim of this pilot study was to explore how the use of a Slovenian-speaking social robot, AlphaMini, impacts student engagement during a university-level session on knowledge management. The results show that students generally responded positively to the robot-led session, especially in terms of emotional and behavioral engagement. One of the most important findings is the strong connection between emotional and behavioral engagement. Students who felt comfortable, curious, or emotionally supported by the robot also participated more actively in the lesson. This supports earlier research that emphasizes the importance of social presence and emotional safety in learning (Belpaeme et al., 2018; Fridin, 2014). Social robots, especially when speaking in the learners' native language, seem to lower barriers to participation and encourage students to be more open and responsive.

The results also showed a moderate improvement in cognitive engagement, meaning that the robot helped students think more about the topic and possibly understand it better. However, this effect was not as strong as for emotional and behavioral aspects. This may be because cognitive engagement depends not only on how the material is presented, but also on content difficulty, prior knowledge, and learning habits. Still, the increase suggests that even complex topics like knowledge management can benefit from

more engaging delivery formats. When comparing students with and without prior experience with social robots, the results clearly show that familiarity helps. Students who had interacted with a robot before were more engaged on all levels.

This suggests that as social robots become more common in classrooms, students may respond to them even more positively over time. It also shows the importance of gradual introduction and training, especially in higher education, where such technologies are still relatively new. Classroom observations further confirmed the questionnaire results. Many students laughed, asked questions, and made positive comments about the robot. A few mentioned that such robots would be especially helpful for children, or in settings where students are shy, anxious, or need more emotional support.

5 Conclusions

This pilot study shows that a Slovenian-speaking social robot supported by generative AI can positively influence student engagement in higher education, particularly in lessons involving abstract or theoretical topics. The strongest impact was seen in emotional and behavioral engagement, with moderate gains in cognitive engagement. Students appreciated the robot's gestures, natural speech, and supportive presence. Due to the non-experimental design and small sample size, we cannot make causal claims regarding the effect of the robot on engagement. Those with previous robot experience showed higher overall engagement in our pilot study, suggesting that familiarity and continued exposure play a key role in acceptance and motivation. Although this was a small-scale study, the results are promising. A potential limitation of this study is also the novelty effect, as students' engagement may have been temporarily elevated by the introduction of the robot (Kennedy et al., 2023). Furthermore, the small sample size and lack of statistical testing limit the generalizability of our findings. Our findings align with previous research suggesting social robots can support engagement in a variety of educational settings (Belpaeme et al., 2018; Augello et al., 2020).

While our study was limited to a university context, future research could explore these applications in primary and inclusive education. Their ability to combine emotional intelligence with structured content delivery makes them a valuable complement to traditional teaching methods. Future studies could include a larger number of participants, test the effects over a longer period, and compare robot-led teaching with traditional instruction. Additional research could also explore learning outcomes, not just engagement, to better understand how social robots impact actual knowledge acquisition. In conclusion, the combination of social robotics and generative AI when adapted to the local language and context offers a promising direction for more inclusive, engaging, and emotionally aware education. Future research should employ longitudinal designs and larger samples to better assess the sustained impact of social robots in education.

References

Aberšek, B., & Flogie, A. (2022). Umetna inteligenca in prihodnost učenja in poučevanja.

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. Journal of School Psychology, 44(5), 427–445. https://doi.org/10.1016/j.jsp.2006.04.002

Augello, A., Daniela, L., Gentile, M., Ifenthaler, D., & Pilato, G. (2020). Robot-assisted learning and education. Frontiers in Robotics and AI, 7, Article 591319. https://doi.org/10.3389/frobt.2020.591319

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. Science Robotics, 3(21), eaat5954. https://doi.org/10.1126/scirobotics.aat5954

Breazeal, C. (2003). Designing sociable robots. MIT Press.

Cai, Y., Pan, X., Wang, L., Yu, Y., & Dong, X. (2024). Exploring the impact of robot interaction on learning engagement: A comparative study of two multi-modal robots. Frontiers in Robotics and AI, 11, Article 1288019. https://www.researchgate.net/publication/388495807_Exploring_the_impact_of_robot_interaction_on_learning_engagement_a_comparative_study_of_two_multi-modal_robots

Fridin, M. (2014). Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. Computers & Education, 70, 53–64. https://www.sciencedirect.com/science/article/pii/S036013151300225X

Hameed, I. A., Quist, J. S., Smedegaard, S., & Juel, K. (2023). How social robots can influence motivation as motivators in learning: A scoping review. Frontiers in Robotics and AI, 10, Article 1204662. https://www.researchgate.net/publication/373065707_How_Social_Robots_can_Influence_Motivation_as_Motivators_in_Learning_A_Scoping_Review

Holmes, W., Bialik, M., & Fadel, C. (2022). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2011). Interactive robots as social partners and peer tutors for children: A field trial. Human–Computer Interaction, 19(1–2), 61–84. https://www.tandfonline.com/doi/abs/10.1080/07370024.2004.9667340

K. K., Nityashree, AryaRa, & Tabassum B. Shivanagi. (2023). Humanoid robotics in artificial intelligence.

Mubin, O., Stevens, C. J., Shahid, S., Mahmud, A. A., & Dong, J.-J. (2013). A review of the applicability of robots in education. Technology for Education and Learning, 1(1), 1–7.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. Journal of Educational Technology, 3(1). https://www.sciencedirect.com/science/article/pii/S2666920X21000199

Exploring Teachers' Acceptance of Digital Textbooks: A Pilot Study in Croatian Primary Schools

Melita Milić, Blaženka Divjak
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{mzahtila, blazenka.divjak}@foi.hr

Abstract. *This pilot study aims to validate a measurement instrument based on an extended UTAUT model, examining factors related to behavioral intention toward digital textbooks (DT) among Croatian primary school subject teachers. Data were analyzed using PLS-SEM, suitable for small samples and non-normal distributions, to assess model adequacy of the proposed model. Fifty-two teachers from 12 counties participated; analysis focused on 39 active DT users. Results show positive attitudes, high perceived usefulness and ease of use, and low perceived workload. Despite non-significant predictors, the model's high explanatory power supports further refinement for broader DT adoption research.*

Keywords. digital textbook, e-textbook, technology acceptance, UTAUT, PLS-SEM, primary education

1 Introduction

Rapid technological changes significantly affect the type of knowledge, skills, and competencies that are imperative for maintaining national competitiveness and individual well-being. The education system is increasingly becoming part of the digital transformation, which in the Republic of Croatia (RC) has been driven by the pilot project "e-Schools" ('E-Škole', n.d.) and the comprehensive curricular reform (School for life, n.d.). As part of this project, in addition to systematic teacher training, **digital textbooks (DT)** were developed for STEM subjects (*Materijali*, n.d.), to support innovative learning and teaching methods. The importance of early technology integration into the teaching process was particularly recognized during the COVID-19 pandemic (Divjak, 2022).

The introduction of DT represents a significant step forward, offering easy access, interactive content, and learning unbound by time or physical constraints (Hermita et al., 2023; Joo et al., 2017; Wijaya et al., 2022). Competent and effective use of DT in teaching can improve educational quality and inclusiveness, facilitating access to materials and supporting students with special educational needs and vulnerable groups (Grönlund et al., 2018; Kabugo, 2020). DT also enable and support personalized learning through adaptable interfaces.

Despite these advantages, implementing DT faces challenges such as infrastructure deficits, the need for additional support and teacher training, and inconsistent improvements in learning outcomes (Sun & Jiang, 2015). Successful implementation depends on various factors, including teacher attitudes and readiness to use them (Wijaya et al., 2022). International studies like Teaching and Learning International Survey (TALIS) (Dekanić et al., n.d.), provide important insights into teachers' work conditions and workload as well as technology acceptance. In Croatia, administrative workload (60% teachers cite it as a major strass factor) and responsibility for student success (56% teachers cite it as a major strass factor) are cited as major sources of stress among teachers.

In general, most studies focus on student attitudes toward DT, while there is a significant lack of research addressing teachers' perspectives (Milić & Divjak, 2023). To address the gap, this paper aims to collect teachers' views and needs in the context of validating a measurement instrument used in a pilot study of factors affecting behavioral intention to use DT in subject-specific teaching in Croatian primary schools.

2 Theoretical Framework and Hypotheses

This pilot study is theoretically grounded in the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003).

2.1 Theoretical Constructs of Intention to Use Digital Textbooks (DT)

UTAUT provides a framework for understanding the factors that influence **behavioral intention (BI)** to adopt and use new technology, as well as actual usage behavior. The model identifies four core constructs: Performance Expectancy (PE), Effort Expectancy

(EE), Social Influence (SI) and Facilitating Conditions (FC) along with Behavioral Intention (BI). In addition to the UTAUT constructs, this study also investigates three more factors: Pedagogical Potential (PP), Workload (WLO) and Attitude (AT). All constructs used in this study are explained below in the context of education. AT, WLO, and PP extend UTAUT with affective, contextual, and pedagogical factors relevant to Croatian schools (Cukurova et al., 2023; Hermita et al., 2023).

BI (Behavioral Intention) is defined as an individual's internal orientation reflecting their willingness to perform a certain behavior (Venkatesh et al., 2003). It is empirically confirmed to be a significant predictor of many behaviors, including technology acceptance (Shukla, 2021).

PE (Performance Expectancy) refers to teachers' perceptions of how DT can enhance quality of learning and teaching, ease preparation and administrative tasks, and improve student acquisition of learning outcomes (Hermita et al., 2023; Shukla, 2021).

EE (Effort Expectancy) encompasses the perceived ease of use of DT, intuitive interface, and minimal effort required for integration into teaching and learning (Hermita et al., 2023).

SI (Social Influence) includes the perceived support and expectations from peers and superiors regarding the use of DT (Hermita et al., 2023)

FC (Facilitating Conditions) relates to the availability of necessary equipment (e.g. computers, tablets), reliable internet access, technical support, and adequate training to use DT (Kahnbach et al., 2024).

PP (Pedagogical Potential) represents teachers' perceptions of how DT can improve teaching and learning and lead to learning outcomes, with a positive influence on BI (Hermita et al., 2023).

WLO (Workload) describes teachers' perceptions of the additional effort & time that using DT may require, e.g. lesson preparation, technical difficulties, extra training, and increased administrative tasks. A negative impact on BI is expected (Cukurova et al., 2023).

AT (Attitude) encompasses teachers' beliefs, feelings, and behavioral tendencies toward DT. A positive attitude is likely to result in stronger BI, while a negative attitude may represent a major barrier to implementation (Hermita et al., 2023).

By including both the core UTAUT constructs and the additional factors, this model aims to better explain the **behavioral intention** of Croatian teachers to use DT in classroom practice.

2.2 Research Questions and Hypotheses

Research Questions: *What are the attitudes, needs, and experiences of teachers in Croatia related to the use of DT? What factors are important for their use?*

Hypotheses linking latent factors to BI to use DT:
H1: Performance Expectancy positively affects Behavioral Intention to use DT.
H2: Effort Expectancy positively affects Behavioral Intention to use DT.
H3: Attitude toward DT positively affects Behavioral Intention to use DT.
H4: Social Influence positively affects Behavioral Intention to use DT.
H5: Facilitating Conditions positively affect Behavioral Intention to use DT.
H6: Workload negatively affects Behavioral Intention to use DT.
H7: Pedagogical Potential positively affects Behavioral Intention to use DT.

Hypotheses H1, H2, H4, and H5 are based on UTAUT (Venkatesh et al., 2003) and widely supported in education (Hermita et al., 2023; Kahnbach et al., 2024; Viberg et al., 2020; Wijaya et al., 2022). Attitude is often added to capture affective factors, while workload may hinder and pedagogical potential promote adoption (Cukurova et al., 2023; Hermita et al., 2023)

3 Methodology

This pilot study employed a quantitative approach to examine factors influencing teachers' behavioral intention to use digital textbooks, focusing on measurement model validation.

3.1 Sample and Data Collection

The study included 52 subject teachers in the upper grades of primary school (grades 5–8 in Croatia) primary school teachers from 12 counties in the Republic of Croatia. In Croatia, subject teachers in the upper grades of primary school (grades 5–8) teach specific subjects such as mathematics, biology, or history. The participants represented various age groups (25–30, 31–40, 41–50, 51+), genders, and levels of work experience (novice: <5 years, mid-career: 5–15 years, experienced: >15 years). Data were collected online during April 2025, and participation was voluntary and anonymous, with informed consent obtained. Of the 52 teachers who participated, analysis focused on 39 active DT users. Given the small sample and non-normal data distribution, PLS-SEM was selected as an appropriate method. Pilot studies with 30–50 participants can still yield valuable insights into measurement model adequacy, especially when hypothesis testing is not the primary aim (Hair et al., 2022).

3.2 Measurement Instrument

The questionnaire consisted of 41 items distributed across seven constructs, along with demographic questions. Most constructs (PE, EE, FC, BI) were measured with three items each, while SI and AT had four items, PP had five, and WLO had two. Responses to items were collected using a 5-point Likert type of

scale (1 = strongly disagree, 5 = strongly agree). Each construct was measured through multiple items forming part of a composite scale. The questionnaire was translated into Croatian and then back-translated into English for accuracy, following international guidelines for cross-cultural scale adaptation (Cruchinho et al., 2024).

3.3 Data Analysis

PE, EE, BI, and AT were modeled reflectively, following the original UTAUT (Venkatesh et al., 2003), while SI, FC, WLO, and PP were specified formatively, as their indicators reflect distinct, non-interchangeable aspects. Despite reflective phrasing, formative modeling better captures their multidimensionality in the Croatian educational context. Data analysis was conducted using the PLS-SEM method, with the **SmartPLS** software (version 4.1.1.2).

For **reflective constructs**, the measurement model assessment included: Item factor loadings (>0.70); Indicator VIF values (<5); Construct reliability (Cronbach's alpha >0.70); Composite reliability (CR >0.70) and Convergent validity using Average Variance Extracted (AVE >0.50).

For **formative constructs**, outer weights and VIF values for indicators were checked (<5). **Discriminant validity** was assessed using: Fornell-Larcker criterion and Heterotrait-Monotrait Ratio (HTMT <0.85). **The structural model assessment** included: Analysis of path coefficients (β) and their statistical significance (bootstrapping with 5,000 samples); Effect sizes (f^2); Multicollinearity among latent variables and Explained variance (R^2) of endogenous constructs.

4 Results

Before assessing the measurement and structural models in detail, a preliminary analysis of the collected data was conducted to ensure suitability for PLS-SEM analysis. The analysis showed no missing data.

4.1 Descriptive Statistics

4.1.1 Sample Description

Although 52 teachers participated in the pilot study, only those who actively use digital textbooks were included in the validation of the instrument, resulting in a final sample of 39 participants (N=39).

Table 1. shows the demographic profile of the sample. Of the 39 respondents, the majority were female (69%, N=27). The most represented age group was 50–59 years (44%, N=17). Regarding work experience, the largest portion had 21–30 years of experience (33%, N=13), followed by those with 31+ years (26%, N=10).

4.1.2 Indicator Statistics

As shown in Table 2., analysis of arithmetic means revealed generally positive attitudes toward the questionnaire items. Most items had average values above 3.50 on a 5-point Likert scale. Particularly high averages (above 4) and medians (4 or 5) were noted for items under: Performance Expectancy (PE1 = 4.18, PE2 = 3.97, PE3 = 4.10), Effort Expectancy (EE2 = 4.28), Social Influence (SI1 = 4.41, SI4 = 4.00), Facilitating Conditions (FC1 = 4.49, FC2 = 4.70) and Behavior Intention (BI1 = 4.51, BI2 = 4.33). The lowest averages were for Workload (WLO1 = WLO2 = 2,59), and BI3 (2.90), followed by Pedagogical Potential (PP4 = 3.54, PP5 = 3.44).

Table 1. Demographic profile

Characteristic	Category	N	%
Gender	M	12	31%
	F	27	69%
Age	21 - 29	4	10%
	30 - 39	5	13%
	40 - 49	10	26%
	50 - 59	17	44%
	60+	3	8%
Work experience (years)	<5	8	21%
	5 - 10	4	10%
	11 - 20	4	10%
	21 - 30	13	33%
	31 +	10	26%

Standard deviations mostly ranged between 0.57 and 1.14, indicating moderate to somewhat pronounced response variability. The lowest variability was seen in FC2 (SD = 0.57), and the highest in EE3 (SD = 1.14) and PP5 (SD = 1.11), suggesting a wide range of opinions.

Kurtosis values were negative for all items except WLO1 and WLO2, suggesting left-skewed distributions typical for positively worded Likert-scale items. The most negatively skewed was FC1 (-1.81). In contrast, WLO1 and WLO2 had slightly positive skewness, indicating concentration of responses on the lower end of the scale. PLS-SEM was chosen given the non-normality of several indicators (Table 2), as it is robust to deviations from multivariate normality (Hair et al., 2022).

4.2 Measurement Model Evaluation

For reflective constructs, high internal consistency was confirmed via: Cronbach's α, Composite reliability (ρ_C), Indicator reliability (ρ_A) and Convergent validity (AVE). Table 3. shows all values for reflective constructs (PE, EE, BI and AT) exceeded thresholds

(α >0,70, ρ_C >0,70, AVE > 0.50). Although high composite reliability values (e.g., PE, ρ_C = 0.965) indicate excellent internal consistency, they may also suggest potential item redundancy, warranting a review of the construct's operationalization in future studies.

Table 2. Statistics per item

Item	Average	Min	Max	SD	Kurtosis
PE1	4.18	1	5	0.90	-1.2468
PE2	3.97	1	5	0.95	-0.8893
PE3	4.10	1	5	0.93	-1.0139
EE1	4.10	1	5	1.00	-1.1535
EE2	4.28	2	5	0.78	-0.8953
EE3	3.85	1	5	1.14	-0.6459
SI1	4.41	3	5	0.74	-0.8561
SI2	3.31	1	5	1.14	-0.3200
SI3	3.36	1	5	1.07	-0.3929
SI4	4.00	1	5	0.91	-1.0768
FC1	4.49	2	5	0.09	-1.8130
FC2	4.70	3	5	0.57	-1.5694
FC3	3.95	1	5	1.13	-0.8907
BI1	4.51	3	5	0.71	-1.1580
BI2	4.33	2	5	0.86	-1.4864
BI3	2.90	1	5	1.08	-0.2940
AT1	4.08	2	5	0.86	-0.6593
AT2	4.08	2	5	0.83	-0.7129
AT3	3.90	2	5	0.93	-0.3879
AT4	3.90	1	5	0.96	-0.5205
WLO1	2.59	1	5	1.01	0.45205
WLO2	2.59	1	5	0.98	0.07852
PP1	3.90	2	5	0.90	-0.4479
PP2	3.87	2	5	0.79	-0.0837
PP3	3.92	2	5	0.83	-0.4131
PP4	3.54	1	5	1.03	-0.5416
PP5	3.44	1	5	1.11	-0.3654

Table 3. Reliability indicators and convergent validity for reflective constructs

Construct	Cronbach α	ρ_C	AVE	ρ_A
EE	0.837	0.902	0.754	0.843
PE	0.946	0.965	0.903	0.947
AT	0.923	0.945	0.812	0.935
BI	0.700	0.833	0.625	0.708

Outer loadings and VIF values for reflective constructs showed satisfactory convergent validity. However, VIF values for PE1 (6.305), PE3 (5.478), and AT3 (5.204) exceeded the threshold of 5, indicating potential multicollinearity within PE and AT constructs (Table 4.).

Table 4. Outer loadings & VIF for reflective constructs

	Outer Loadings	T Stat.	*p*-value	VIF
PE1	0.960	48.783	<0.001	6.305
PE2	0.937	36.520	<0.001	3.768
PE3	0.953	41.585	<0.001	5.478
EE1	0.865	24.155	<0.001	1.816
EE2	0.893	21.478	<0.001	2.272
EE3	0.846	15.023	<0.001	1.959
AT1	0.849	10.822	<0.001	2.519
AT2	0.926	28.157	<0.001	4.832
AT3	0.935	49.168	<0.001	5.204
AT4	0.891	25.318	<0.001	3.012
BI1	0.817	16.011	<0.001	1.420
BI2	0.824	9.720	<0.001	1.496
BI3	0.727	6.345	<0.001	1.266

For formative constructs, VIF values were acceptable (<5), confirming no significant multicollinearity. However, indicator contributions to formative constructs (Table 5.) varied: Only SI4 (weight = 0.694, p = 0.002) significantly contributed to SI; In constructs WLO, FC, and PP, no indicators were statistically significant; Some indicators (PP2, PP5) had negative weights, which is theoretically dubious and indicates a need for revision.

Table 5. Outer weights & VIF for formative constructs

Item	Outer Weights	T stat	*p*-value	VIF
SI1	0.261	0.924	0.356	1.461
SI2	0.131	0.462	0.644	1.770
SI3	0.129	0.400	0.689	2.591
SI4	0.694	3.122	0.002	2.006
WLO1	0.817	0.837	0.402	2.684
WLO2	0.220	0.223	0.823	2.684
FC1	0.396	0.973	0.331	1.209
FC2	0.444	1.026	0.305	1.259
FC3	0.594	1.430	0.153	1.047
PP1	0.495	1.367	0.172	3.761
PP2	-0.156	0.576	0.565	2.168
PP3	0.716	2.120	0.034	3.631
PP4	0.166	0.444	0.657	2.187
PP5	-0.289	1.090	0.276	1.988

Discriminant validity (Table 6.) using Heterotrait-Monotrait Ratio (HTMT) values showed: HTMT between EE and PE = 0.883 (acceptable); HTMT between AT and EE = 0.912 (above 0.90), suggesting a problem with discriminant validity between those constructs.

Table 6. HTMT for discriminant validity

Construct	EE	PE	AT	BI
EE				
PE	0.883			
AT	0.912	0.802		
BI	1.017	0.877	0.971	

4.3 Structural Model Evaluation

Structural model evaluation began with path coefficients and their significance. The **R^2 value for BI was 0.712**, meaning that exogenous constructs explained 71.2% of the variance in Behavioral Intention, i.e. a high level of explained variance. Graphical representation of the evaluated structural model with R^2 values is given on **Fig. 1.**

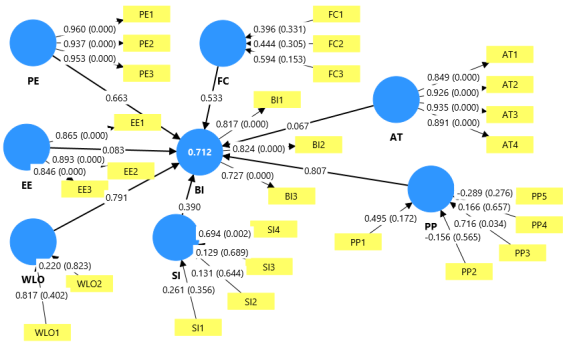


Figure 1. Evaluated structural model

However, none of the tested hypothesis, i.e. paths to BI were statistically significant ($p < 0.05$), as shown in Table 7. The path EE → BI came closest ($\beta = 0.316$, $p = 0.083$).

Table 7. Path coefficients, p-value, f^2

Relationship	Path coefficients (b)	T stat.	p-value	f^2
SI® BI	0.167	0.860	0.390	0.044
EE® BI	0.316	1.735	0.083	0.087
WLO® BI	0.024	0.266	0.791	0.002
FC® BI	0.087	0.623	0.533	0.020
PE® BI	0.087	0.436	0.663	0.008
PP® BI	-0.047	0.245	0.807	0.003
AT® BI	0.380	1.831	0.067	0.121

Effect sizes **(f^2)** were small. Despite the fact that the total explained variance ($R^2 = 0.712$) is high, individual effects of predictors measured with f^2 are small. The largest were: EE → BI (0.087); AT → BI (0.121) and SI → BI (0.044). Other paths had very small effects ($f^2 < 0.02$), suggesting that although the model explains BI collectively, no single predictor has a strong individual contribution.

Finally, multicollinearity among predictors in the inner model was checked (Table 8.). All VIF values were below 5 (maximum = 4.148 for AT), indicating no multicollinearity issues and supporting the reliability of path coefficient estimates.

Table 8. VIF for constructs in the inner model

Constructs	BI
SI	2.205
EE	3.986
WLO	1.192
FC	1.336
PE	3.521
PP	2.420
AT	4.148

5 Discussion

This study was designed as a pilot project with the primary aim of validating a newly developed measurement instrument within the extended UTAUT framework. As such, the findings should be viewed as preliminary and interpreted in the context of instrument testing rather than generalization.

5.1 Teachers' Attitudes, Needs and Experiences

Especially high average values and medians were recorded for all **Performance Expectancy** (PE) items, as well as certain Effort Expectancy (EE2), Social Influence (SI1, SI4), Facilitating Conditions (FC1, FC2), and Behavioral Intention (BI1, BI2**)** items. These results indicate highly positive teacher perceptions regarding the benefits, ease of use, and available support for using DT. This aligns with prior findings (Hermita et al., 2023; Wijaya et al., 2022), which showed PE as a significant positive predictor of teachers' intentions to use DT. While (Wijaya et al., 2022) found EE was not significant, (Masango, 2022) reported that participants considered digital resources "easy to use and navigate." (Hermita et al., 2023) found EE positively and significantly influenced teachers' intentions to use DT.

Low averages for **Workload** items (WLO1, WLO2) further reinforce the perceived ease of use, suggesting that teachers do not see DT as a significant source of stress or additional work.

High averages for **Social Influence (SI1, SI4)** reflect a strong perception that principals and students encourage teachers to use DT which is consistent with UTAUT, where SI is a key factor. SI was identified by (Wijaya et al., 2022) as the second most important factor influencing BI, and by (Hermita et al., 2023) as the most important.

High values for **Facilitating Conditions (FC1, FC2)** indicate that teachers believe their schools and technical infrastructure support DT use. UTAUT defines FC as the degree to which one believes that organizational and technical systems support

technology use. Other studies (Hermita et al., 2023; Masango, 2022; Wijaya et al., 2022) confirm FC's importance in enabling DT adoption. Namely, (Wijaya et al., 2022) found that Facilitating Conditions (FC) have a significant effect on the actual use of digital textbooks, highlighting the need for schools to provide all necessary support and training. (Masango, 2022) also reported adequate support from technical staff in schools. (Hermita et al., 2023) likewise found that FC positively and significantly influence teachers' intention to use digital textbooks.

High values for **Behavioral Intention (BI1, BI2)** suggest a strong intention among teachers to continue using DT and recommend them to others. This supports the UTAUT model, where BI is the strongest predictor of actual usage. (Wijaya et al., 2022) confirmed that BI had the most significant effect on actual DT use by teachers.

A lower average for item BI3 ("I think most of my teaching will be conducted using DT.") suggests teacher hesitation or disagreement with the idea that DT will be used as the most prominent source in teaching and learning. This implies that teachers see DT as **a complement rather than a replacement** for traditional methods which is consistent with (Moundy et al., 2022), who found that digital tools support teachers and students in teaching and learning but do not replace face-to-face teaching. On the other hand, (Grönlund et al., 2018) noted that teachers often use DT traditionally, without leveraging their interactive or collaborative features. Additionally, (Håkansson Lindqvist, 2019) observed that students frequently use printed materials in the classroom and digital ones outside it — pointing to a potential divide between social and educational use.

Lower-rated items were also found in the **Pedagogical Potential (PP4, PP5)** area. This may suggest some hesitancy or uncertainty about DT's full pedagogical value, despite their perceived usefulness and ease of use. (Grönlund et al., 2018) noted that teachers often view DT as static books, while (Håkansson Lindqvist, 2019) highlighted challenges teachers face in finding time to explore and implement DT effectively, which may limit their understanding of DT's pedagogical potential.

5.2 Validity and Reliability of Constructs

High **VIF values** were observed for indicators PE1 (6.305), PE3 (5.478), and AT3 (5.204), suggesting potential **multicollinearity** within those constructs and indicating a need to re-examine content redundancy in these items (see Table 4.).

For the **formative constructs**, although VIF values were within acceptable limits (Table 5.), **most indicators did not show statistically significant weights**. This was especially true for the constructs Workload (WLO), Facilitating Conditions (FC), and Pedagogical Potential (PP), where none of the indicators significantly contributed to construct formation. These findings indicate potential weaknesses in content validity, suggesting that some indicators may not meaningfully contribute to the conceptual domain of the construct. However, given the limited sample size and the exploratory nature of this pilot study, no changes to the instrument are planned at this stage. Instead, further validation with a larger sample will help confirm whether the observed issues (such as non-significant or negative indicator weights) persist, and whether any refinement is ultimately warranted. Furthermore, some indicators had **negative weights**, such as PP2 (-0.156), and PP5 (-0.289), which raises theoretical concerns and calls for thorough **review or reformulation** of these items before the main study.

In the case of WLO, the similarity in content between WLO1 and WLO2, despite differing weights, further suggests the need to revise these indicators to avoid redundancy and ensure that each item contributes clearly and consistently to the concept of teacher workload.

The low perceived workload in this study may partially explain the absence of a negative effect of WLO on Behavioral Intention, which is consistent with **TALIS 2018** results for Croatia (Dekanić et al., n.d.). This showed that workload is luckily not perceived as a barrier to professional development. This finding is important because it suggests that teachers **do not perceive DT as a major source of stress or additional work**, which is a key positive aspect in their integration into teaching and learning.

Given that Croatian teachers often report a **high workload**, especially in administrative tasks and lesson planning, our preliminary results indicate that DT **do not add to this burden**. On the contrary, low levels of perceived workload may be a **key factor** in successful acceptance and broader implementation of DT which is contrary to common expectations that new technologies inherently increase teacher workload. Therefore, these aspects should be explored further in the main study to more precisely identify elements that reduce or increase workload.

5.3 Discriminant Validity

Discriminant validity assessment using the Heterotrait-Monotrait Ratio (HTMT) showed an overlap between **Attitude (AT)** and **Effort Expectancy (EE)**, with a value of 0.912 that is exceeding the recommended threshold of 0.90 (Table 6.).

This suggests that participants **did not clearly distinguish** between these two constructs. While this is statistically problematic for discriminant validity, the overlap is not theoretically unexpected: both constructs come from established technology acceptance models and are often closely linked and mutually influential (Hermita et al., 2023; Venkatesh et al., 2003). Nevertheless, this calls for a **revision of indicators** to avoid redundancy and ensure better discriminant validity in the main study. In the main study, the item

sets for Attitude and Effort Expectancy should be reviewed and potentially reworded or reorganized to reduce conceptual overlap and improve discriminant validity. Similarly, the HTMT value between PE and EE was 0.883 — slightly under the 0.90 threshold but still indicating a strong correlation, which is theoretically expected due to their conceptual proximity.

The HTMT value between BI and EE (1.017) also exceeds the recommended threshold, indicating substantial conceptual overlap that may reflect item redundancy or insufficient differentiation in item wording. This further supports the need to review both constructs' operationalization in future studies.

5.4 Structural Model Results

The structural model analysis revealed that although the total explained variance (R^2) for BI is high at 0.717, **none of the tested hypothesis, i.e. paths** to BI were statistically significant (see Table 7.). The closest was EE → BI (p = 0.098), aligning with previous studies (Kahnbach et al., 2024; Wijaya et al., 2022). Other paths were far from significance thresholds, likely due to the **small sample size (N = 39)** in this pilot study, which limits the statistical power of the model and its ability to detect significant relationships (Hair et al., 2022).

The f^2 effect size values further confirmed that **no individual predictor** had a strong influence on BI. This opens room to explore the potential role of **moderating variables** in the main study. VIF values (Table 8.) confirmed the **absence of multicollinearity** among latent variables in the inner model, supporting the reliability of the path coefficients.

6 Conclusion

This pilot study investigated the factors influencing the acceptance of DT among primary school subject teachers in Croatia, focusing on attitudes, experiences, and behavioral intentions. Based on the UTAUT framework, extended with pedagogical, attitudinal, and workload-related constructs, the study aimed to validate a measurement instrument for broader research.

The findings revealed generally positive attitudes among teachers toward the use of DT. Teachers recognize the value, ease of use, and support associated with DT implementation. Further, low workload perceptions indicate that DT are not viewed as a source of stress or additional burden. However, the study also highlighted the need for strategies that foster deeper understanding and full use of DT's pedagogical potential, and their integration beyond traditional methods.

Despite these encouraging perceptions, none of the hypothesized relationships with behavioral intention were statistically significant, which is likely due to the limited sample size in this pilot phase. Nonetheless, the overall explained variance for behavioral intention was high, pointing to the collective strength of the model and its potential in the broader study.

Despite a small sample and limited theoretical clarity in some constructs, the study provides valuable input for instrument refinement. The main study will include a larger sample, revised items, and mediation testing to strengthen model validity. The lack of a priori power analysis limits interpretation of non-significant paths in this relatively complex model.

In conclusion, the pilot study confirmed strong readiness among teachers to adopt DT as well as the relevance and applicability of the extended UTAUT model in this context, and the need for instrument revision.

References

Briz-Ponce, L., Pereira, A., Carvalho, L., Juanes-Méndez, J. A., & García-Peñalvo, F. J. (2017). Learning with mobile technologies – Students' behavior. *Computers in Human Behavior*, *72*, 612–620. https://doi.org/10.1016/j.chb.2016.05.027

Cruchinho, P., López-Franco, M. D., Capelas, M. L., Almeida, S., Bennett, P. M., Miranda da Silva, M., Teixeira, G., Nunes, E., Lucas, P., & Gaspar, F. (2024). Translation, Cross-Cultural Adaptation, and Validation of Measurement Instruments: A Practical Guideline for Novice Researchers. *Journal of Multidisciplinary Healthcare*, *Volume 17*, 2701–2728. https://doi.org/10.2147/jmdh.s419714

Cukurova, M., Miao, X., & Brooker, R. (2023). Adoption of Artificial Intelligence in Schools: Unveiling Factors Influencing Teachers' Engagement. In *Artificial Intelligence in Education* (Vol. 13916, pp. 151–163). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9_13

Dekanić, A. M., Gregurović, M., & Batur, M. (n.d.). *TALIS 2018: Učitelji, nastavnici i ravnatelji – cijenjeni stručnjaci*.

Divjak, B. (2022). *Balanced Learning Design Planning: Concept and Tool*. *46*(2).

E-škole. (n.d.). *E-Škole*. Retrieved 16 May 2022, from https://pilot.e-skole.hr/hr/e-skole/opis-projekta/

Grönlund, Å., Wiklund, M., & Böö, R. (2018). No name, no game: Challenges to use of collaborative digital textbooks. *Education and Information Technologies*, *23*(3), 1359–1375. https://doi.org/10.1007/s10639-017-9669-z

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Sage Publications.

Håkansson Lindqvist, M. (2019). Talking about digital textbooks. The teacher perspective. *The*

International Journal of Information and Learning Technology, *36*(3), 254–265. https://doi.org/10.1108/IJILT-11-2018-0132

Hermita, N., Wijaya, T. T., Yusron, E., Abidin, Y., Alim, J. A., & Putra, Z. H. (2023). Extending unified theory of acceptance and use of technology to understand the acceptance of digital textbook for elementary School in Indonesia. *Frontiers in Education*, *8*, 958800. https://doi.org/10.3389/feduc.2023.958800

Joo, Y. J., Park, S., & Shin, E. K. (2017). Students' expectation, satisfaction, and continuance intention to use digital textbooks. *Computers in Human Behavior*, *69*, 83–90. https://doi.org/10.1016/j.chb.2016.12.025

Kabugo, D. (2020). Utilizing Open Education Resources to Enhance Students' Learning Outcomes during the COVID-19 Schools Lockdown: A Case of Using Kolibri by Selected Government Schools in Uganda. *Journal of Learning for Development (JL4D)*, *7*(3), 447–458. https://doi.org/10.56059/jl4d.v7i3.465

Kahnbach, L., Hase, A., Kuhl, P., & Lehr, D. (2024). Explaining primary school teachers' intention to use digital learning platforms for students' individualized practice: Comparison of the standard UTAUT and an extended model. *Frontiers in Education*, *9*, 1353020. https://doi.org/10.3389/feduc.2024.1353020

Liebenberg, J., Benade, T., & Ellis, S. (2018). *Acceptance of ICT: Applicability of the Unified Theory of Acceptance and Use of Technology (UTAUT) to South African Students*. *10*(3). https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1499&context=ajis

Masango, M. M. (2022). *A Paperless Classroom: Importance of Training and Support in the Implementation of Electronic Textbooks in Gauteng Public Schools*. *20*(3), 15.

Materijali. (n.d.). Retrieved 13 May 2025, from https://edutorij.carnet.hr/materijali

Milić, M., & Divjak, B. (2023). *Innovative learning design and teachers' competencies: Prerequisites for meaningful use of digital material in schools*. 189–197. https://bit.ly/innovative_learning_design

Moundy, K., Chafiq, N., & Talbi, M. (2022). Digital Textbook and Flipped Classroom: Experimentation of the Self-Learning Method Based on the Development of Soft Skills and Disciplinary Knowledge. *International Journal of Emerging Technologies in Learning (IJET)*, *17*(07), 240–259. https://doi.org/10.3991/ijet.v17i07.28933

Shukla, S. (2021). M-learning adoption of management students': A case of India. *Education and Information Technologies*, *26*(1), 279–310. https://doi.org/10.1007/s10639-020-10271-8

Škola za život—Program 'Škola za život' je jedan od prvih koraka. (n.d.). Retrieved 15 June 2025, from https://skolazazivot.hr/

Sun, Z., & Jiang, Y. (2015). How the young generation uses digital textbooks via mobile learning terminals: Measurement of elementary school students in China: How the young generation uses digital textbooks. *British Journal of Educational Technology*, *46*(5), 961–964. https://doi.org/10.1111/bjet.12299

Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540

Viberg, O., Mavroudi, A., Khalil, M., & Bälter, O. (2020). Validating an Instrument to Measure Teachers'Preparedness to Use Digital Technology in their Teaching. *Nordic Journal of Digital Literacy*, *15*(1), 38–54. https://doi.org/10.18261/issn.1891-943x-2020-01-04

Wijaya, T. T., Zhou, Y., Houghton, T., Weinhandl, R., Lavicza, Z., & Yusop, F. D. (2022). Factors Affecting the Use of Digital Mathematics Textbooks in Indonesia. *Mathematics*, *10*(11), 1808. https://doi.org/10.3390/math10111808

Appendix A

Latent Construct	Item	Measurement	Reference
Performance expectancy	PE1	Digital textbooks improve my teaching quality.	(Hermita et al., 2023; Shukla, 2021)
	PE2	Teaching using digital textbooks increases my productivity.	
	PE3	Digital textbooks help me when teaching elementary materials.	
Effort expectancy	EE1	It is easy for me to teach using a digital textbook	(Hermita et al., 2023)
	EE2	I find the digital textbook easy to use.	
	EE3	Using a digital textbook for teaching does not require much preparation	
Social influences	SI1	In general, the school supports me using digital textbooks.	(Hermita et al., 2023)
	SI2	Colleagues affecting my work think I should use digital technology.	
	SI3	Parents support the use of digital textbook in teaching.	(Kahnbach et al., 2024)
	SI4	Students support the use of digital textbook in teaching.	(Wijaya et al., 2022)
Facilitating conditions Infrastructure Knowledge Support	FC1	I have all the necessary resources (technology, Internet access, licenses) to be able to use digital learning digital learning platforms for practice.	(Kahnbach et al., 2024)
	FC2	I have the knowledge necessary to use digital learning platforms for practice.	
	FC3	I have a contact person available to help me with difficulties in using digital learning platforms for practice	
Behavior intention	BI1	I will continue to use digital textbooks to teach.	(Hermita et al., 2023)
	BI2	I will use the digital textbook whenever the situation allows.	
	BI3	I think most of my teaching will be conducted using digital textbooks	
Attitude	AT1	Using digital textbooks to teach in elementary schools is a good idea.	(Hermita et al., 2023; Liebenberg et al., 2018)
	AT2	Digital textbooks make lessons more interesting.	
	AT3	I like to teach using a digital textbook.	
	AT4	Using digital textbooks makes teaching and learning activities fun	(Hermita et al., 2023; Briz-Ponce et al., 2017)
Workload	WLO1	Since I started using digital textbooks, I have significantly more workload in Lesson planning and preparation (i.e. setting instructional goals, understanding student levels, previewing lessons etc.)	(Cukurova et al., 2023)
	WLO2	Since I started using digital textbooks, I have significantly more workload in Classroom management (i.e. balancing self-regulated learning and collaboration, behavioural management, orchestration of learning tasks etc.)	
Pedagogical potential	PP1	Digital textbooks provide opportunities to teach in new ways.	(Hermita et al., 2023)
	PP2	Digital textbooks provide opportunities to interact with students.	(Hermita et al., 2023)
	PP3	Digital textbooks provide opportunities for creative thinking	(Hermita et al., 2023)
	PP4	Digital textbooks allows for the adaptation of content and activities to the individual needs and abilities of students.	(Wijaya et al., 2022)
	PP5	Digital textbooks allows for the monitoring of student progress.	(Wijaya et al., 2022)

The Impact of Different Game Modes in an Educational Web Application on Student Engagement and Performance

Mia Šagovac, Dorotea Potoč, Tomislav Jagušt
University of Zagreb
Faculty of Electrical engineering and Computing
Unska 3, 10000 Zagreb
{mia.sagovac, dorotea.potoc, tomislav.jagust}@fer.unizg.hr

Abstract. *This Work-in-Progress paper presents the design and preliminary classroom evaluation of Edu Bingo, a gamified web application developed to enhance motivation and engagement among lower primary school students. The application features three distinct game modes, Endless (practice-oriented), Reward (badge and XP-based), and Multiplayer (competitive with leaderboard), each tailored to different motivational profiles and learner preferences. Core game elements such as points, badges, and avatars are integrated to promote active learning, while minimizing dependence on extrinsic feedback alone. The initial study involved 24 second-grade students (ages 7–8) using school-provided desktop computers. After engaging with all three modes, students completed a post-session questionnaire, and gameplay data (e.g., task accuracy, XP earned, and mode selection) was collected and analysed. Preliminary findings indicate that offering multiple gamified approaches can improve student engagement and accommodate diverse learning styles. These early results support the potential of personalized gamification as a valuable tool in early primary education.*

Keywords. educational technology, gamification, personalization, game-based learning, engagement

1 Introduction

Gamification has become a widely discussed and increasingly adopted strategy for enhancing student motivation and engagement in educational contexts. Broadly defined as the use of game-like elements in non-game settings (Deterding et al., 2011), gamification has been successfully applied across educational levels to make learning more interactive and rewarding. Numerous studies have demonstrated its positive effects on student motivation, participation, and long-term knowledge retention (Gee, 2003). Commonly used gamified elements include points, badges, leaderboards, avatars, and levels, all designed to transform routine learning tasks into meaningful and enjoyable challenges (Werbach and Hunter, 2012).

In technology-enhanced learning environments, gamification is often embedded within existing e-learning platforms or developed as standalone educational games. For instance, Duolingo incorporates experience points and daily streaks to support consistent language learning (Falk and Götz, 2016), while Prodigy integrates mathematical problem-solving into a fantasy game narrative to sustain learner interest through progression mechanics (Himmawan and Juandi, 2023). New pedagogical frameworks emphasize that gamification should not merely be added onto content but should align with deeper learning principles and student-centric design (Fathi Najafi et al., 2025).

Although most research highlights the benefits of gamification, a growing body of literature points to its limitations. Not all students respond positively to the same game elements; for example, competitive students may be motivated by leaderboards, whereas others may find them demotivating or anxiety-inducing (Buckley and Doyle, 2016). More recent studies have reinforced these concerns, in a systematic mapping of 87 educational software studies, found that elements such as points, badges, and leaderboards can lead to increased anxiety, decreased intrinsic motivation, and even a decline in performance when not properly aligned with learner profiles (Almeida et al., 2023). Furthermore, the novelty effect of gamification tends to wear off over time, especially if the system lacks adaptability or meaningful long-term progression (Hanus and Fox, 2015). In a longitudinal study involving 756 students in a STEM learning environment, observed an initial peak in engagement followed by a sharp drop after approximately four weeks, with only a partial recovery over time (Rodrigues et al., 2022). These findings underscore the importance of designing gamified learning tools that are not only engaging at the outset but also sustainable and responsive to diverse learner needs over time. These issues suggest that "one-size-fits-all" gamified designs may fail to meet the diverse needs of learners in a classroom setting.

To address these challenges, recent research advocates for more personalised and multimodal approaches to gamification, allowing learners to choose

the mode of interaction that best fits their preferences and learning styles (Oliveira et al., 2022). Such designs may offer more inclusive and effective learning environments by better aligning with individual differences in personality, motivation, and cognitive engagement.

In parallel, recent developments in learning analytics highlight the importance of monitoring student progress through visual dashboards, which can provide teachers with actionable insights to support personalised instruction and adaptive intervention strategies (Balaban et al., 2024).

A recent study further reinforces these insights by demonstrating that personalised gamification, tailored to learners' profiles, significantly outperforms uniform approaches in terms of motivation, engagement, and learning outcomes, particularly in online learning environments (Xiao and Hew, 2024)

This paper presents findings from a small-scale exploratory study conducted in a second-grade classroom using a custom-developed gamified web application, Edu Bingo. The application includes three distinct game modes: Endless (focused on practice and repetition), Reward (featuring badges and experience points), and Multiplayer (emphasizing real-time competition via a leaderboard), each designed to reflect a different motivational approach. Students used the application during a classroom session, after which they completed a short questionnaire. Simultaneously, gameplay data (e.g., accuracy, playtime, and mode selection) were logged for further analysis.

The study seeks to contribute to the growing field of personalised gamification by exploring how different game mechanics influence engagement, motivation, and learning outcomes in early primary education. This study is guided by the following research questions:

- **RQ1:** Which of the three game modes (Endless, Reward, or Multiplayer) is most preferred by second-grade students in a gamified learning environment?
- **RQ2:** How do different game modes affect students' motivation and engagement during gameplay?
- **RQ3:** How does task accuracy differ across the Endless, Reward, and Multiplayer game modes?

2 Related work

Gamification has been widely recognised as an effective strategy for enhancing learning outcomes. It is defined as "a goal-oriented approach that integrates game design elements into learning environments to increase motivation and engagement" (Sailer and Homner, 2020). Meta-analyses highlight that elements such as points, feedback, badges, leaderboards, and narrative are particularly effective in supporting both cognitive and motivational benefits.

Systematic reviews further confirm the positive impact of gamification on student motivation, participation, and academic performance, especially when grounded in clear pedagogical objectives and supported by purposeful and varied game mechanics (Zainuddin et al., 2020).

However, several studies caution that the mere presence of game elements does not necessarily lead to improved learning. Research has shown that individual components, such as points or levels, have little effect on intrinsic motivation unless they are meaningfully embedded within the learning process (Mekler et al., 2017). Similarly, leaderboards have been found to improve performance only when learners understand how their actions influence rankings and progress (Landers et al., 2017). These findings underscore the importance of aligning game mechanics with learning goals and ensuring transparency in their function.

A distinction has also been made between surface-level gamification, where elements like points and badges are simply layered over content, and deeper, gameful designs where mechanics are integrated directly into instructional structures (Seaborn and Fels, 2015). Such findings highlight the need for pedagogically sound and thoughtfully designed gamification strategies.

Traditional gamified systems are often based on the assumption that all learners respond similarly to the same features. Yet longitudinal studies reveal varied engagement patterns, while competitive elements may motivate some students, they can be demotivating for others (Barata et al., 2017). Related research further argues that although extrinsic motivators such as points and badges may boost short-term participation, long-term engagement requires meaningful tasks and user-centred design (Hamari et al., 2014).

Gamification appears particularly effective in procedural and skill-based domains. Integrating game elements into activities such as problem-solving, grammar, and spelling has been shown to enhance both performance and learner satisfaction (Domínguez et al., 2013). These insights align with the focus of this study, which centres on mathematics and language-based tasks. Additionally, gamified tools that align with curriculum content can foster comprehension, creativity, and active engagement among primary school students (Al Ghozali et al., 2024).

Beyond task type, learner personality also plays a key role in shaping engagement with gamified environments. It has been demonstrated that competitive learners benefit from public ranking and recognition, whereas others perform better in non-competitive settings where progress is privately tracked (Buckley and Doyle, 2017). Further studies indicate that extroverted students are more responsive to public-facing features like leaderboards, while introverted and ambiverted learners prefer personalised goals, badges, and private feedback mechanisms. Many in the latter group report

discomfort with publicly visible tracking and favour systems that support internally visible progress (Tlili et al., 2019).

3 Methodology

The classroom study was conducted with a group of 24 second-grade students (ages 7 to 8) from a primary school in Zagreb. The group consisted of 10 girls and 14 boys, all of whom had previous experience using digital devices in class.

3.1 Application

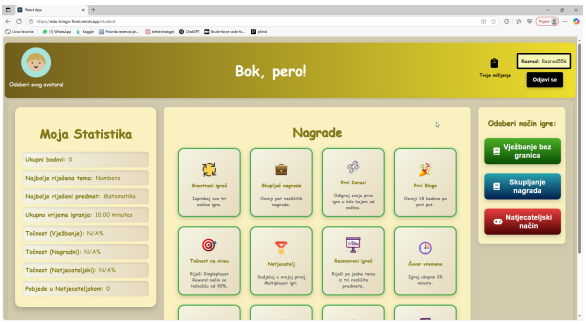


Figure 1. Student dashboard

The web application was developed to support second-grade students in practising educational content through an interactive and gamified e xperience. The application features three distinct game modes, Singleplayer Endless, Singleplayer Reward, and Multiplayer Competitive, which differ in motivational mechanics but share a consistent interface and game logic. Each mode is accessed through the student's profile page which is shown in Fig. 1, where students can also track their statistics and earned badges and choose their avatar. All modes are based on a bingo-style gameboard consisting of a 3×3 grid of numbers, behind which tasks are hidden. Each gameboard is automatically generated with three easy, medium, and complex questions, balancing content difficulty. Easy questions carry 1 point, medium 2 points, and hard questions 3 points. A random number between 1 and 90 is generated upon clicking a button, and students uncover the corresponding task if that number appears on their board. Questions are subject- and theme-specific (e.g., grammar, addition, multiplication, etc.) and must be answered correctly to mark the tile as complete. If the student answers incorrectly as shown in Fig. 2, they automatically see that and the task is replaced with a new one of the same difficulty l evel, a nd t he s tudent may try again. In the Endless Mode, students play continuously with no scores, rewards, or pressure, this mode is intended for repetition and practice. The Reward Mode introduces gamification b y a warding XP points and badges for performance; these elements boost motivation and promote goal setting.



Figure 2. Student view of the Multiplayer mode interface

Earned badges (e.g., "First Bingo") and XP totals are visible on the student dashboard and the gameboard during gameplay immediately when earned.

Multiplayer Mode offers a competitive experience in real time. Students join the same game via a code. The game ends when one player achieves bingo, and all participants are shown a final l eaderboard. A live leaderboard tracks player progress during the game, encouraging quick thinking and friendly competition. Student accuracy and gameplay time are logged for every session. The app also includes a Teacher Dashboard as shown in Fig. 3, allowing educators to view student performance by game mode, monitor progress, and initiate multiplayer games. However, teachers cannot modify or create tasks; the application uses a predefined d atabase o f c urriculum-aligned q uestions prepared in advance. This ensures that all students receive standardised content across game modes. The app aims to promote active learning and analyse how different game mechanics influence e ngagement and educational outcomes.

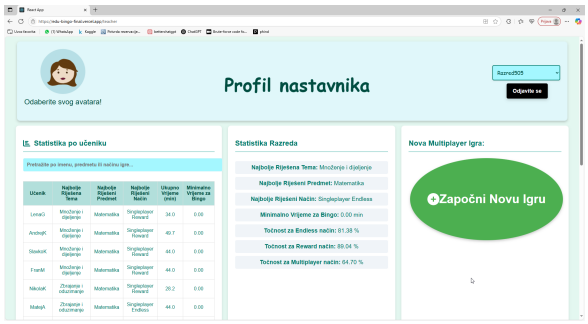


Figure 3. Teacher dashboard

3.2 Experiment Design

After a brief introduction to the application and its purpose, students were given access to the Edu Bingo web application using school-provided desktop computers. All students played through each of the three game modes (Endless, Reward, and Multiplayer) in that order, spending approximately 10 minutes on each mode. Instructions were intentionally kept minimal to encourage intuitive and independent interaction with the game. After completing all three modes, students were allowed to choose their preferred mode. The majority opted for the Multiplayer mode, and the class proceeded to play several additional rounds. This unstructured session was considered and discussed in the interpretation of the results.

3.3 Questionnaire

After the experiment, a short questionnaire with three groups of questions was completed by the students via Google Forms. The first group consisted of multiple-choice questions about gameplay preferences (e.g., most fun, most difficult, most useful mode). The second group included 5-point Likert scale questions (1 – strongly disagree, 5 – strongly agree) related to overall enjoyment and specific elements like competition. The final group contained open-ended questions, allowing students to explain which topic or game mode they found most interesting, and whether any part of the application was confusing and why.

3.4 Context and Learning Outcomes

The Edu Bingo activities were based on the official second-grade curriculum and designed to support learning in several core subject areas, namely:

- **Mathematics** (addition, subtraction, multiplication, division),
- **Croatian Language** (capital letters and sentence structure),
- **Environmental Studies** (e.g., sorting different types of waste).

The intended learning outcomes (LOs) of the Edu Bingo activities included:

- Correctly solving arithmetic problems of varying complexity,
- Recognizing and applying capitalization rules in written language,
- Demonstrating understanding of basic environmental responsibility through correct waste sorting.

These topics were selected in consultation with the classroom teacher. Mathematics topics had been actively practiced in the weeks prior to the experiment, while grammar and environmental topics had been covered earlier or less extensively. This variation in prior knowledge was considered when interpreting students' performance across different themes (see Section 4.3).

4 Results

Students showed high levels of enthusiasm and curiosity when introduced to the concept of gamified learning.

To explore students' preferences, performance and engagement across different game modes, data were collected from both application logs and a post-experiment questionnaire. Gameplay data included the number of tasks completed and task accuracy per game mode and theme, while the survey captured students' subjective experiences and opinions.

4.1 Game Mode Preferences (RQ1)

Prior research has shown that students differ significantly in how they respond to gamification elements, highlighting the importance of offering diverse interaction styles to match varied motivational profiles (Santos et al., 2021).

Student responses revealed a range of preferences regarding the available game modes. As presented in Table 1, Multiplayer was selected as the most liked mode by 48% of students, followed by Endless mode (36%) and Reward mode (16%). These results suggest that students were generally motivated by social interaction and competition, but a significant number also valued a more relaxed, individual approach to practice.

Table 1. Student preferences and perceptions of game modes

Question	Endless	Reward	Multiplayer
Which game mode did you like the most?	36%	16%	48%
Which game mode was the easiest to understand?	56%	20%	24%
Which game mode helped you best remember what you learned at school?	36%	8%	56%

Beyond enjoyment, students were also asked to indicate which game mode they found easiest to understand. Over half of the class (56%) selected Endless mode, indicating that the absence of points, badges, or time pressure may have helped them focus on the mechanics of the game itself. Multiplayer and Reward modes were considered less intuitive, with 24% and 20% of students respectively identifying them as easiest to follow. This finding may reflect the additional

complexity introduced by competitive or reward-based elements.

When asked which mode best supported their learning and helped them remember the content, Multiplayer again stood out with 56% of responses. Endless mode was the second most helpful in this regard (36%), while only a small percentage (8%) found Reward mode to be most beneficial for knowledge retention. This may be explained by the fact that the excitement and active participation involved in real-time competition helped reinforce memory for many students.

Such variability in preferences aligns with prior research showing that different learner types respond to different gamification elements. For instance, study of 271 students by Pinto et al., 2022, who found that more than one-third identified as "Achievers", who prefer progression mechanics like points and badges, while a smaller share identified as "Socialisers", favouring communication, driven element.

These findings are consistent with recent work by I. Santos et al., 2024, who found that elements such as quests, points, and badges were among the most appreciated by students, whereas competitive elements and time pressure were less well received.

This finding aligns with recent research by A. C. G. Santos et al., 2021, who found that learners demonstrate highly diverse user profiles, such as Achievers, Philanthropists, or Free Spirits, each associated with distinct motivational patterns and preferences for gamification features. Their study with over 300 participants showed that no single design suits all learners equally, underscoring the need for flexible and adaptable gamification systems.

Our own results echo these conclusions. While nearly half of the students preferred Multiplayer, a considerable portion favored other modes. This diversity suggests that a one-size-fits-all approach to gamification may be insufficient, and that effective learning tools should allow flexibility and personalization to support different motivational profiles.

4.2 Motivation and Engagement (RQ2)

While preferences reflect functional and cognitive evaluations, students' motivational responses provided further insight into how they experienced the gamified environment. Survey results indicate that students found EduBingo both enjoyable and engaging across game modes.

As shown in Table 2, 84% of students reported that they would like to play EduBingo again, while 64% felt they learned something new, and the same percentage agreed that learning was easier when they were competing with others.

In addition, Likert-scale responses presented in Table 3 show that 64% strongly agreed that EduBingo was more interesting than regular classroom learning, and 56% strongly agreed that they enjoyed the competitive Multiplayer mode. These results suggest that the inclusion of playful and interactive mechanics enhanced students' perception of the learning experience.

Table 2. Student responses to selected survey questions

Question	**Yes (%)**
Do you think learning was easier when competing with others?	64%
Did you learn something new while playing EduBingo?	64%
Would you like to play EduBingo again?	84%

Observationally, many students voluntarily chose to replay the Multiplayer mode during the unstructured portion of the lesson, indicating sustained engagement beyond the scope of the formal experiment. While Multiplayer appears to foster emotional excitement and social involvement, the Reward mode's link to higher accuracy suggests that progress-tracking features may help students stay focused and goal-oriented.

Overall, the findings support prior research indicating that students are not uniformly motivated. While some thrive on social comparison and competition, others are more engaged by achievement markers or individual progress tracking. This diversity reinforces the need to design gamified learning tools that include varied motivational mechanisms.

4.3 Performance and Accuracy (RQ3)

Accuracy was measured as the proportion of correctly solved tasks out of the total attempted tasks, calculated separately for each game mode and academic topic. Log data analysis revealed differences in performance across modes. As shown in Fig. 4, Reward mode resulted in the highest overall accuracy with 54% of students scoring between 91-100%, followed closely by Endless mode, where 50% of students achieved the same level of accuracy. In contrast, Multiplayer mode showed the lowest high-accuracy rate, with fewer than 20% of students scoring in 81-90%. This difference likely stems from time pressure and the competitive nature of Multiplayer play, which may have encouraged students to prioritize speed over correctness. Reward mode, which included XP points and badges, likely supported more careful and goal-oriented gameplay.

Beyond game modes, task performance also varied by academic theme. As illustrated in Fig. 5, students achieved the highest overall accuracy in Addition and Subtraction (54%), followed by Multiplication and Division (38%). Other themes such as Waste Sorting and Capital Letters accounted for only 4% each of total correct answers.

In this context, "overall accuracy" per topic refers

Table 3. Student responses to Likert-scale items (1 = Strongly Disagree, 5 = Strongly Agree)

Question	1	2	3	4	5
Was EduBingo more interesting than regular classroom learning?	8%	4%	8%	16%	64%
Did you enjoy competing with other students in the Competitive mode?	12%	8%	8%	16%	56%

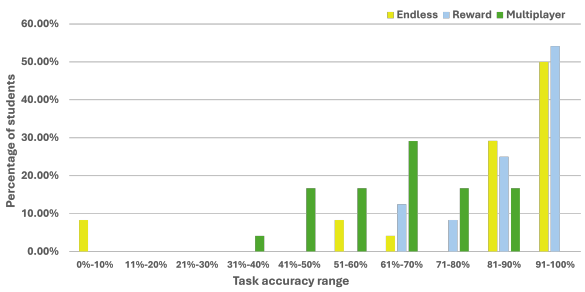


Figure 4. Accuracy per Mode

to the total number of correct answers across all game modes divided by the total number of tasks attempted in that topic. Topics with a larger number of included tasks (such as addition and subtraction) yield more reliable indicators of performance due to increased sample size. In contrast, topics with fewer tasks, such as language and environmental themes, should be interpreted more cautiously, as there were not enough data points to ensure reliable conclusions.

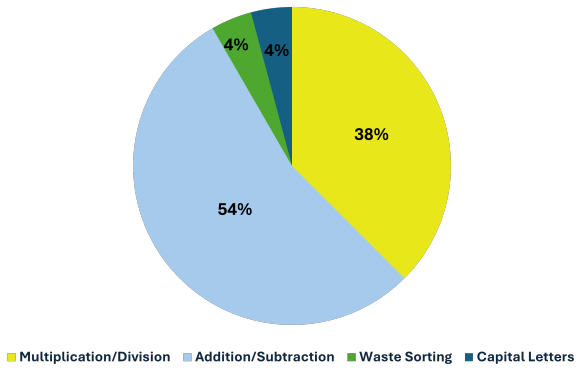


Figure 5. Best performing topic

These findings reflect both the students' curriculum familiarity and the perceived difficulty of each theme. At the time of the study, mathematics topics, particularly addition, subtraction, multiplication, and division, had been actively covered in class. This curricular alignment likely contributed to higher performance in math-based tasks. Moreover, anecdotal feedback from the classroom teacher confirmed that these topics were fresh in students' minds and well-practiced, which may have made solving them feel both accessible and enjoyable.

In contrast, tasks involving capitalisation rules (e.g., recognising proper nouns or sentence starts) were generally more challenging. The teacher noted that many students regularly struggle with this grammar topic, which is consistent with the low percentage of correct answers in that category. Additionally, these language-based tasks tended to be slightly more complex in structure, which may have compounded the difficulty for early readers.

Although the app included a range of topics, these four were among the most frequently encountered and statistically relevant during the session. The results underline the importance of content familiarity and perceived difficulty when evaluating performance in gamified learning environments.

While 4 offers useful insight into students' accuracy per game mode, it cannot fully represent the extent to which the targeted learning outcomes (LOs) were achieved.

5 Conclusion and Future Work

This paper presented the design and classroom testing of Edu Bingo, a gamified web-based learning tool for second-grade students. The application features three game modes—Endless, Reward, and Multiplayer—designed to support different learning styles through varied mechanics such as points, badges, and leaderboards.

Classroom results showed that gamification enhanced student interest and participation. Reward mode produced the highest accuracy, while Multiplayer was rated most engaging. Survey responses confirmed students' preference for gamified learning over traditional instruction, with many reporting that they learned something new.

Findings support the value of offering multiple game modes to address diverse motivational profiles. While some students responded best to competition, others preferred structured progression or relaxed practice, highlighting the need for adaptable gamification strategies in primary education.

The study is limited by its small sample size (N=24) and short duration. The custom-designed questionnaire also requires further validation.

Future work will involve broader testing and the development of adaptive features that personalise the learning experience. Planned improvements include integrating a built-in chat system to support collaboration, peer communication, and group tasks. Such functionality could enhance engagement and provide valuable insights into student thinking and social interaction.

Additionally, enabling teachers to create and cus-

tomise their own tasks, rather than relying solely on a fixed question base, would allow for greater curricular alignment and pedagogical flexibility. Additional features, such as team-based play and teacher-facing analytics, could further support cooperative learning and formative assessment.

This is a noted limitation of the current work. Future research will incorporate formal pre- and post-testing to evaluate learning progress in a more rigorous manner and better align observed in-game performance with curriculum-based outcomes.

References

Al Ghozali, M. I., Mustoip, S., & Sulkhah, S. (2024). The role of gamification methods in increasing creativity and innovation of elementary school students in the digital era. *JPS: Journal of Primary School*, *1*(2), 21–27.

Almeida, C., Kalinowski, M., Uchôa, A., & Feijó, B. (2023). Negative effects of gamification in education software: Systematic mapping and practitioner perceptions. *Information and Software Technology*, *156*, 107142.

Balaban, I., Zlatović, M., & Matus, M. (2024). Using learning analytics dashboards to monitor student progress: The case of a blended computer science university course. *2024 IEEE International Conference on Advanced Learning Technologies (ICALT)*, 72–74. https://doi.org/10.1109/ICALT61570.2024.00027

Barata, G., Gama, S., Jorge, J., & Gonçalves, D. (2017). Studying student differentiation in gamified education: A long-term study. *Computers in Human Behavior*, *71*, 550–585. https://doi.org/https://doi.org/10.1016/j.chb.2016.08.049

Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments*, *24*(6), 1162–1175. https://doi.org/10.1080/10494820.2014.964263

Buckley, P., & Doyle, E. (2017). Individualising gamification: An investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market. *Computers & Education*, *106*, 43–55. https://doi.org/https://doi.org/10.1016/j.compedu.2016.11.009

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*, *11*, 9–15. https://doi.org/10.1145/2181037.2181040

Domínguez, A., Saenz-de-Navarrete, J., de-Marcos, L., Fernandez-Sanz, L., Pagés, C., & Martínez, J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education*, *63*, 380–392. https://doi.org/10.1016/j.compedu.2012.12.020

Falk, S., & Götz, S. (2016, January). Interactivity in language learning applications: A case study based on duolingo.

Fathi Najafi, T., Andaroon, N., Bolghanabadi, N., et al. (2025). Gamification in midwifery education: A systematic review. *BMC Medical Education*, *25*, 297. https://doi.org/10.1186/s12909-025-06880-x

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Comput. Entertain.*, *1*(1), 20. https://doi.org/10.1145/950566.950595

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? – a literature review of empirical studies on gamification. *2014 47th Hawaii International Conference on System Sciences*, 3025–3034. https://doi.org/10.1109/HICSS.2014.377

Hanus, M., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, *80*. https://doi.org/10.1016/j.compedu.2014.08.019

Himmawan, D., & Juandi, D. (2023). Games based learning in mathematics education: A systematic literature review. *Union: Jurnal Ilmiah Pendidikan Matematika*, *11*, 41–50. https://doi.org/10.30738/union.v11i1.13982

Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior*, *71*, 508–515. https://doi.org/https://doi.org/10.1016/j.chb.2015.08.008

Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, *71*, 525–534. https://doi.org/https://doi.org/10.1016/j.chb.2015.08.048

Oliveira, W., Hamari, J., Joaquim, S., Toda, A., Toledo Palomino, P., Vassileva, J., & Isotani, S. (2022). The effects of personalized gamification on students' flow experience, motivation, and enjoyment. *Smart Learning Environments*, *9*. https://doi.org/10.1186/s40561-022-00194-x

Pinto, F. M., Jaftha, N., Borg, S., Micallef, M. Z., & Chircop, T. (2022). Students' learning and gaming preferences and their expectations of gamification. *MCAST Journal of Applied Research & Practice*, *6*(1), 60–78.

Rodrigues, L., Pereira, F. D., Toda, A. M., Palomino, P. T., Pessoa, M., Carvalho, L. S. G., Fernandes, D., Oliveira, E. H., Cristea, A. I., & Isotani, S. (2022). Gamification suffers from the novelty effect but benefits from the familiarization effect: Findings from a longitudinal study. *International Journal of Educational Technology in Higher Education*, *19*(1), 13.

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, *32*, 77–112. https://doi.org/10.1007/s10648-019-09498-w

Santos, A. C. G., Oliveira, W., Hamari, J., Rodrigues, L., Toda, A. M., Palomino, P. T., & Isotani, S. (2021). The relationship between user types and gamification designs. *User modeling and user-adapted interaction*, *31*(5), 907–940.

Santos, I., Felizardo, K. R., Gerosa, M. A., & Steinmacher, I. (2024). Game elements to engage students learning the open source software contribution process. *2024 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 59–70.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, *74*, 14–31. https://doi.org/https://doi.org/10.1016/j.ijhcs.2014.09.006

Tlili, A., Chen, N.-S., Denden, M., Essalmi, F., Jemni, M., Kinshuk, D., & Chang, M. (2019). Automatic modeling learner's personality using learning analytics approach in an intelligent moodle learning platform. *Interactive Learning Environments*, *31*, 15. https://doi.org/10.1080/10494820.2019.1636084

Werbach, K., & Hunter, D. (2012, January). *For the win: How game thinking can revolutionize your business*.

Xiao, Y., & Hew, K. F. (2024). Personalized gamification versus one-size-fits-all gamification in fully online learning: Effects on student motivational, behavioral and cognitive outcomes. *Learning and Individual Differences*, *113*, 102470.

Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review*, *30*, 100326. https://doi.org/https://doi.org/10.1016/j.edurev.2020.100326

Disrupting Reading through the Invisible Presence of AI: Rethinking Credibility and Critical Reading in Primary Education

Maja Kerneža
Faculty of Education, University of Maribor
Department of Elementary Education
Koroška cesta 160, 2000 Maribor, Slovenia
maja.kerneza1@um.si

Daniel Hari, Andrej Flogie, Dejan Zemljak
Faculty of Natural Sciences and Mathematics, University of Maribor
Institute for Contemporary Technologies
Koroška cesta 160, 2000 Maribor, Slovenia
{daniel.hari, andrej.flogie, dejan.zemljak1}@um.si

Abstract. *Primary school students increasingly encounter artificial intelligence (AI) in subtle, often invisible ways—through search engines, recommendation systems, and automated summaries. Although not direct users of generative AI, they are regularly influenced by its presence. This paper conceptualizes AI as a pedagogical disruption that reshapes reading from decoding to critical inquiry and evaluation. It highlights the need to foster critical reading skills that enable students to question sources, detect bias, and understand algorithmic influence, redefining literacy as an active, reflective process in an AI-shaped digital environment.*

Keywords. Critical reading, digital literacy, educational disruption, invisible presence of artificial intelligence, primary education

1 Introduction

In today's digitally mediated environments, children encounter AI technologies from early on–often before formal schooling. Voice assistants (e.g., Siri, Alexa), curated video content on YouTube, and adaptive learning apps are just some of the ways AI enters children's daily lives (Arora & Arora, 2022). These interactions may be direct–when children actively engage with AI tools–or indirect, through systems designed for adults but accessible to children. In both cases, children become active participants in AI-shaped digital ecosystems, though the mechanisms often remain opaque to young users (Gampe et al., 2023).

By primary school, many children have already developed habitual ways of interacting with digital content, shaped by personalization, filtering, and prediction (Voulgari et al., 2024). In school, students engage with AI-powered platforms (Butler & Starkey, 2024) both explicitly (e.g., digital textbooks with audio-summarization, adaptive learning systems) and implicitly (e.g. search engines, autocomplete, recommendation algorithms).

A growing share of content is no longer solely human-generated; it is created, curated, or filtered by AI–often without students realizing (Zhou et al., 2024). While they may not use generative AI tools directly, students are nonetheless regularly influenced by them – through personalized search results, automated suggestions, and algorithmic summaries that shape what they read, how they read, and how they evaluate its credibility.

This *invisible presence of AI* alters how students access, interpret, and evaluate information (Steele, 2023). The term refers to the systemic, often unnoticed influence of algorithms and automated systems that shape reading experiences—even when AI is not perceived as a technological actor. In this context, the traditional view of reading as decoding and comprehension are no longer sufficient. Contemporary reading literacy must also include the ability to identify sources, assess reliability, interpret intent, and distinguish between human and AI-generated content (Sanchez-Acedo et al., 2024).

These complex skills do not develop spontaneously, especially in young readers still mastering basic literacy. Their acquisition requires targeted pedagogical support and well-designed didactic strategies (Kerneža & Kordigel, 2022; Kordigel Aberšek & Kerneža, 2023).

This paper argues that the pervasive, often invisible integration of AI into reading environments constitutes a *productive disruption* in reading education. This shift requires rethinking the aims, methods, and assumptions of reading instruction. Rather than viewing AI solely as a tool, we conceptualize it as a pedagogical challenge that redefines reading as an interpretive, critical, and evaluative act. Special emphasis is placed on fostering critical reading literacy, enabling students to navigate automated

information landscape with awareness, independence, and responsibly.

2 Invisible AI and the Changing Nature of Reading

AI increasingly shapes children's reading experiences in ways that are not immediately visible or recognized as technologically mediated. While children are often labeled indirect users of AI, this distinction can be misleading. They frequently engage directly with AI-powered tools–voice assistants, educational platforms, or embedded recommendation systems. The key difference lies in visibility and agency: tools like ChatGPT require active input, while others work in the background, shaping information without user's awareness.

Many digital environments used by children are not designed with their best interest in mind. On platforms commonly accessed by primary school students, AI operates discreetly—through search suggestions, automated summaries, and recommendation algorithms (e.g., Nini & Kong, 2021; Zammit et al., 2022). This hidden integration creates a new reading reality: students often read algorithmically shaped content without realizing it.

Examples include search engines where autocomplete suggestions appear instantly, and personalized results vary by user profile. Featured snippets—automatically generated summaries—are often accepted uncritically (e.g., Al-Rawi et al., 2022; Huyen et al., 2019).

A similar dynamic occurs on platforms like YouTube Kids, where recommended content is based on previous viewing history (Rieder et al., 2020), rather than on didactic value (Vogelman-Natan, 2021). Digital textbooks increasingly offer AI-driven tools—auto-summaries, answer generators, and predictive suggestions. While convenient, these can lead to shallow understanding, as students may skip cognitive processing. AI becomes a shortcut to an answer, not a path to comprehension (Panit, 2024).

AI also appears in everyday tools such as web browsers and office software. Chrome, Word, and Gmail include features like Smart Compose, grammar checking, and stylistic corrections, enabling students to generate full sentences with little effort (Wang, 2024). Voice assistants (e.g., Siri, Alexa) provide general answers, often without sources or content (Hocutt, 2021).

In all these cases, AI silently co-constructs the reading experience. Content appears neutral, though it is filtered, summarized, and predicted by opaque algorithms. This encourages passive intake and discourages questioning and verification.

Thus, reading today is not only comprehension—it requires awareness of the technology behind the text. Students must learn to detect how AI shapes content develop strategies for critical evaluation. Without such skills, they risk information overload, manipulation, and superficial literacy.

In some cases, AI doesn't just reshape how students read–it reduces the need to read. Voice interfaces, audio summaries, and predictive tools can replace reading with listening or auto-generated responses. While beneficial for accessibility, such features raise concern: reduced reading motivation, less exposure to complex texts, and increased passivity. These trends underscore the urgency of cultivating reflective and critical reading habits in AI-mediated environments.

3 Critical Reading as an Educational Imperative

In a digital environment where readers rarely encounter a visible author, critical reading becomes essential. AI is no longer just a tool—it acts as a co-constructor of the reading process, with algorithms silently shaping how information is presented (Zemljak & Kerneža, 2024).

Decoding remains fundamental in early reading, but today's environment demands more. Students must develop skills in source evaluation, contextual understanding, and bias detection. Without questions like *Who created this? Why this source? Why now?*; reading risks becoming superficial skimming, where impressions replace understanding (Wexler et al., 2019).

Critical reading now requires attention to context: Who produced the content? For what purpose? Based on which data? In what environment? Stefanov Dimov, 2024). This context increasingly includes AI, which—often without the reader's awareness—influences selection, tone, and structure of what is read (Jiang et al., 2024). Reading thus becomes an act of interpretation, where student shifts from passive recipient to active meaning-maker. While reading has always been shaped by its context, the novelty lies in AI's dynamic, real-time shaping of content – requiring new forms of critical awareness.

The need is particularly urgent in primary education, where younger students often lack skills to verify sources or recognize misinformation (Kerneža & Kordigel, 2022; Kordigel Aberšek & Kerneža, 2023). They are rarely aware that AI is not a neutral transmitter but an active participant in shaping knowledge (Jiang et al., 2024). Teachers, meanwhile, may hesitate to address AI due to unclear curricula, lack of training, or doubts about developmental readiness (Su, 2024).

This paper argues that critical reading is not an advanced literacy skill–but its foundation. It includes comprehension, questioning, verification, justification, and awareness and digital mediation. These must be cultivated from the earliest school years.

When student ask–*Why did the algorithm show me this? Can I trust it? Who is behind it?*—they are not merely completing a task but engaging in democratic literacy.

In this way, reading instruction transforms–from linear knowledge transmission to a dialogue between student, text, and technology. AI is not just a topic of instruction but a didactic challenge, calling on the teacher to act not as a transmitter, but as a mentor of interpretation and critical judgment.

4 AI as a Pedagogical Opportunity for Fostering Critical Thinking

In schools, AI is often seen as a shortcut to answers. Yet this convenience can replace the effort of thinking. When students receive ready-made outputs from systems that don't think but combine, they may stop asking questions altogether (Mah et al., 2024). But this disruption offers a didactic opportunity: if addressed pedagogically, AI can become a catalyst for developing critical literacy.

Unlike traditional tools, AI acts as a co-creator. It can serve as a writer, explainer, dialogue partner—or a source of error and bias. This complexity invites reflection: *Who created this information? For what purpose? How can it be verified?* (Walker et al., 2025).

Research shows that less experienced often accept AI-outputs uncritically (Cail, 2024) and struggle to distinguish them from human ones (Fritz, 2025). Younger students are particularly vulnerable—they tend to trust structured, seemingly "definitive" answers without understanding how they were produced (Horodenko et al., 2024; Johnston, 2020).

Avoiding technology is not the answer. Instead, AI should be used as an imperfect partner in thinking, prompting comparison, verification and justification. Here, the teacher becomes a moderator of interpretation, not a supplier of answers.

To support this shift, researchers and practitioners have proposed tasks–aligned with critical digital literacy frameworks (Revell et al., 2024; Yue Yim, 2024)–which may be adapted for primary classrooms:

- **Dialogues with AI.** Students test how different prompts affect responses, exploring how wording, tone, or specificity changes AI output. This develops metacognitive awareness and shows that answers are shaped by interaction. In early grades, teachers can scaffold this through shared discussions and modeled prompting (Walter, 2024).
- **Human vs. AI comparison tasks**. Students compare AI-generated and human-written texts, focusing on tone, structure, sources, and reasoning. While AI texts often appear formal, they typically lack depth and nuance—a gap that becomes clear through classroom comparisons (Revell et al., 2024). Younger students might begin with simple narrative variations, progressing toward critical evaluation.
- **Detecting algorithmic bias**. Students analyze AI outputs for patterns—what is omitted, repeated, or simplified? They may notice missing perspectives or overused formats (like lists or definitions). Teachers guide discussions on how design choices and training data shape bias. Inclusive AI literacy frameworks emphasize integrating these ethical and epistemic questions early (Yue Yim, 2024).
- **Combining AI with trusted sources**. Students compare AI-generated content with information from books, encyclopedias, or reliable websites. They **synthesize both into a reflective summary**, citing each source. This fosters triangulation, evaluation, and the understanding that AI is **just one source**—not an authority. In early stages, this may take the form of comparing a print text with AI's version, guided by teacher questioning.

In this way, AI shifts from threat to provocation—inviting students to question, analyze, and co-construct meaning. The teacher becomes a mentor of doubt, opening space for distance, reflection, and dialogue with texts–even when the texts is machine-shaped.

5 Conclusion

This paper has argued that AI is not just a technical tool but a disruptive force that reshapes children's reading experiences from the earliest stages. The content they encounter is no longer neutral–it is filtered, summarized, or generated by invisible algorithmic systems. This opacity demands a new kind of reading attention.

Reading today extends beyond decoding text; it requires recognizing technological mediation. Questions such as *Who created this information? Why in this form? How can it be verified?* are now central to reading literacy. While not entirely new, such questions have become essential in an AI-shaped information ecosystem, where algorithms invisibly influence content selection, tone, and visibility.

Literacy in this context becomes a capacity for judgment, verification, and ethical orientation, calling for new didactic strategies and a redefined teacher role. The shift is not just methodological but pedagogical: from delivering knowledge to fostering critical engagement with information.

These insights align with European education policies, including the *Digital Education Plan* (2021–2027) (European Commission, 2020), and *DigComp 2.2* (Vuorikari et al., 2022), both of which emphasize early development of digital competence, critical thinking, and the ability to distinguish human from automated content. Embedding these frameworks in

classroom practice can help align national curricula with shared European goals.

Nonetheless, this study has limitations. It does not offer large-scale empirical data or classroom-based interventions. Its claims are conceptual and theoretical. Further research is needed on the developmental readiness of young learners for AI literacy, and on the types of scaffolding and teacher competencies required to support this meaningfully.

This paper does not call for a radical overhaul of reading education, but rather a recalibration. AI's invisible presence invites us to rethink not only what students read–but how and why it is presented. If critically framed and contextually supported, AI can become not just a challenge, but a pedagogical opportunity.

Acknowledgments

The authors would like to thank the research program P5-0433, Digital Restructuring of Deficit Occupations for Society 5.0 (Industry 4.0), founded by the Slovenian Research Agency (ARRS), for the support that significantly contributed to the progress of our study and was crucial for achieving the research objectives.

References

Al-Rawi, A., Celestini, C., Stewart, N., & Worku, N. (2022). How Google Autocomplete Algorithms about Conspiracy Theorists Mislead the Public. *M/C Journal, 25*(1). https://doi.org/10.5204/mcj.2852

Arora, A., & Arora, A. (2022). Effects of small voice control devices on children: Current challenges and future perspectives. *Archives of Disease in Childhood, 107*, 1129–1130. https://doi.org/10.1136/archdischild-2022-323888

Butler, L., & Starkey, L. (2024). OK Google, help me learn: An exploratory study of voice-activated artificial intelligence in the classroom. *Technology, Pedagogy and Education, 33*(2), 135–148. https://doi.org/10.1080/1475939X.2024.2311779

Cail, J. (2024). Visualization of AI Accuracy: A Novel Assignment for the Teaching of Critical Thinking and Science Writing. *Teaching of Psychology*, *52*(3), 285-290. https://doi.org/10.1177/00986283241289551

European Commission. (2020). *Digital Education Action Plan 2021–2027*. European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0624

Fritz, J. (2025). Understanding authorship in artificial intelligence-assisted works. *Journal of Intellectual Property Law & Practice*. https://doi.org/10.1093/jiplp/jpae119

Gampe, A, Zahner-Ritter, K., Joys Muller, J., & Schmid, S. R. (2023). How children speak with their voice assistant Sila depends on what they think about her. *Computers in Human Behavior, 143*,107693. https://doi.org/10.1016/j.chb.2023.107693

Hocutt, D. (2021). Interrogating Alexa: Holding voice assistants accountable for their answers. *39th ACM International Conference on Design of Communication (SIGDOC '21)* (pp. 142–150). Association for Computing Machinery. https://doi.org/10.1145/3472714.3473634

Horodenko, L., Tsymbalenko, Y., & Tsymbalenko, M. (2024). *Increasing trust in information among high school students through media literacy training*. *Вісник Книжкової палати, 11*(340), 15–22. https://doi.org/10.36273/2076-9555.2024.11(340).15-22

Jiang, J., Vetter, M., & Lucia, B. (2024). Toward a 'more-than-digital' AI literacy: Reimagining agency and authorship in the postdigital era with ChatGPT. *Postdigital Science and Education, 6*, 922–939. https://doi.org/10.1007/s42438-024-00477-1

Johnston, N. (2020). Living in the World of Fake News: High School Students' Evaluation of Information from Social Media Sites. *Journal of the Australian Library and Information Association*, *69*(4), 430–450. https://doi.org/10.1080/24750158.2020.1821146

Kerneža, M., & Kordigel Aberšek, M. (2022). Online reading in digital learning environments for primary school students. *Problems of Education in the 21st Century, 80*(6), 836–850. https://dx.doi.org/10.33225/pec/22.80.836

Kordigel Aberšek, M., & Kerneža, M. (2023). *Age conditioned online research and comprehension skills in primary school students*. In L. Gómez Chova, C. González Martínez, & J. Lees (Eds.), *INTED2023: 17th Annual International Technology, Education and Development Conference, 6–8 March 2023, Valencia, Spain* (pp. 5496–5504). IATED Academy. https://doi.org/10.21125/inted.2023.1440

Le, H., Maragh, R., Ekdale, B., High, A., Havens, T., & Shafiq, Z. (2019). Measuring political personalization of Google News search. *World Wide Web Conference (WWW '19)* (pp. 2957–2963). Association for Computing Machinery. https://doi.org/10.1145/3308558.3313682

Mah, C., Walker, H., Phalen, L., Levine, S., Beck, S. W., & Pittman, J. (2024). Beyond CheatBots:

Examining Tensions in Teachers' and Students' Perceptions of Cheating and Learning with ChatGPT. *Education Sciences*, *14*(5), 500. https://doi.org/10.3390/educsci14050500

Nini, H., & Kong, D. (2021). Research on the Application of children's reading analysis based on artificial intelligence – take small "Racoon Reading" and "Jiao Jiao Reading" as examples. *Journal of Physics: Conference Series, 1848*, 012121. https://dx.doi.org/10.1088/1742-6596/1848/1/012121

Panit, N. M. (2024). Can critical thinking and AI work together? Observations on science, mathematics and language instructors. *Environment and Social Psychology, 9*(11), 3141. https://doi.org/10.59429/esp.v9i11.3141

Reider, B., Gordon, G., & Sileno, G. (2020). Mapping value(s) in AI: The case of YouTube. *AoIR Selected Papers of Internet Research*, *2020*. https://doi.org/10.5210/spir.v2020i0.11315

Revell, T., Yeadon, W., Cahilly-Bretzin, G., Clarke, I., Manning, G., Jones, J, Mulley, ., Pascual, R. J., Bradley, N., Thomas, D., & Leneghan, F. (2024). ChatGPT versus human essayists: An exploration of the impact of artificial intelligence for authorship and academic integrity in the humanities. *International Journal of Educational Integrity, 20,* 18. https://doi.org/10.1007/s40979-024-00161-8

Sanchez-Acedo, A., Carbonell-Alcocer, A., Gertrudix, M., & Rubio-Tamayo, J.-L. (2024). The challenges of media and information literacy in the artificial intelligence ecology: deepfakes and misinformation. *Communication & Society*, *37*(4), 223-239. https://doi.org/10.15581/003.37.4.223-239

Steele, J. L. (2023). To GPT or not GPT? Empowering our students to learn with AI. *Computers and Education: Artificial Intelligence, 5*. 100160. https://doi.org/10.1016/j.caeai.2023.100160

Stefanov Dimov, P. (2024). An online tool for misinformation detection: Enhancing media literacy and critical thinking skills. *Online Reading in Educational Research, 4*(2), 61–69. https://doi.org/10.58524/oler.v4i2.413

Su, J. (2024). Kindergarten teachers' perceptions of AI literacy education for young children. *International Journal of Technology and Design Education, 34*, 1665–1685. https://doi.org/10.1007/s10798-024-09876-8

Vahedian Movahed, S., & Martin, F. (2025). *Ask Me Anything: Exploring children's attitudes toward an age-tailored AI-powered chatbot* (arXiv:2502.14217) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.14217

Vogelman-Natan, K. (2021). The algorithm playground: a content analysis of user-produced children's videos on YouTube. *AoIR Selected Papers of Internet Research*, *2021*. https://doi.org/10.5210/spir.v2021i0.12071

Voulgari, I., Lavidas, K., Aravantinos, S., Sypsa, S., & Sfyroera, M. (2024). Exploring the role, implementation, and educational implications of AI in early childhood Education. In A. M. Olney, I. A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (eds.), *Artificial intelligence in education. Posters and late breaking results, workshops and tutorials, industry, and innovation tracks, practitioners, doctoral consortium and blue sky. AIED 2024. Communications in Computer and Information Science, 2151*. Springer, Cham. https://doi.org/10.1007/978-3-031-64312-5_4

Vuorikari, R., Kluzer, S., & Punie, Y. (2022). *DigComp 2.2: The digital competence framework for citizens – With new examples of knowledge, skills and attitudes*. European Union. https://doi.org/10.2760/115376

Walker, P. B., Haase, J. J., Mehalick, M. L., Steele, C. T., Russell, D. W., & Davidson, I. N. (2025). Harnessing Metacognition for Safe and Responsible AI. *Technologies*, *13*(3), 107. https://doi.org/10.3390/technologies13030107

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*, 15. https://doi.org/10.1186/s41239-024-00448-3

Wang, C. Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers. *Tech Know Learn* (2024). https://doi.org/10.1007/s10758-024-09744-3

Wexler, J., Swanson, E., Kurz, L. A., Shelton, A., & Vaughn, S. (2019). Enhancing Reading Comprehension in Middle School Classrooms Using a Critical Reading Routine. *Intervention in School and Clinic*, *55*(4), 203-213. https://doi.org/10.1177/1053451219855738

Yue Yim, I. H. (2024). A critical review of teaching and learning artificial intelligence (AI) literacy: Developing an intelligence-based AI literacy framework for primary school education. *Computers and Education: Artificial Intelligence, 7*, 100319. https://doi.org/10.1016/j.caeai.2024.100319

Zammit, M., Voulgari, I., Liapis, A., & Yannakakis, G. N. (2022). Learn to machine learn via games in the classroom. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.913530

Zemljak, D., & Kerneža, M. (2024). Embracing doubt: Critical thinking in the age of artificial intelligence in education. *International Conference on Science, Innovations and Global Solutions* (pp. 330–334). Futurity Research Publishing. https://futurity-publishing.com/international-conference-on-science-innovations-and-global-solutions-archive

Zhou, X., Zhou, Y., Gong, Y., Cai, Z., Qiu, A., Xiao, Q., Antle, A. N., & Bai, Z. (2024). "Bee and I need diversity!" Break filter bubbles in recommendation systems through embodied AI learning. *23rd Annual ACM Interaction Design and Children Conference* (pp. 44–61). ACM. https://doi.org/10.1145/3628516.3655802

Teachers' digital competences and the connection of lifelong learning in higher education

Jasenka Begić, Goran Đambić, Ana Lokas Ćošković, Vanja Šebek, Ana Tecilazić, Karolina Zelenika
Algebra Bernays University
Department of Interdisciplinary Sciences
Gradišćanska Street 24, 10 000 Zagreb, Croatia
{jasenka.begic, goran.dambic, ana.lokascoskovic, vanja.sebek, ana.tecilazic, karolina.zelenika}@algebra.hr

Abstract*. The aim of the research was to examine the relationship between age and gender of teachers, their field of teaching, location of the institution and the participation in lifelong learning with the level of acquired digital competences. 215 teachers from higher education institutions in the Republic of Croatia completed an online questionnaire based on the DigCompEdu model. The largest number of teachers is at levels A2 (Explorer) and B1 (Integrator), with the highest results in the dimension* Digital sources and materials*, and the lowest in the dimension* Monitoring and evaluation*. The overall level of competence corresponds to level B1. Statistically significant differences were found for the location of the institution and the participation in international training programs, while differences by age, gender and the field of teaching were not found. The results indicate the need for adapted and purposeful professional development programs in the field of digital competences.*

Keywords. digital competence, DigCompEdu, higher education, lifelong learning

1 Introduction

The development of information and communication technologies (ICT) has profoundly influenced all aspects of human activity, including the field of education. Over the past decade, digital technologies have become more than mere "tools"; they now represent a key element in the transformation of society, and virtually every area of activity now relies on them (European Commission, 2020). The introduction of ICT into education has significantly transformed traditional approaches to learning and teaching, as well as the way these processes are understood and structured. This development has also enabled innovative possibilities that enhance the educational experience. The application of ICT in education today enables, among other elements, the creation of interactive and personalized learning environments, facilitates access for students with disabilities and those living in remote areas, as well as enhances collaboration and communication between students and teachers (Iqbal, 2024). Today's students, who have grown up immersed in a digital environment, are accustomed to constant access to information and the active use of technology. Furthermore, they expect their education to equip them with the knowledge and skills required for employability, continuous professional development and active participation in society. Consequently, adapting teaching methods becomes essential, with digital technologies significantly contributing to improving the quality of teaching and enabling the better achievement of learning outcomes (Kučina Softić, 2020).

2 Literature review

In the context of today's higher education, the appropriate, purposeful and effective integration of digital technologies into the teaching process largely depends on the level of teachers' digital competences, systematic institutional support and an ongoing professional development (Basilotta-Gómez-Pablos et al., 2021; Howard & Tondeur, 2023). Teachers' digital competence encompasses far more than technical proficiency; it represents a complex set of knowledge, skills and attitudes that includes pedagogical expertise, methodological adaptability and the ability to apply digital tools critically, reflectively and purposefully across diverse educational contexts (Basilotta-Gómez-Pablos et al., 2021; Kučina Softić, 2020). Moreover, current approaches emphasize the importance of structured and personalized professional development programs that integrate technology with pedagogical principles, improve teaching quality, achieve learning outcomes and more effectively adapt to diverse learning styles and student needs (Basilotta-Gómez-Pablos et al., 2021; Howard & Tondeur, 2023; Palvia et al., 2024). Ultimately, the traditional perception of

the teacher as the sole source of knowledge and skills is being replaced by new professional roles in which the teacher acts as a mentor, facilitator, advisor and the designer of the educational process. Accordingly, it is essential that educators continuously acquire new knowledge and competences (Instefjord & Munthe, 2017; Jurčić, 2014; Kučina Softić, 2020).

2.1 Lifelong learning in higher education

In the recent decades, the concept of lifelong learning has become a central pillar of European education policy, particularly as a strategic response to accelerating social and technological change (European Commission, 2021). Lifelong learning and continuous professional development are essential mechanisms for fostering personal and professional growth, preventing stagnation and enhancing adaptability in dynamic educational environments (OECD, 2021). For educators, these processes are critical in maintaining the quality of teaching and in responding effectively to the evolving demands of educational systems.

Lifelong learning among teachers includes a broad spectrum of activities, ranging from participation in professional training and reflective practice to research engagement and the development of professional autonomy (Day, 1999; Trevisan et. al., 2024). Formal and informal learning opportunities, such as seminars, workshops, conferences and mentoring within professional learning communities, are recognized as vital for sustaining teachers' professional growth. Reflective practice is emphasized as a core component of professional development that enables teachers to critically evaluate their teaching approaches and adapt to diverse classroom needs (Vizek-Vidović, 2005).

Since the early 2000s, policy shifts within the EU have encouraged teacher education institutions to reconfigure their curricula with a stronger emphasis on research-oriented teaching, problem-based learning and the development of digital competence (European Commission, 2021). Given the rapid pace of technological advancement, educators are now expected to continuously update their digital skills and adopt innovative teaching practices. Hower, the development of digital competence does not occur in isolation: it requires coherent institutional strategies, adequate infrastructure and a high level of internal motivation from educators themselves (Howard & Tondeur, 2023; Instefjord & Munthe, 2017; Vizek-Vidović, 2005).

2.2 Digital competences and DigCompEdu model

The concepts of *digital literacy* and *digital competence* are increasingly recognized, both within the context of education and much broader societal development, as fundamental prerequisites for building a knowledge-based society rooted in innovation, technological advancement and social inclusion. These constructs extend far beyond the realm of purely technical skills related to the creation and management of digital tools: they encompass a complex spectrum of abilities associated with information, media and communication literacy, as well as critical thinking and the ethical use of digital technologies (Ferrari, 2013; Ilomäki et al., 2011). Accordingly, digital competence is not viewed in isolation, but rather as an integral component of the broader framework of 21st-century skills, which also includes social, civic, cultural and metacognitive dimensions of knowledge and practice (Erstad & Voogt, 2018).

The Lisbon Strategy (European Council, 2000) first identified digital competence as essential to building a competitive, knowledge-based economy. This goal was further developed in strategic documents such as the *Digital Agenda for Europe*, *Europe 2020*, and the *Digital Education Action Plans* (2018; 2021-2027) (European Commission, 2018; 2020). These frameworks stress the need for systematic, institutionally supported development of digital competences across all educational levels, highlighting both early education and the ongoing professional development of educators (Redecker, 2017).

Since teachers and educators are increasingly recognized as key agents of change in the digital age (Jurčić, 2014; Kučina Softić, 2020), there is a growing need to clearly define educators' digital competence, enhance levels of digital literacy and ensure the high-quality education in this area. The overarching aim of this approach is to foster the development of the knowledge, skills and attitudes necessary for active and equitable participation in contemporary digital society (Redecker, 2017). Within this context, several theoretical frameworks have been developed to assess and support the advancement of teachers' digital competence. Among the most prominent are the TPACK model (Technological Pedagogical Content Knowledge), the UNESCO ICT Competency Framework for Teachers, the PEAT model (Pedagogical, Ethical, Attitudinal, Technical), and the DigCompEdu framework (European Framework for the Digital Competence of Educators). These models highlight the importance of integrating pedagogy with technology and facilitate a structured and comprehensive understanding of the specific components of digital competence in the teaching practice (Tomczyk & Fedeli, 2021).

The DigCompEdu model provides a detailed description of 22 digital competences, which are organized into six distinct areas and classified across six levels of proficiency: A1, A2, B1, B2, C1 and C2. The model assesses various dimensions of digital competence and outlines what it means to be a digitally competent educator. The six defined areas include *Professional Engagement*, *Digital Resources*, *Teaching and Learning*, *Assessment*, *Empowering Learners* and *Facilitating Learners' Digital Competence*. In addition, a seventh area, *Open*

Education, has been introduced (European Commission, 2023). Each competency is further operationalized across six proficiency levels, from the introductory level (A1) to the highly advanced level (C2), thereby supporting an individualized approach to teachers' professional development and contributing to clearly defined objectives within the institutional training strategies. The purpose of the model is to support and motivate educators and researchers to use digital tools to enhance the teaching process and foster educational innovation (European Commission, 2020). Ultimately, DigCompEdu now plays a key role in guiding the development, assessment and advancement of educators' digital competence, while contributing to the overall quality of education and the effective integration of digital technologies into teaching and learning (Suzer & Koc, 2024).

3 Methodology

The research aims to determine whether there is a connection between factors such as teachers' age, field of teaching, regional positioning of the employing institution and the participation in lifelong learning activities with the acquired level of digital competences. In addition, the research aims to identify key challenges in this field and explore factors that are associated with the level of acquired digital competences of teachers.

The specific objectives of the research are as:

1. To examine whether teachers' digital competences differ according to their age and gender.
2. To examine differences in the level of digital competences between teachers of different fields of teaching.
3. To examine the connection of the regional positioning of the employing institution with the level of teachers' digital competences.
4. To examine the connection between teachers' participation in lifelong learning activities and the level of acquired digital competences, whereby different aspects of lifelong learning will be analysed: local vs. international participation and digitally focused vs. general professional development.

The set hypotheses concerning the objectives are:

- Hypothesis 1: Younger teachers have a statistically significantly higher level of digital competences compared to older teachers.
- Hypothesis 2: Female teachers have a statistically significantly higher level of digital competences compared to male teachers.
- Hypothesis 3: Teachers working in technical and natural science disciplines have statistically significantly more developed digital competences compared to teachers working in social, humanities and art disciplines.
- Hypothesis 4: Teachers working in institutions located in larger urban centers have a statistically significantly higher level of digital competence compared to teachers working in less developed regions.
- Hypothesis 5: Teachers who participate in lifelong learning activities have a statistically significantly higher level of digital competence compared to teachers who do not participate in such activities.
  Within this hypothesis, additional aspects will be examined: (a) teachers participating in international lifelong learning programs are more digitally competent than those participating only in local programs, and (b) teachers participating in programs focused on digital skills achieve higher levels of digital competence compared to those participating in professional development programs focused on general skills.

3.1 Procedure and participants

The research was conducted in March and April 2025 via online survey created in Google Forms, using the snowball method. The survey was distributed to teachers at higher education institutions in Croatia and then circulated among colleagues until a sufficient sample was reached. Participation was anonymous and voluntary, with prior information about the purpose of the research and the time to complete it (up to 20 minutes). The data was statistically processed using the Mann-Whitney U test.

215 respondents participated, of whom 65.6% were women and 34.4% were men. The most represented age group was 40-49 years (34%), and the least represented was 25-29 years (4.2%). Most respondents work at universities (45.1%), most in the City of Zagreb (25.6%), while the least in Bjelovar-Bilogora County (1.4%). The most common job titles are lecturer (18.6%) and senior lecturer (18.1%), with the highest number of respondents coming from the social sciences (50.7%). Most respondents (87.4%) teach undergraduate students who have had no previous career.

3.2 Research instrument

The online questionnaire used for data collection consisted of two parts. The first part gathered data to assess teachers' digital competence using a self-assessment tool based on the European Framework for the Digital Competence of Educators (DigCompEdu model), which includes 22 competences organized in six areas: *Professional Engagement*, *Digital Resources*, *Teaching and Learning*, *Assessment*, *Empowering Learners* and *Facilitating Learners'*

Digital Competence. An additional, seventh area titled *Open Education* was also included (European Commission, 2023). The competences are defined across six proficiency levels (A1, A2, B1, B2, C1, C2):

- A1: Newcomer – I recognize that digital technologies can enhance my professional practice.
- A2: Explorer – I have experimented with using digital technologies in my professional context.
- B1: Integrator – I regularly apply digital technologies in my teaching and professional routines.
- B2: Expert – I use digital technologies confidently and critically to enhance and transform my professional practice.
- C1: Leader – I apply digital technologies in strategic ways and actively share my expertise with students and colleagues.
- C2: Pioneer – I initiate and lead innovative, evidence-based practices for the integration of digital technologies within my academic institution and the wider educational community (Luić & Rončević, 2023; Redecker, 2020).

In line with the research objectives, the questionnaire was adapted by omitting the area *Facilitating Learners' Digital Competence* and adding a new area titled *Working Conditions and Lifelong Learning*. The decision to omit *Facilitating Learners' Digital Competence* was based on two key considerations. While the omitted area is undoubtedly important in the broader context of digital education, it was considered less relevant for the aims of this study (educators' self-assessed skills, training and working conditions). Second, the inclusion of all originally proposed areas made the questionnaire excessively long and potentially burdensome for respondents. The added area *Working Conditions and Lifelong Learning* was included to analyse contextual factors, such as access to technology, institutional support and opportunities for professional development in the context of the development and application of digital competences. This adaptation ensured the instrument remained focused, manageable and aligned with the study's goals.

To conclude, this study offers a theoretical contribution by situating the DigCompEdu framework within the context of Croatian higher education and empirically examining its adaptability. Specifically, the additional area, *Working Conditions and Lifelong Learning*, was integrated to analyse the socio-institutional factors connected with the development of digital competences that align with established lifelong learning policies within the EU (European Commission, 2023).

4 Results

The findings revealed that 24.7% of respondents have been using digital technologies in teaching for 6 to 9 years. Only 4.2% of respondents have the longest experience, 20 years or more, while 4.7% have been using digital technologies for less than one year. Regarding the use of digital technologies in teaching over the past three months, the largest percentage of respondents (26%) reported using digital technologies between 76% and 100% of their teaching time. Following this group, 22.3% of respondents used digital technologies during 26% to 50% of their teaching time. The smallest percentage, up to 10% of teaching time, was reported by 10.7% of respondents. Among digital tools used for teaching, presentations were the most common (98.1%), followed by video and audio content (80%), and digital quizzes or surveys (50.2%). In contrast, online simulators and AI tools were the least frequently used (0.5%).

4.1 Self-assessment of digital competences

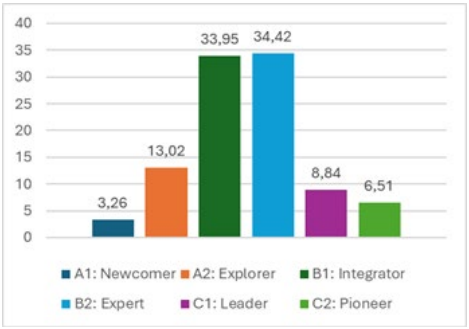


Figure 1. Self-assessed levels of digital competence before completing the questionnaire

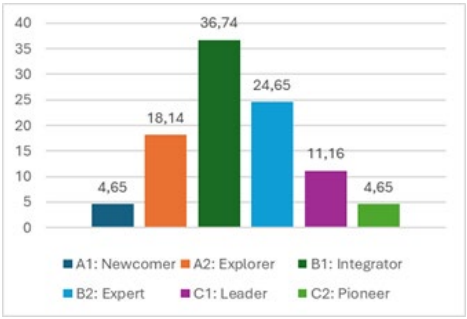


Figure 2. Self-assessed levels of digital competence after completing the questionnaire

Respondents evaluated their digital competences both before and after completing the entire questionnaire. As shown in Fig. 1 and 2, compared to their initial self-assessment, the percentage of respondents rating their digital competences at levels A1, A2, B1 and C1 notably increased after completing the questionnaire. On the other hand, while the largest percentage of respondents (34.4%) initially assessed their digital competency at level B2 (expert level), this decreased to 24.7% after completing the entire questionnaire. Similarly, before completing the questionnaire, 6.5% of respondents assessed their digital competency at level C2 (leader), but this percentage dropped to 4.6% afterward.

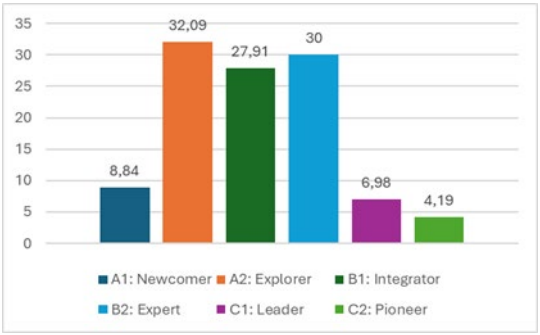


Figure 3. Current levels of digital competence among respondents

Fig. 3 shows the percentages of respondents' current digital competences, determined through an analysis of their responses across each of the six digital competency areas, which were then used to calculate the overall digital competency score for each respondent. The highest percentage of respondents (33%) is at level A2 (Researcher), followed by level B1 (Experimenter) at 27.9%. The lowest percentage of respondents is found at level C2 (Leader), with only 4.2%. Starting from level A2, results reveal a general downward trend in the percentage of respondents as the level of digital competency increases.

Table 1. Basic descriptive statistics for each of the six analyzed areas of digital competence and the overall digital competence score for all respondents

Area	N	M	SD	DigCompEdu level
Professional Engagement	215	3,05	1,075	B1
Digital Resources and materials	215	3,61	1,290	B2
Teaching and Learning	215	2,77	1,442	B1
Assessment and Feedback	215	2,42	1,349	A2
Learner Empowerment	215	2,80	1,493	B1
Open Education	215	2,48	1,506	B1
Digital competence	215	2,85	1,152	B1

Table 1 shows basic descriptive parameters for each of the six analysed areas of digital competences, as well as the overall score across all digital competency areas for all respondents. The highest average response value was recorded in Digital Re-sources and Materials (M = 3.61, SD = 1.290), while the lowest average was noted in Monitoring and Evaluation (M = 2.42, SD = 1.349). In other words, considering all respondents collectively, their highest level of digital competency (B2) was in Digital Resources and Materials, while the lowest competency level (A2) was observed in Monitoring and Evaluation. The mean value across all analysed area, representing respondents' overall digital competences is M = 2.85 (SD = 1.152), corresponding to level B1.

4.2 Participation in lifelong learning

Most respondents (56.7%) participated in programs focused on various pedagogical and scientific topics, while slightly fewer respondents (53%) took part in programs aimed at general teaching methods. Additionally, 52.1% of respondents participated in programs focused specifically on digital skills. Regarding the frequency of participation in professional development activities related to digital technologies over the past three years, half of the respondents (50.7%) participated fewer than 5 times, 28.4% participated between 5 and 10 times, 10.2% participated more than 10 times, and 8.8% did not participate at all.

Regarding the types of professional development in digital technologies over the past three years, 67% of respondents participated in online webinars and virtual conferences. Formal educational seminars, workshops and training sessions were attended by 43.3% of respondents, compared to 36.3% who participated in informal educational seminars, workshops and training. National professional seminars and conferences attracted 24.7% of respondents, while international professional seminars and conferences had a participation rate of 25.6%. More than half of the respondents (59.5%) received training focused on technology use and e-learning, 41.9% underwent training in the use of office tools for lesson preparation, and 40.5% were trained in identifying plagiarized student work. The lowest percentage of respondents participated in training related to the use of blogs and wikis.

4.3 Hypothesis testing

Table 2. Mann-Whitney U Test results (Digital competence: younger and older respondents)

	Age	N	Mean rank	Sum of ranks	M-W U	p-value
Digital competence	Under 39	47	97,57	4586,00	3458,000	0,194
	Over 40	168	110,92	18634,00		

To test the first hypothesis of this study, the Mann-Whitney U test was used, and the results are presented in Table 2. Given the determined p-value ($p > 0.05$), it can be concluded that there is no statistically significant difference between younger and older respondents in their level of digital competences, thus the H1 hypothesis is rejected.

Table 3. Mann-Whitney U Test results (Digital competence: male and female respondents)

	Gender	N	Mean rank	Sum of ranks	M-W U	p-value
Digital competence	F	141	103,30	14566,00	4555,000	0,127
	M	74	116,95	8654,00		

The results of the Mann-Whitney U test for examining differences in digital competences based on gender are presented in Table 3. No statistically significant difference was found between male and female respondents in terms of digital competences ($p > 0.05$), therefore the H2 hypothesis is rejected.

Table 4. Mann-Whitney U Test results (Digital competence: field of teaching)

	In which of the following fields do you teach?	N	Mean rank	Sum of ranks	M-W U	p-value
Digital competence	Technical and Natural Sciences	86	109,08	9380,50	5368,500	0,760
	Social Sciences, Humanities and Arts	128	106,44	13624,50		

Table 4 presents the results of testing for a statistically significant difference in digital competences between respondents teaching technical and natural sciences and those teaching in social sciences, humanities and arts. No statistically significant difference was found between these two groups of respondents ($p > 0.05$), therefore the H3 hypothesis is rejected.

Table 5. Mann-Whitney U Test results (Digital competence: teaching/work location)

	County of employment	N	Mean rank	Sum of ranks	M-W U	p-value
Digital competence	City of Zagreb and Zagreb County	59	123,85	7307,00	3667,000	0,022
	Other counties	156	102,01	15913,00		

The results of the Mann-Whitney U test for examining differences in digital competences based on teaching/work location are presented in Table 5. Respondents working in the City of Zagreb and Zagreb County have a statistically significantly higher level of digital competences compared to respondents working in other counties (M-W U = 3667.000, $p < 0.05$). Based on this result, the H4 hypothesis is accepted.

Table 6. Mann-Whitney U Test results (Digital competence: forms of professional development)

	In which forms of professional development in the field of digital technologies have you participated in the past three years?	N	Mean rank	Sum of ranks	M-W U	p-value
Digital competence	International seminars and conferences	56	132,37	7412,50	2359,500	0,000
	National seminars and conferences	146	89,66	13090,50		

Table 6 presents the results of the Mann-Whitney U test for determining whether there is a statistically significant difference in digital competences between respondents who, in the past three years, participated in international professional seminars and conferences and those who participated in national professional seminars and conferences. Since the p-value is less than 0.05, it can be concluded that there is a statistically significant difference between these two groups of respondents. Respondents who attend international professional seminars and conferences report a higher level of digital competences compared to those who attended national professional seminars and conferences.

Table 7. Mann-Whitney U Test results (Digital competence: types of professional development programs)

	In which types of professional development have you participated in the past three years?	N	Mean rank	Sum of ranks	M-W U	p-value
Digital competence	Programs focused on digital skills	156	105,44	16448,50	3587,500	0,410
	Programs focused on general teaching methods	50	97,45	4872,50		

As a part of the final hypothesis, the existence of a statistically significant difference in the level of digital competences was also examined concerning the types of professional development programs respondents participated in over the past three years. The results of the Mann-Whitney U test are presented in Table 7. Based on the findings ($p > 0.05$), it can be concluded that there is no statistically significant difference between respondents who participated in programs focused on digital skills and those who participated in programs focused on general teaching methods. Therefore, the final H5 hypothesis is partially accepted.

5 Discussion

The observed diversity in teachers' experiences with the use of digital technologies in teaching is consistent with findings from previous studies, which indicate that the acquired level of digital competence is significantly associated with factors such as access to technology, opportunities for professional development, as well as personal attitudes and perceptions of the usefulness of technology in teaching (Ertmer & Ottenbreit-Leftwich, 2010; Instefjord & Munthe, 2017). Regarding the frequency of digital technology use in teaching over the past three months, the data indicates a certain polarization: 26% of respondents use digital technologies in more than 75% of their teaching time, while 10.7% use them in less than 10% of time. These results suggest the existence of a group of teachers who have highly integrated technology into their daily teaching practice, whereas a significant group of respondents use digital tools only occasionally. At the same time, numerous studies confirm that the intensity of technology use is not a sufficient indicator of the quality of its integration into the teaching process. Teachers often use technology in ways that do not transform the fundamental structure of teaching, but rather enhance it on a technical level; this is considered the first level, or the basic one, of technology implementation (Ertmer & Ottenbreit-Leftwich, 2010). In this context, the relatively low intensity of use among some teachers may result from

various barriers, including insufficient professional support, underdeveloped skills and low self-confidence in using digital tools (European Commission, 2020; Krumsvik, 2014).

The analysis of the frequency of use of specific digital tools in teaching shows that most respondents primarily use various presentation tools (98.1%) and multimedia content such as video and audio materials (80%). In contrast, advanced digital tools such as online simulators or artificial intelligence-based systems are rarely used (0.5%). These findings indicate a strong orientation among teachers toward basic and traditional forms of digital technologies, primarily focused on the frontal content delivery (e.g. replacing the blackboard and textbooks with PowerPoint presentations). However, this approach remains within a learning model that emphasizes passive reception of information and does not necessarily lead to meaningful enhancement of pedagogical processes (Laurillard, 2012). Additionally, the use of advanced tools such as gamified learning systems or AI-powered adaptive learning technologies can significantly increase student engagement, foster higher cognitive skills such as critical thinking and problem-solving, and ultimately improve learning outcomes (Holmes et al., 2019; Means et al., 2013). The limited integration of such technologies in teaching is often attributed to a lack of specific digital competences, constrained institutional resources and the absence of systematic professional support (Voogt et al., 2011; OECD, 2021).

The results of the self-assessment of digital competences before and after completing the questionnaire indicate a notable adjustment among respondents, particularly among those who initially rated their competence level as high (levels B2 - Expert and C2 - Pioneer). This finding suggests that the use of structured and validated self-assessment models, such as DigCompEdu, can support respondents in gaining metacognitive insight into their strengths and weaknesses, leading to a more realistic self-evaluation that may serve as a foundation for further professional development (Koehler et al., 2014; Redecker, 2017).

The final analysis of digital competence levels shows that the largest group of respondents fall within levels A2 (33%) and B1 (27.9%), which correspond to basic to moderately developed digital competences. In contrast, only 4.2% of respondents believe they have reached level C2, the highest level in the implementation of digital technologies in the educational process. This distribution confirms findings from previous studies that indicate a low representation of advanced digital competences among teachers in the European Union (European Commission, 2020; Ferrari, 2013). According to the DigCompEdu framework, the average digital competence level of European educators is B1, which represents a moderately advanced level of competence. According to Suzer and Koc (2024), educators operating at this proficiency level implement digital tools regularly and in diverse ways; however, this use is not embedded within a structured pedagogical framework nor subjected to systematic evaluation of effectiveness. In practice, digital tools are incorporated into occasional projects and include platforms such as learning management systems (LMS), webinars, digital presentations, as well as basic communication and task-oriented applications. Nevertheless, these practices often lack comprehensive integration, ongoing assessment and sustained professional support.

Although the proportion of educators at the highest level of digital competence (C2) remains low, the presence of post-assessment recalibration indicates a degree of metacognitive awareness and a willingness to engage in professional growth. To advance digital competence levels towards B2, C1 or C2, a coordinated strategy is necessary, one that includes systematic professional development, structured collaboration among stakeholders, continuous evaluation and reflective practice, along with more robust institutional support (Janssen et al., 2013; Redecker, 2017).

In line with these findings, within the area of *Teaching and Learning* (corresponding to level B1 - Integrator: “I regularly apply digital technologies in my professional practice”), educators demonstrated the ability to effectively search for, select, create and share digital resources. They also showed awareness of copyright regulations, ethical principles and the adaptation of content to various instructional contexts. This competence is considered foundational in the digital transformation of teaching and is frequently developed informally through the routine use of online content, open educational resources and multimedia tools such as presentations, videos and e-textbooks (European Commission, 2020; Instefjord & Munthe, 2017). On the other hand, the lowest average competence level was recorded in *Assessment*, which corresponds to level A2 - Explorer: “I have experimented with using digital technologies in my professional context.” The limited use of digital tools for assessment and feedback may stem from several factors: insufficient knowledge of how to apply digital assessment platforms, quizzes, or learning analytics, and a general lack of trust in the reliability and fairness of digital assessment formats (Schildkamp et al., 2020). Furthermore, existing research indicates that teachers often lack adequate institutional support specifically in digital assessment. This area is frequently underrepresented in professional development programs, despite its potential to enhance monitoring of student progress, provide timely feedback and support personalized learning approaches (Redecker & Punie, 2017). It can therefore be concluded that effective digital assessment requires higher levels of competence, which most educators have not yet fully attained.

An analysis of the tested hypotheses indicates that certain demographic variables, such as age, gender and academic field of teaching are not statistically significantly associated with educators’ levels of

digital competence. In contrast, significant differences were observed for institution location and participation in international programs. While previous studies have documented a considerable digital divide across genders and suggested that teachers in technical fields tend to use digital technologies more extensively than those in the social sciences (Müller & Aleksa Varga, 2020), recent findings point to a narrowing of gender-based disparities in the application of digital skills in educational settings (Sánchez Prieto et al., 2020). These results are consistent with a growing body of research suggesting that digital competence is increasingly determined by factors such as access to professional development, institutional support and individual motivation, rather than inherent demographic characteristics (Instefjord & Munthe, 2017; Siddiq et al., 2016). Given the wider availability of digital tools and the improvement of professional training opportunities for educators across various profiles, the findings of this study reflect ongoing shifts in the educational landscape and contribute to a more nuanced understanding that challenges persist beyond gender and disciplinary stereotypes.

Although the data indicate that most respondents report frequent internet use and general satisfaction with the availability of institutional resources, high levels of internet usage do not necessarily translate into pedagogically meaningful or purposeful integration of digital tools in education (OECD, 2021). Ultimately, the overall intensity of teacher participation in professional development related to digital technologies remains low: 50.7% of respondents stated that they had participated in fewer than five such activities over the past three years. This finding stands in contrast to international trends. According to OECD research (2019), in many countries worldwide, most teachers engage in some form of lifelong learning activity, with an average of 94% of teachers across OECD countries having participated in at least one professional development activity within a single year.

Given the regional disparities in resource availability and institutional infrastructure, the results revealed statistically significant differences in favour of respondents from the City of Zagreb and Zagreb County. According to findings by the European Commission (2020) and the OECD (2021), more urban and economically developed areas, such as Zagreb, tend to benefit from more advanced technical infrastructure, greater access to both formal and informal education opportunities, and increased participation in international and research projects. Ultimately, these contextual factors may play a critical long-term role in shaping the development of educators' digital competences.

Furthermore, educators who have participated in international professional seminars and conferences demonstrate higher levels of digital competence compared to those whose experience is primarily based on national training initiatives. Participation in international education programs provides access to contemporary trends and current research insights, fosters intercultural knowledge exchange, and facilitates critical reflection on one's teaching practices within a global context. It also offers exposure to information and practices that are often unavailable through national-level activities. Engagement with innovative pedagogical approaches and international experiences can significantly enhance the adoption of new knowledge, the implementation of innovative methods, and alignment with global educational standards. As a result, such experiences tend to have a far greater transformative impact on educators' professional growth (Darling-Hammond et al., 2017; Revuelta-Domínguez et al., 2022).

Interestingly, no difference was found in the acquired level of digital skills between respondents who attended programs focused on digital competences and those who participated in programs oriented toward general scientific and pedagogical methods. According to Redecker (2017), in order to truly contribute to the development of specific competences, digital skills training must be pedagogically relevant, applicable and connected to real teaching scenarios and practical examples. This result suggests a potential link between training quality and competence development, methodological approach, as well as the context of the training, while the thematic focus itself plays a secondary role.

6 Conclusion

The study confirms that the acquired level of basic digital competences among higher education teachers in the Republic of Croatia has reached a high level. However, while basic digital competences are widely present, competences related to student assessment and evaluation in digital environments, the implementation of individualized teaching approaches and the use of AI technologies are still underdeveloped. A lack of systematic and purposeful integration of technology into the educational process was identified, as well as a shortage of targeted professional development programs that address these specific needs. It can be concluded that factors such as resource availability, institutional support and the quality and scope of professional development are more relevant for acquiring digital competences than demographic characteristics such as gender, age or professional field of practice. The positive role of international training experiences stands out, as does the pedagogical relevance of selected digital technologies. In conclusion, the results confirm the need for professional development programs focused on implementing contemporary evaluation models, responding to real pedagogical needs, as well as being supported by institutional resources.

The scientific contribution of this research is reflected in several areas. Unlike previous studies that emphasized differences in digital competences based

on gender or academic discipline, this study shows that such differences are no longer statistically significant. Furthermore, this is the first study in a national context to empirically confirm a strong connection between participation in international professional development programs and higher levels of acquired digital competence. A specific scientific contribution lies in highlighting the low level of digital competence in the area of student assessment and evaluation, as well as the presence of regional disparities, which provides a basis for the development of targeted national education policies. Given the observed adjustment in teachers' self-assessment of their digital competences at the beginning and end of the study, the value of the DigCompEdu model is further confirmed, both in identifying actual competence levels and in supporting teachers' metacognitive reflection.

Since the study was conducted within a specific time frame, a longer data collection period could potentially result in a higher response rate and more general results. Additionally, as this research focuses on the dimensions of the DigCompEdu model, future studies may consider analysing and comparing data using other instruments. The use of qualitative research methods, such as interviews or focus groups, is also recommended for a deeper understanding of teachers' experiences and attitudes.

Reference

Basilotta-Gómez-Pablos, V., Matarranz, M., Casado-Aranda, L. A., & Otto, A. (2022). Teachers' digital competencies in higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education, 19*(8). https://doi.org/10.1186/s41239-021-00312-8

Day, C. (1999). *Developing teachers: The challenges of lifelong learning*. Falmer Press

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development*. Learning Policy Institute.

Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). *Teacher Technology Change: How Knowledge, Confidence, Beliefs, and Culture Intersect*. Journal of Research on Technology in Education, 42(3), 255–284. DOI:10.1080/15391523.2010.10782551

Erstad, O., & Voogt, J. (2018). The twenty-first century curriculum: Issues and challenges. In J. Voogt et al. (Eds.), *Second Handbook of Information Technology in Primary and Secondary Education* (pp. 19–36). Springer.

European Commission. (2018). *Digital Education Action Plan (2018)*. Publications Office of the European Union. https://education.ec.europa.eu/focus-topics/digital-education/action-plan

European Commission. (2020). Digital Education Action Plan 2021–2027: Resetting education and training for the digital age. Publications Office of the European Union. https://education.ec.europa.eu/hr/focus-topics/digital-education/action-plan

European Commission. (2020). *2nd survey of schools: ICT in education*. Publications Office of the European Union. https://data.europa.eu/doi/10.2759/23401

European Commission. (2023). *Check-in survey on higher education*. Data collection instrument. https://ec.europa.eu/eusurvey/runner/CheckIn_HE_v2023_HR?startQuiz=true&surveylanguage=EN

European Parliament. (2000). Lisbon strategy. https://www.europarl.europa.eu/meetdocs/2009_2014/documents/empl/dv/lisbonstrategybn_/lisbonstrategybn_en.pdf

Ferrari, A. (2013). *DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe*. Publications Office of the European Union. https://doi.org/10.2788/52966

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign. ISBN: 978-1794

Howard, S. K., & Tondeur, J. (2023). Higher education teachers' digital competencies for a blended future: Cultural and regional perspectives. *Educational Technology Research and Development, 71*(1), 1–6.293700

Iqbal, J. (2024). Role of information and communication technology (ICT) in education. *International Journal of Humanities, Social Science and Management*, *4*(4), 617–625

Ilomäki, L., Kantosalo, A., & Lakkala, M. (2011). What is digital competence? *Linked portal of European Schoolnet*.

Instefjord, E. J., & Munthe, E. (2017). Preparing teachers for digital competence: A study of strategies and actions in teacher education. *Teaching and Teacher Education, 67*, 1–10. https://doi.org/10.1016/j.tate.2017.05.016

Janssen, D., Stoyanov, S., Ferrari, A., Punie, Y., & Pannekeet, K. (2013). Experts' views on digital competence: Commonalities and differences. *European Journal of Education, 48*(4), 514–535.

Koehler, M. J., Mishra, P., & Cain, W. (2014). *What is Technological Pedagogical Content Knowledge (TPACK)?*. Contemporary Issues in Technology and Teacher Education, 14(1), 1–13.

Krumsvik, R. J. (2014). Teacher educators' digital competence. *Scandinavian Journal of Educational Research, 58*(3), 269–280. https://doi.org/10.1080/00313831.2012.726273

Jurčić, M. (2014). Kompetentnost nastavnika – pedagoške i didaktičke dimenzije. *Pedagogijska istraživanja, 11*(1), 77–91. https://hrcak.srce.hr/139572

Kučina Softić, S. (2020). *Digitalne kompetencije nastavnika za primjenu e-učenja u visokom obrazovanju*. Doktorska disertacija. Sveučilište u Zagrebu, Filozofski fakultet.

Laurillard, D. (2012). *Teaching as a Design Science: Building Pedagogical Patterns for Learning and Technology*. Routledge.

Luić, L., & Rončević, A. (2023). *Analiza recentnih rezultata samoprocjene digitalnih kompetencija visokoškolskih nastavnika u Hrvatskoj*.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2013). *The Effectiveness of Online and Blended Learning: A Meta-Analysis of the Empirical Literature*. Teachers College Record, 115(3), 1–47.

Müller, M., & Aleksa Varga, M. (2020). Digital competences of teachers and associates at higher educational institutions in the Republic of Croatia. *Informatologia*, *53*(1–2), 10–18. https://doi.org/10.32914/i.53.1-2.2

Redecker, C. (2017). *European Framework for the Digital Competence of Educators: DigCompEdu*. Publications Office of the European Union. https://doi.org/10.2760/159770

Revuelta-Domínguez, F.-I., Guerra-Antequera, J., González-Pérez, A., Pedrera-Rodríguez, M.-I., & González-Fernández, A. (2022). Digital teaching competence: A systematic review. *Sustainability, 14*(11), 6428. https://doi.org/10.3390/su14116428

Tomczyk, Ł., & Fedeli, L. (2021). Digital literacy among teachers – Mapping theoretical frameworks: TPACK, DigCompEdu, UNESCO, NETS-T, DigiLit Leicester. In *Proceedings of the 38th International Business Information Management Association (IBIMA), Innovation Management and Sustainable Economic Development in the Era of Global Pandemic* (pp. 244–252). ISBN 978-0-9998551-7-1

OECD. (2019). *TALIS 2018 Results (Volume I): Teachers and school leaders as lifelong learners*. OECD Publishing.

OECD. (2021). *Teachers and leaders in vocational education and training*. OECD Publishing.

OECD. (2021). *21st-century readers: Developing literacy skills in a digital world*. OECD Publishing.

Palvia, S., Aeron, P., Gupta, P., Mahapatra, D., Parida, R., Rosner, R., & Sindhi, S. (2018). Online education: Worldwide status, challenges, trends, and implications. *Journal of Global Information Technology Management, 21*(4), 233–241. https://doi.org/10.1080/1097198X.2018.1542262

Sánchez Prieto, J., Trujillo Torres, J. M., Gómez García, M., & Gómez García, G. (2020). Gender and digital teaching competence in dual vocational education and training. *Education Sciences, 10*(3), 78. https://doi.org/10.3390/educsci10030084

Siddiq, F., Scherer, R., & Tondeur, J. (2016). Teachers' emphasis on developing students' digital information and communication skills (TEDDICS): A new construct in 21st-century education. *Computers & Education, 92–93*, 1–14. https://doi.org/10.1016/j.compedu.2015.10.006

Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. P. (2020). Formative assessment, teacher knowledge, and the impact on student achievement. *Assessment in Education: Principles, Policy & Practice, 27*(6), 527–544.

Suzer, E., & Koc, M. (2024). Teachers' digital competency level according to various variables: A study based on the European DigCompEdu framework in a large Turkish city. *Education and Information Technologies, 29*, 22057–22083. https://link.springer.com/article/10.1007/s10639-024-12711-1

Trevisan, O., Christensen, R., Drossel, K., Friesen, S., Forkosh-Baruch, A., & Phillip, M. (2024). Drivers of digital realities for ongoing teacher professional learning. *Technology, Knowledge and Learning, 29*, 1851–1868. https://doi.org/10.1007/s10758-024-09771-0

Vizek-Vidović, V. (2005). *Cjeloživotno obrazovanje učitelja i nastavnika: Višestruke perspektive*. Institut za društvena istraživanja.

Voogt, J., Knezek, G., Cox, M., Knezek, D., & ten Brummelhuis, A. (2011). *Under which conditions does ICT have a positive effect on teaching and learning? A call to action*. Journal of Computer Assisted Learning, 31(5), 465–471. DOI:10.1111/j.1365-2729.2011.00453.

Learning path definition and usage in the context of learning analytics: literature review

Valentina Kirinić, Mihaela Bosak
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
{valentina.kirinic, mlaljek}@foi.unizg.hr

Abstract. *This paper reviews literature on learning paths (LP) within learning analytics, aiming to clarify LP definitions, contexts, and research methods. From 50 studies, 16 were analysed. Findings show that the term "learning path" is rarely precisely defined and usually refer to sequences of learning resources, activities, or challenges. Purposes of LPs usage range from guiding student learning, predicting outcomes to supporting adaptive interventions. Data sources are often learning management system logs, with statistical methods like logistic regression and discriminant analysis commonly applied. In the paper is highlighted the need for clearer conceptualization and further research into LP definitions and applications.*

Keywords. Learning path, learning analytics, definition

1 Introduction

Understanding and guiding student's behaviour when acquiring knowledge and skills is an important aspect of teaching/learning process, oriented to effective achievement of learning outcomes.

Learning paths – LPs (also learning pathway, journey, roadmap, trajectory) is a term which may be defined by different elements, used in different contexts emphasizing different purposes.

For example, according to Wu (2025) „an effective learning path can guide students to master the necessary knowledge at an appropriate pace, thereby improving learning efficiency and outcomes."

On the Digital Skills and Jobs Platform (n.d.) (learning) paths are defined, in the context of "the new feature designed to enrich the learning experience for users on the platform", as "a set of learning content, training opportunities, skills resources or skills publications that correspond to a specific learning purpose and are structured in a guided way for the user to embark on a small learning journey".

In the ECTS Users' Guide (2015), the term learning pathway is used and defined as "a route taken by a learner allowing him/her to build knowledge progressively and acquire the desired set of competences. The learning pathway may be 'signposted' through institution guidance and regulations (including the recognition of prior learning and experience) and different learning pathways may lead to the award of the same qualification. In essence the concept of a 'learning pathway' emphasises the choice of the student in reaching the desired educational goals."

Reason to do scoping literature review on the topic of „learning path" was to clarify the key concept/definition and context of its use in the literature.

2 Research questions

Scoping literature review of studies/papers dealing with learning paths is based on the following research questions:

- RQ1: What are the approaches to definition of learning path(s) and context of its usage?
- RQ2: What are dominant research worldviews, types of research and research methods used in relevant research dealing with learning path(s)?

3 Methodology

In order to explore and systematize important aspect of defining learning path for area of education, learning and learning analytics, related scoping literature review has been done.

The following steps in scoping literature review have been followed:

- Identification of the research questions
- Identification of relevant studies
- Selection of studies to be included in the review
- Data extraction
- Summarizing and reporting the results

As a starting point, we precisely **defined the research questions**, which had been outlined in the previous section. The questions were defined to

provide the base of understanding the term learning path, needed to serve the goal of the project TRUELA dealing with the identification of learning patterns of more and less successful students.

Steps of **identification of relevant studies** and selection of studies to be included in the review are presented by the Fig. 1.

Identification of relevant studies was done in 19th of February, 2025 by retrieving records from the Web of Science (WoS) collection based on the query:

((((ALL=(learning path*)) OR ALL=(learning trajector*")) AND ALL=("learning analytics")) AND ALL=("education")) AND ALL=("student*")

The query resulted in 50 records and all 50 articles/studies were downloaded and all relevant information were recorded.

The full set of 50 articles/studies found in WoS were published between years 2013 and 2025. 32 were published in journals, while 18 in conference proceedings

Selection of studies to be included in the review, i.e. eligibility check of the 50 articles/studies relevance to the topic of the literature review, has been done by 4 researchers in two rounds:

- firstly, based on the article abstract 26 articles were excluded, and
- then based on the full-text article insight additional 8 articles were excluded.

Eligibility check of 50 papers resulted in 16 papers to be analysed.

Analysis/**data extraction** was done by the same 4 researchers checking eligibility of the articles/studies.

Besides some common elements/information about articles to be included (such as the type of the article journal/conference, publication year,…), for each research question elements of analysis have been defined as follows:

- RQ1: What are the approaches to definition of learning path(s) and context of its usage?

  To answer the first research question in each article/study, Definition(s) of learning path were checked as well as its (Definition) elements (as a sequence of *something* – for example activities, tasks…). Furthermore, Learning path usage context, Use/application of learning path, Data used, and Methods used were identified (listed in the Table 1).

- RQ2: What are dominant research worldviews, types of research and research methods used in relevant research dealing with learning path(s)?

  To answer the second research question in each article/study, Research paradigm, Research approach, Research methodology, and Research methods were checked and identified (listed in the Table 2).

The steps of summarizing and reporting the results are incorporated in the following section of the paper.

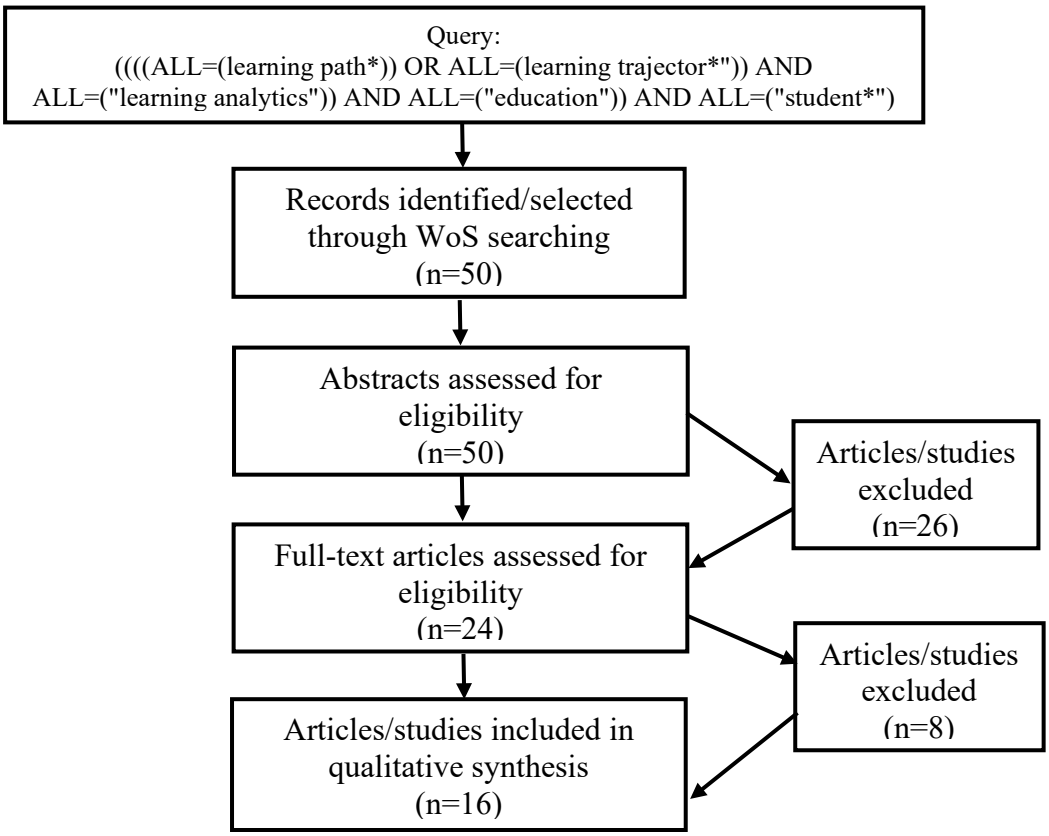


Figure 1. Flow diagram of the identification and selection of relevant studies to be included in the review

4 Results

After two rounds of eligibility checking, the set of 50 articles/studies was downsized to 16 published between years 2014 and 2025, 10 published in journals and 6 in conference proceedings.

Studies presented in the articles were done in 11 countries (Brazil, Bulgaria, Canada, Chile, China, India, South Korea, Spain, Turkey, UK, USA), mostly in higher education (1 primary and 1 in secondary education) and in the fields related to computer/ICT/STEM and engineering.

Students were the research participants in the analysed studies, with the exception of one case where both students and teachers were involved.

4.1 Results of the literature review to answer the first research question on approaches to definition of learning path(s) and context of its usage

To answer the first research question (**RQ1 What are the approaches to definition of learning path(s) and context of its usage?**), the sources of the definition of the term "learning path" were analysed.

Definition(s) of learning path:

In 2 articles/studies out of 16, author(s) definition (no quotation/citation/reference) was used. In 9 (out of 16) articles/studies implicit definitions could be concluded/recognized, while in 3 (out of 16) articles/studies there was no definition of the term "learning path" used.

Some authors use **own definitions**, as follows:

- "A learning path is the linear list of LOs, organized based on their knowledge relation" (**Raj & Renumol, 2024**)
- "The results from the social network analysis provide the learning path during a course." (**Choi & Cho, 2020**).

In two articles authors use/quote definition provided by other authors, from other sources:

- "Learning paths can be defined by various factors, such as their preferred learning styles (according Felder & Silverman 1988). Learning styles could be influenced by how students perceive information (sensing or intuitive), acquire information (visual or verbal), organize information (inductive or deductive), process information (active or reflective), and understand information (sequential or global)" (**Govindarajan, Kumar & Kinshuk, 2016**)
- "At course level, LPs are defined as "a sequence of learning tasks or activities which are designated to assist the student in improving their knowledge or skill in the particular subject" (according Yang, Li & Lau, 2010)" (**Martínez-Carrascal, Munoz-Gama & Sancho-Vinuesa, 2023**).

In 9 articles/studies **no explicit, just implicit definition** could be identified**:**

- Creating optimal learning path in order to achieve higher students' motivation and thus learning outcomes and efficiency; based to different learning styles models identified by using dedicated psychological questionnaires, educational data mining (EDM) (Bayesian Networks (BN) and Case-Based Reasoning (CBR) (**Kurilovas, 2019**)
- "monitoring learners involves observing their interactions within the *(learning)* system, which can result in learning paths (Ramos et al., 2021). These paths can be analysed from multiple perspectives, and the success of learning strategies and tactics is influenced by several factors (Wang, 2021). According to López-Pernas et al. (2021), learning strategies are defined as any thoughts, behaviours, beliefs, or emotions that facilitate the acquisition, understanding, or subsequent transfer of new knowledge and skills. Closely related, learning tactics refer to cognitive routines-actions that learners take to achieve a learning goal or perform a task; in other words, they are the methods a learner uses to learn something. Data-driven analysis methods enable the identification of learning patterns and associations between digital traces (learning paths) and learning outcomes, which can be 'mapped back' to the planned learning activities. This contextualization highlights the need to develop new approaches for interpreting log data to understand better students' learning interactions (Wang, 2021). (**Real & Pimentel, 2025).**
- Implied as a structured sequence of learning activities shaped by instructional design: "two learning paths to our students, traditional and novel, the second one with the aim of increasing the motivation and the engagement of the students and improving the learning results." (**Sousa-Vieira at al., 2023**)
- Implicitly, individualized sequence of strategies and activities based on behavioural analysis and prediction. (**Ouyang at al., 2023**)
- "As a result, the teacher's assistance is critical in correcting his learning path, assisting him, directing him to the appropriate path, identifying and resolving learning hurdles for him, and encouraging him to continue on his learning path and **attain his intended goals**. "This investigation illuminates the value of learning analytics provided by the learning analytics dashboard in AEs (*adaptive environments*), how it aids educators in determining when to intervene with students to alter their learning paths" (**Abouelenein, Selim & Aldosemani, 2025**).
- "**Students' primary choice of learning activity were characterized as 3 learning paths**: i) No use of video lectures (i.e., the course book readers) ii) Below average use of video lectures iii) Above

average use of video lectures." Reading the course book and mix of both (**Angrave at al., 2020**)

- "This work presents the study of **curricular trajectories** as processes (i.e., sequence of events) using process mining techniques. Specifically, the Backpack Process Model (BPPM) is defined as a novel model to unveil student trajectories, not by the courses that they take, but according to the courses that they have failed and have yet to pass" (**Salazar-Fernandez at al., 2021**).
- "Smart learning could mean customized learning that optimizes learning pathways, engages learners in positive interactions, and guides instruction in a goal-oriented fashion." (**Kumar at al., 2014**)
- No explicit definition given/used. In the context of Gamification and Digital Game-Based Learning (DGBL): "From learning perspective, challenges can be mapped to learning goals, levels to learning path, points and feedback to positive reinforcement, leaderboards to learning analytics." (**Iliev, 2018, p. 10656**).

There are no identified definitions used in 3 articles/studies ((Wu, Guo & Zhu, 2023), (Cooper, Ferguson & Wolff, 2016) (Baneres, 2016)).

More information on the elements/information collected are presented in the Table 1.

4.2 Results of the literature review to answer the second research question on dominant research worldviews, types of research and research methods used in relevant research dealing with learning path(s)

The second goal in the focus of the research and literature review is to answer the question formulated as types of research and research methods used in relevant research dealing with learning path(s).

With regard to research paradigms, the most common are post-positivist in 6 out of 16 (37.50%) and pragmatic - in 5 out of 16 (31.25%) articles/studies.

Dominant research approach is quantitative – in 9 out of 16 (56.25%) articles/studies, followed by mixed - in 5 out of 16 (31.25%) articles/studies and qualitative – in 1 out of 16 (6.25%) articles/studies reviewed.

The most used research methodologies are case study research – in 6 out of 16 (37.50%) and experimental research – in 5 out of 16 (31.25%) articles/studies. Research methodologies of quasi-experiment research, empirical study, action research, systematic literature review and design research are used in one article each – in 1 out of 16 (6.25%) articles/studies reviewed.

Research methods are diverse and presented in the last column of Table 2, which contains the results of the literature review related to RQ2.

5 Conclusions with future research

Based on the review of the literature results, the following conclusions can be drawn:

- Definitions of the term "learning path" are rarely exactly given or clarified or clearly indicated in the articles /studies.
- Different elements are used and/or emphasized when defining the term: as a sequence of objects/resources, learning tasks or activities, and as challenges in blended learning environment (gamification in situated learning).
- Definition and/or term "learning path" are used in the context of: a course, a part of the study program (set of courses) and/or in a particular field/group of courses, and task.
- Use/application/purpose of the "learning path" varies from to guide student learning and to predict learning success/failure, to compare the critical learning paths of two groups of students (students with or without disabilities), to construct a method of data tracking and recording in interactive learning environment, to propose a learning path recommendation approach (model) focused on knowledge building and learning performance analysis, to illuminate the value of learning analytics in AEs (adaptive environments) and how it aids educators in determining when to intervene with students to alter their learning paths.
- Data used are mostly collected from dashboard, video, LMS, task or multimodal logs, learning objects (meta)data, psychological or preference questionnaires.
- Methods used to analyse data differ from clustering, behavioral analytics, statistical comparisons, prediction models, Bayesian networks, Case-Based Reasoning, educational data mining (EDM), process mining, etc.
- Regarding research on the topic, dominant research paradigms are post-positivist and pragmatic, research approach is quantitative, research methodologies are case study research and experimental research, encompassing diverse research methods.

Based on the results, it is evident that the related field would benefit from a clearer conceptualization of the term "learning path," particularly in light of its varied and often ambiguous usage across studies. Despite the increasing relevance of learning paths in the context of learning analytics and instructional design, the absence of consistent definitions and theoretical grounding limits comparability among studies.

Table 1. RQ1 related results of the literature review

	Definition(s) of learning path:	Learning path as a sequence of:	Learning path used in the context of:	Use/application	Data used	Methods used
(Abouelenein, Selim & Aldosemani, 2025)	implicit	activities influenced by adaptive interventions	science teacher education (pre-service teachers)	analyze how interventions change self-regulation behavior	dashboard logs, self-regulation indicators	clustering, learning analytics, questionnaires
(Angrave at al., 2020) -	implicit	video segments accessed by learners	system programming course	analyze how learning paths affect learning outcomes and engagement	video logs, search events, student grades	behavioral analytics, statistical comparisons
(Baneres, 2016) -	no definition	learning tasks and exercises	course in digital systems design	adaptive support based on predicted success	performance data from tutoring systems	prediction models, intelligent tutoring systems
(Choi & Cho, 2020)	own	learning tasks and interactions	Educational psychometrics, statistics courses	diagnose learning progression and provide feedback	task logs, social network structures	SNA, Bayesian networks, ECD framework
(Cooper, Ferguson & Wolff, 2016)	no definition	learning activities	programs or courses for disabled students	compare the critical learning paths of disabled students and nondisabled students and to identify barriers to accessibility	completion rates, LMS data, accessibility indicators	comparative analysis
(Govindarajan, Kumar & Kinshuk, 2016)	quoted	learning activities	Programming course	predict optimal paths based on competence and meta-competence	competence scores, learning object data	clustering, learning analytics
(Iliev, 2018)	implicit	challenges and learning tasks	challenges and learning tasks	dynamic adaptation of tasks, gamification, and instructor support	multimodal logs (gesture, text, audio, video)	real-time tracking, dynamic assessment
(Kumar at al., 2014)	implicit	learning tasks	programming course	personalized feedback, competence analysis	trace data from coding activities	competence analytics

Table 1. RQ1 related results of the literature review (cont.)

	Definition(s) of learning path:	Learning path as a sequence of:	Learning path used in the context of:	Use/application	Data used	Methods
(Kurilovas, 2019)	implicit	learning activities and content tailored to learning styles	learning paths in courses, adapted for individuals	increase motivation and learning outcomes by matching LPs to learning styles	Psychological questionnaires, learning styles inventories	Bayesian networks, Case-Based Reasoning, educational data mining (EDM)
(Martínez-Carrascal, Munoz-Gama & Sancho-Vinuesa, 2023)	quoted	learning activities	mathematics course	assess effectiveness of recommended LPs through process mining	LMS logs	Process mining, log skeletons
(Ouyang at al., 2023)	implicit	learning resources and tasks	engineering courses	predict at-risk students and establish personalized learning pathways	Student interaction data, collaborative learning data	deep learning, learning analytics
(Raj & Renumol, 2024)	own	learning objects	computer science and engineering courses	recommend adaptive LPs based on learner profiles and real-time performance	learner logs, knowledge graphs, LO metadata	ontology, real-time analytics
(Real & Pimentel, 2025)	implicit	learning tasks	course	identify patterns, adjust course design	LMS logs	process mining
(Salazar-Fernandez at al., 2021)	implicit	courses, with special focus on failed and retaken courses	curricular analytics at the engineering study program level	inform curricular changes	student enrolment and grade data	process mining, directly-follows graphs
(Sousa-Vieira at al., 2023)	implicit	learning tasks	ICT course	compare traditional and gamified approaches, analyze impact on engagement and outcomes, predict learning success/failure	LMS logs, forum participation, task completion	descriptive statistics, correlation analysis
(Wu, Guo & Zhu, 2023)	no definition	micro-operations within 3D design software	3D design course	Evaluate spatial thinking and design skills	xAPI logs from GeekCAD	learning analytics, xAPI tracking

Table 2. RQ2 related results of the literature review

	Research paradigm	Research approach	Research methodology	Research methods
(Abouelenein, Selim & Aldosemani, 2025)	pragmatic	quantitative	experimental research	descriptive-longitudinal method
(Angrave at al., 2020)	positivist	quantitative	experimental research	log analysis, statistical tests (a Mann-Whitney U test)
(Baneres, 2016)	post-positivist	quantitative	case study	predictive modelling
(Choi & Cho, 2020)	post-positivist	quantitative	empirical study	data mining method using social network analysis, an analytic method using a Bayesian network
(Cooper, Ferguson & Wolff, 2016)	pragmatic	mixed	case study	(survey and historical module) data analysis, comparative analysis, statistical method (odds ratios)
(Govindarajan, Kumar & Kinshuk, 2016)	positivist	quantitative	experimental research	simulation, cluster analysis/clustering
(Iliev, 2018)	constructivist	mixed	action research	questionnaire
(Kumar at al., 2014)	post-positivist	quantitative	case study	descriptive statistics methods (visualization)
(Kurilovas, 2019)	pragmatic	qualitative	systematic literature review	literature review
(Martínez-Carrascal, Munoz-Gama & Sancho-Vinuesa, 2023)	constructivist	quantitative	experimental research	process mining
(Ouyang at al., 2023)	pragmatic	mixed	quasi-experiment research	predictive modelling, social network analysis (SNA), quantitative content analysis (QCA), thematic analysis method,
(Raj & Renumol, 2024)	post-positivist	mixed	experimental research	ontology-based method (of/for modelling), simulation (of data)
(Real & Pimentel, 2025)	positivist	quantitative	design research	educational process mining (EPM),
(Salazar-Fernandez at al., 2021)	post-positivist	quantitative	case study	process mining
(Sousa-Vieira at al., 2023)	post-positivist	mixed	case study	social network analysis (SNA), machine learning/deep learning (ML/DL)
(Wu, Guo & Zhu, 2023)	pragmatic	quantitative	case study	validation

Acknowledgments

This work has been supported by the Croatian Science Foundation under the project IP-2022-10-2854 (*Trustworthy Learning Analytics and Artificial Intelligence for Sound Learning Design* – TRUELA).

References

EU. (n.d.). Learning Path. Digital Skills and Jobs Platform, Retrieved from https://digital-skills-jobs.europa.eu/en/learning-path?page=0

European Commission. (2015). ECTS users' guide 2015. Publications Office of the European Union. Retrieved from

https://education.ec.europa.eu/sites/default/files/document-library-docs/ects-users-guide_en.pdf

Abouelenein, Y. A. M., Selim, S. A. S., & Aldosemani, T. I. (2025). Impact of an adaptive environment based on learning analytics on pre-service science teacher behavior and self-regulation. *Smart Learning Environments*, *12*(1), 8.

Angrave, L., Zhang, Z., Henricks, G., & Mahipal, C. (2020, February). Who benefits? positive learner outcomes from behavioral analytics of online lecture video viewing using classtranscribe. In *Proceedings of the 51st acm technical symposium on computer science education* (pp. 1193-1199).

Baneres, D. (2016, October). Towards a particular prediction system to evaluate student's success. In *International Conference on P2P, Parallel, Grid, Cloud and Internet Computing* (pp. 935-945). Cham: Springer International Publishing.

Choi, Y., & Cho, Y. I. (2020). Learning analytics using social network analysis and Bayesian network analysis in sustainable computer-based formative assessment system. *Sustainability*, *12*(19), 7950.

Cooper, M., Ferguson, R., & Wolff, A. (2016, April). What can analytics contribute to accessibility in e-learning systems and to disabled students' learning?. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 99-103).

Felder, R.M. & Silverman, L.K. (1988). Learning and Teaching Styles in Engineering Education*, J. Eng. Education,* vol. 78, no. 7, pp. 674-681

Govindarajan, K., Kumar, V. S. & Kinshuk (2016, December). Dynamic learning path prediction—A learning analytics solution. In *2016 IEEE eighth international conference on technology for education (T4E)* (pp. 188-193). IEEE.

Iliev, T. (2018). Tracking and Adaptation of Situated Blended Learning with Challenge Unicorder™. In *EDULEARN18 Proceedings* (pp. 10652-10662). IATED.

Kumar, V., Boulanger, D., Seanosky, J., Kinshuk, Panneerselvam, K., & Somasundaram, T. S. (2014). Competence analytics. *Journal of Computers in Education*, *1*, 251-270.

Kurilovas, E. (2019). Different methods to identify students preferences to learning styles and learning paths. In *ICERI2019 Proceedings* (pp. 2914-2922). IATED.

Martínez-Carrascal, J. A., Munoz-Gama, J., & Sancho-Vinuesa, T. (2023). Evaluation of recommended learning paths using process mining and log skeletons: Conceptualization and insight into an online mathematics course. *IEEE Transactions on Learning Technologies*, *17*, 555-568.

Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, *20*(1), 4.

Raj, N. S., & Renumol, V. G. (2024). An improved adaptive learning path recommendation model driven by real-time learning analytics. *Journal of Computers in Education*, *11*(1), 121-148.

Ramos, D., Ramos, I., Gasparini, I., & Teixeira de Oliveira, E. (2021). A new learning path model for e-learning systems. (Vol. 19, pp. 34–54).

Real, E., & Pimentel, E. (2025). An Educational Process Mining Model on Students' Paths Data from Virtual Learning Environments. *Technology, Knowledge and Learning*, 1-26.

Salazar-Fernandez, J. P., Munoz-Gama, J., Maldonado-Mahauad, J., Bustamante, D., & Sepúlveda, M. (2021). Backpack Process Model (BPPM): A Process Mining Approach for Curricular Analytics. *Appl. Sci*, *11*, 4265.

Sousa-Vieira, M. E., López-Ardao, J. C., Fernández-Veiga, M., & Rodríguez-Rubio, R. F. (2023). Study of the impact of social learning and gamification methodologies on learning results in higher education. *Computer Applications in Engineering Education*, *31*(1), 131-153.

Wang, F.H. (2021). Interpreting log data through the lens of learning design: Second-order predictors and their relations with learning outcomes in fipped classrooms. (Vol. 168).

Wu, Y. (2025). Learning Path Analysis of Optimizing Educational Data Mining Based on Genetic Algorithm. In *SHS Web of Conferences* (Vol. 213, p. 02031). EDP Sciences.

Wu, Y., Guo, S., & Zhu, L. (2023). Design and implementation of data collection mechanism for 3D design course based on xAPI standard. In *Cross Reality (XR) and Immersive Learning Environments (ILEs) in Education* (pp. 64-81). Routledge.

Yang, F., Li, F. W. B., & Lau, R. W. H. (2010). An open model for learning path construction," in Advances in Web-Based Learning – ICWL 2010 *Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*. Berlin, Germany: Springer, 2010, doi: 10.1007/978-3-642-17407-0_33.

Strategic Planning and Decision Making

Marta Mas-Machuca, Frederic Marimon, Anna Akhmedova
Generative AI and Workplace Productivity: A Qualitative Study in Spain

Marta Mas-Machuca, Alba Manresa, Jasmina Berbegal-Mirabent
The Impact of Environmental Training on Employee Motivation and Corporate Sustainability in Spanish Companies

Predrag Trpeski, Violeta Cvetkoska
Bridging Systems, Shaping Futures: A Multilevel Strategy for Curriculum Transformation for Innovation and Resilience in Higher Education

Marko Bohanec
Using Large Language Models for Creating, Modifying and Interpreting Decision Tables in Multi-Criteria Method DEX

Sabina Rako, Nina Begičević Ređep, Barbara Šlibar, Diana Šimić, Bart Rienties
Students' Cognitive Engagement in Blended Learning: Observations from Event-contingent Self-report and Log Data

Enrique Mu, Anna Florek-Paszkowska, Santiago Luis Risco-Martinez
Exploring the Impact of Digital Infrastructure Challenges on Learning Outcomes and Student Satisfaction

Sanja Kelemen, Nemanja Mišljenović, Goran Knežević, Stjepan Babić
Application of Multi-Objective Optimization for Short-Term Operation Planning of an Energy Community

Krisztián Attila Bakon, Tibor Holczinger
Reactive Scheduling under Due-Date and Intermediate Storage Constraints: A Conceptual Multi-Objective Approach

Teresa Monllau, Jaime Lacueva, EnricoTesti, Silvia Sordi, Nikola Kadoić, Ivan Malbašić
A conceptual model to assess and develop digitalisation in the social economy's people care sector: the case of Prato (Italy), Mataro (Spain) and Varaždin (Croatia)

Nikica Daraboš, Violeta Vidašek Hainš, Nevena Ladić
Gender equality in the fields of information and communication technology in healthcare

Jin Eun Yoo, Nam-Hwa Kang, Suna Ryu, Jun-ki Lee, Youngsun Kwak, Hyeong Gwan Kim
Developing an LLM-specific Coding Framework for Prediction and Feedback Modeling of Science Teachers' Instructional Expertise

Generati/ve AI and Workplace Productivity: A Qualitative Study in Spain

Marta Mas-Machuca, Anna Akhmedova, Frederic Marimon
International University of Catalonia
Faculty of Economics and Social Science
Immaculada, 22 08017 Barcelona, Spain
{mmas, aakhmedova, fmarimon}@uic.es

Abstract. *The growing use of generative Artificial Intelligence (AI) in the workplace is undeniable, with varying views on its adoption and impact. This article explores how Generative AI influences employee performance and the role of hedonics and job satisfaction. A qualitative analysis of 16 Spanish managers and employees (April 2024) reveals that AI boosts productivity by enhancing satisfaction, saving time, and improving output quality. Positive employee experiences and the integration of AI into daily routines are essential for maximizing productivity. These findings provide valuable theoretical and practical insights for improving employee satisfaction and performance in the workplace.*

Keywords. Generative Artificial Intelligence, employee productivity, efficiency, hedonics, Job satisfaction, future of work.

1 Introduction

The adoption of generative AI (Gen AI) is expected to boost economic growth (Lal, 2023), yet its implications remain uncertain, as Gen AI is viewed both as a facilitator and inhibitor of human potential (Mäkelä & Stephany, 2024). Its integration in the workplace is reshaping industries, particularly in functional areas, with controversial impacts on employees. Negative effects (Hu & Min, 2023) include reduced demand for low-skilled jobs (Xie et al., 2021), increased job insecurity (Wang et al., 2019), burnout (Möhlmann et al., 2021; Hessari & Daneshmandi, 2024), and a negative competitive climate (Li et al., 2019). However, positive outcomes also exist, such as higher demand for skilled labor (Li et al., 2021), new job opportunities (Acemoglu & Restrepo, 2020), improved work performance (Zhou & Wang, 2021), increased job autonomy (Xiaomei et al., 2021) and creativity (Jia et al., 2024, De Smet et al., 2024).

Technology adoption also affects employees' psychological state, intrinsic motivation (Watchravesringkan et al., 2010, Cerasoli et al., 2014; Bernardo et al., 2005), and well-being (Xu et al., 2023). For instance, smartphones increase stress through constant email checks and reduced face-to-face interactions (Przybylski & Weinstein, 2013). While technology design can impact emotions (Nguyen et al. 2025), focusing solely on engagement doesn't necessarily lead to greater job satisfaction (Marimon et al., 2025).

This study aims to assess how employees' use of Gen AI affects productivity (Keltanen, 2024) focusing on the role of hedonics and job satisfaction. The key research question is: How does Gen AI integration impact employee performance across various organizational settings? Is improving productivity purely a technical matter, or does it also involve personal aspects? Additionally, we explore how Gen AI influences different aspects of Employee Performance (EP), including hedonics and job satisfaction.

This research addresses two gaps in the literature: the emotional and psychological effects of disruptive technologies like Gen AI on employees, a topic not yet fully explored (Wisskirchen et al., 2017), and how the implementation of Gen AI will shape day-to-day work. Understanding its impact on employee productivity in the context of Gen AI's emergence is essential (Callari & Puppione, 2025).

This study extends existing literature by applying the Self-Determination Theory framework to Gen AI use at work. It focuses on factors influencing employee satisfaction—autonomy, competence, and relatedness—and examines how early Gen AI adoption impacts productivity. This theoretical approach helps explore new ways to adapt human-machine interactions, offering insights into the future of work.

2 Background research

The Technology Acceptance Models, such as the Technology Acceptance Model (Davis, 1989), are widely used to predict technology adoption behaviors (Alturas, 2021). These models identify key factors that influence the use of technology, including Gen AI, which affects employee motivation, job satisfaction,

and performance. In this context, the Self-Determination Theory (SDT) explains the relationship between intrinsic motivation (hedonics), job satisfaction, and performance. SDT highlights the importance of autonomy, competence, and relatedness in promoting high-quality performance and well-being (Ryan & Deci, 2000a). When employees feel autonomous, competent, and connected, they are more motivated and engaged, leading to higher job satisfaction and productivity. Gen AI can enhance these aspects by enabling more efficient task completion, facilitating collaboration, and improving task relevance and utility.

SDT (Deci, 1976, Ryan & Deci, 2017, 2000b) offers a valuable framework for understanding how motivational factors influence Gen AI adoption and performance. It also emphasizes three essential psychological needs: 1) autonomy, or the sense of control over one's actions; 2) competence, reflecting the ability to perform effectively; and 3) relatedness, which involves feeling connected to others. Gen AI can enhance these needs by improving human-computer interactions and fostering a sense of connection at work.

Previous studies have shown that Gen AI enhances efficiency and productivity, especially for developers and knowledge workers (Sira, 2023). For instance, using tools like ChatGPT has improved productivity in professional writing (Noy & Zhang, 2023) and consultancy tasks (Dell'Acqua et al., 2023). Gen AI's ability to automate tasks, provide personalized suggestions, and improve decision-making can boost workplace efficiency (Aishwarya, 2023). Also, it is relevant to identify the internal flow from risk identification to knowledge integration and strategic outlook, highlighting generative AI's role in enhancing decision-making and competitive advantage (Söllner et al., 2025). However, issues like quality control, job displacement, and privacy concerns must be addressed (Wach et al., 2023) to fully capitalize on Gen AI's potential.

This research builds on existing literature by exploring the impact of Gen AI on employee performance, specifically through the lens of SDT. It suggests that the usability of Gen AI positively affects productivity, hedonics, and job satisfaction, as employees are more likely to engage with tools that are enjoyable and effective. Additionally, creating a psychologically safe environment where employees feel supported in using AI tools fosters a culture of experimentation and innovation, ultimately enhancing performance (Coetzee, 2019).

The role of hedonic motivation—pleasure derived from technology use—has been shown to influence technology acceptance and performance (Brown & Venkatesh, 2005). If employees enjoy using Gen AI tools, they are more likely to adopt and utilize them, enhancing productivity. Furthermore, intrinsic motivation and satisfaction, shaped by factors like autonomy and competence, have been linked to better job performance (Deci et al., 2017). This study also explores how hedonic motivation (Tamilmani et al., 2019) impacts job satisfaction and ultimately employee performance, proposing a positive relationship between these factors.

Fostering a work environment that promotes job satisfaction alongside the use of Gen AI is essential for enhancing productivity and well-being. By reducing repetitive tasks and focusing on higher-value work, employees can experience greater satisfaction and contribute more effectively to organizational success (Nazareno & Schiff, 2021; Bankins et al., 2023).

Based on the literature and the proposed theoretical framework, the following propositions are presented to explore the impact of Generative AI (Gen AI) on employee performance, job satisfaction, and motivation. These propositions aim to highlight the key relationships between technology usability, hedonics, and employee outcomes in the workplace.

Proposition 1: The usability of Generative AI (Gen AI) positively impacts employee performance (productivity).

Proposition 2: The usability of Generative AI (Gen AI) positively impacts employees' hedonic motivation.

Proposition 3: Hedonic motivation positively impacts job satisfaction among employees.

Proposition 4: Job satisfaction positively impacts employee performance (productivity).

Proposition 5: The usability of Generative AI (Gen AI) positively impacts job satisfaction among employees.

Proposition 6: Hedonic motivation positively impacts employee performance (productivity).

3 Method (focus groups)

Four focus groups were conducted, involving 16 professionals, with each session consisting of 2 to 6 participants. The sessions lasted between 50 and 60 minutes and took place during April 2024. At least two researchers were present in every session. The focus group discussions were guided by semi-structured questions (see guidelines for more details). These questions were designed based on the literature review.

Participants were selected from Spain using a purposive, non-probabilistic sampling approach. To minimize the influence of cultural factors on participants' perceptions and evaluations (Lobuono et al., 2016), individuals with similar cultural backgrounds were chosen, but with variation in characteristics such as age, gender, area of expertise, and openness to adopting new technologies. Table 1 provides an overview of the focus group composition. The groups continued until the saturation point was reached (Malterud et al., 2016), ensuring comprehensive exploration of the research questions.

To prepare for the sessions, the participants received the questions a few days in advance, and the sessions were held online via Google Meet. All

Table 1. Profile of participants of Focus Groups.

Nº	FC	Gender	Age	Frequency of use	Industry or area of expertise	Main use (personal/professional)
1	FC1	F	36-45	Twice per month	Headhunting	Professional
2	FC1	M	36-45	Daily	Logistics	Personal (The company prohibits employees from using Generative AI)
3	FC1	F	21-25	Intensively every day	Audit firm (Big Four)	Professional
4	FC1	F	21-25	Intensively every day	Consultancy (Big Four)	Professional
5	FC2	F	>55	Once per month	Retail	Personal
6	FC2	M	>55	Weekly	Engineering	Both
7	FC2	M	21-25	Daily	Legal & taxation	Professional
8	FC2	F	46-55	Weekly	Logistics	Personal (low level of digitalization in the organization)
9	FC2	F	36-45	Intensively and advanced daily use	Entrepreneur & software developer	Professional
10	FC3	M	46-55	Weekly	IT industry	Both
11	FC3	M	26-35	Daily	Technological start-up	Both
12	FC3	F	21-25	Twice per month	Digital banking	Professional (but with caution)
13	FC3	M	36-45	Twice per month	International currency transfer	Personal
14	FC3	M	>55	Once per month	Consultant & lecturer	Professional
15	FC4	M	26-35	Daily	Dealer manager system	Both
16	FC4	M	26-35	Daily	Digital marketing	Both

discussions were recorded and transcribed, and the content was analyzed using ATLAS.TI 24. The session durations were as follows: the first focus group (FC1) lasted 53:05 minutes, the second (FC2) lasted 1 hour and 4 minutes the third (FC3) lasted 58:48 minutes, and the fourth (FC4) lasted 46:34 minutes. The analysis and coding were initially done by one researcher, and then cross-validated by the rest of the research team to ensure reliability.

4 Results

The adoption of GenAI is still in its early stages in Spain, but its advantages are already apparent to both employees and managers, although these benefits vary across industries. In some sectors, such as consultancy, auditing, and consumer goods, companies actively promote GenAI usage and provide corresponding training. However, in other sectors, like shipping and logistics, the level of digitalization is low, and GenAI is often viewed as a threat, with its use even prohibited in some cases.

In terms of autonomy, competence, and interpersonal relations, participants consistently reported increased autonomy through the use of GenAI. They noted the ability to independently find answers, reducing their reliance on colleagues or supervisors, which in turn enhanced their knowledge acquisition. One participant shared, "I feel more autonomous and rely less on others for certain information." Another said, "It makes me more independent because I can access knowledge outside my area of expertise using this tool." These findings suggest that GenAI fosters greater individual empowerment.

Competence, however, yielded mixed opinions. Some participants felt GenAI inspired them, helping them uncover new insights and improve their understanding of work-related topics. One participant remarked, "Yes, it enhances my competence because it sometimes reveals things I wasn't aware of." Others appreciated how the tool supported their professional competence, stating, "With GPT chat, it's a new way of learning. The quick, structured responses help me access a lot of information." On the other hand, some participants viewed it merely as a time-saving tool,

depending on the task at hand. As for interpersonal relations, most agreed that GenAI had little effect. One participant noted, "Maintaining conversation and connectivity with teams remains unchanged. AI doesn't disrupt how we interact at work; it's just another way of communicating."

There was unanimous agreement that these tools are easy to use and generally helpful for work (Noy & Zhang, 2023). However, they have not yet been integrated into companies' information systems, meaning their use is not standardized and remains up to individual employees. Some companies encourage the use of GenAI and provide training, while others lack internal policies regarding its adoption. Regardless of frequency or intensity of use, employees tend to trust certain tools over others, depending on their personal experience.

Curiosity is a key factor driving employees to use these tools. One participant shared, "Sometimes it's just curiosity to see what it can do." Additionally, the 'wow effect' often plays a role in initial use, as expressed by one participant: "It's amazing, isn't it? The amount of information it processes and how it can answer anything. If I had this during my university days…" This enthusiasm tends to diminish when GenAI is used in personal contexts, though it is deeply embedded in daily life, often unnoticed.

The majority of participants expressed enjoyment in using these tools (van der Heijden, 2004). One participant commented, "I enjoy seeing how it works, but it also depends on which tools I'm using." In terms of enjoyment, one participant brought up an interesting point about the gamification aspect of using GenAI. She said, "Humans love to play; that's why tools like ChatGPT are so popular—people enjoy using them for entertainment purposes, like generating images or videos." However, despite this enjoyment, participants viewed these tools primarily as work tools, stating, "I see it as a tool, just like any other. It helps me, but I don't use it beyond what's necessary."

When it comes to productivity, most participants confirmed that GenAI does enhance their work performance (Vrinda & Jacob, 2015, Kessler et al., 2020). This was particularly evident in two areas: time savings and improved output quality. Time-saving was frequently mentioned in all four focus groups, with participants noting that it allowed them to focus on more valuable, less repetitive tasks. One participant stated, "Saving time on simpler tasks lets me dedicate more time to things that generate more value." Another added, "It frees up time for more interesting tasks, like research or reflecting on solutions."

Regarding the quality of work, most participants agreed that using GenAI improved their output, especially in tasks like text translation, writing, spelling, code review, and report preparation. One participant mentioned, "The work-time ratio is unmatched. The quality is higher, and I spent less time on it." Another participant added, "For me, it's about ensuring quality when I don't know how to do something—GenAI helps me with tasks I've never done before." In some cases, GenAI expanded employees' capacity and allowed them to broaden the scope of their work.

Moreover, GenAI is viewed as a tool for personal growth and continuous learning. One participant explained, "It helps me organize and learn, and ultimately, it saves me time. Time savings alone improve my well-being." Despite being in its early stages, GenAI's positive impact on employee performance and satisfaction was clear.

Lastly, participants recognized that these technologies will shape the future of work (Valeriya et al., 2024). By streamlining manual and administrative tasks, saving time, and offering new opportunities, GenAI will change how work is done and adapt as new technologies continue to evolve.

5 Discussion and limitations

The results obtained from the focus groups confirm most of the proposed propositions. First, Proposition 1, which states that the usability of Gen AI positively impacts employee performance (productivity), is confirmed, as participants highlighted improvements in work efficiency and the quality of outcomes thanks to the tool. Proposition 2, suggesting that the usability of Gen AI positively impacts hedonic motivation, is also supported, as participants expressed enjoyment in using Gen AI, especially when exploring its capabilities. Proposition 3, linking hedonic motivation with job satisfaction, is reinforced, as the enjoyment derived from using Gen AI translated into higher job satisfaction. Regarding Proposition 4, which posits a positive relationship between job satisfaction and employee performance, participants confirmed that satisfaction from autonomy and the improvement in work quality contributed to increased productivity. Proposition 5, which asserts that the usability of Gen AI positively impacts job satisfaction, is also validated, as participants reported greater satisfaction in their work when using this technology. Finally, Proposition 6, suggesting that hedonic motivation positively impacts performance, was corroborated, as the motivation derived from using Gen AI was reflected in increased employee performance in tasks related to the technology.

The Fig. 1 presents a conceptual framework illustrating the influence of generative artificial intelligence (Gen AI) usability on hedonic motivation, job satisfaction, and employee performance. The model posits that higher usability of Gen AI tools enhances users' hedonic motivation—defined as the intrinsic enjoyment derived from using the technology—which subsequently contributes to increased job satisfaction. Both hedonic motivation and job satisfaction are proposed to positively impact employee performance. Furthermore, the model suggests a bidirectional relationship between Gen AI

usability and job satisfaction, indicating that employees who are more satisfied with their work may also perceive the technology as more usable. Overall, the framework underscores that the effective integration of generative AI in the workplace yields not only operational improvements but also psychological and motivational benefits that can foster higher employee performance.

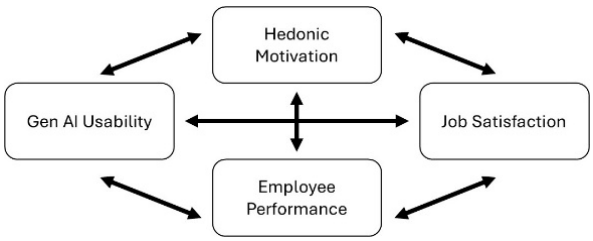


Figure 1. Relations among the analysed variables

This study contributes both theoretically and practically by advancing our understanding of how the usability of Gen AI influences employee performance, job satisfaction, and motivation in the workplace. It also provides valuable insights into the integration of these technologies within organizational settings. However, the study has some limitations, including its focus on a specific geographic region (Spain) and its reliance on qualitative data from a relatively small sample of professionals, which may not fully capture the diverse impacts of Gen AI across different industries and cultures. Future research could expand the sample size and scope to include other regions and sectors, explore the long-term effects of Gen AI on employee motivation and performance, and examine the role of organizational culture in shaping the adoption and effectiveness of these technologies.

References

Acemoglu, D., & Restrepo, P. (2020). The wrong kind of AI? Artificial intelligence and the future of labour demand. Cambridge Journal of Regions, Economy and Society, 13(1), 25-35.

Alturas, B. (2021). Models of acceptance and use of technology research trends: Literature review and exploratory bibliometric study. Recent Advances in Technology Acceptance Models and Theories, 13-28.

Aishwarya Gowda; Hui-Kai; Kuo Wen-Kai. Unleashing Potential of Employees through Artificial Intelligence. In 2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS). IEEE, 2023. p. 204-206.

Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2023). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. Journal of Organizational Behavior.

Bernardo, M., Marimon, F., & del Mar Alonso-Almeida, M. (2012). Functional quality and hedonic quality: A study of the dimensions of e-service quality in online travel agencies. Information & management, 49(7-8), 342-347.

Brown, S. A., & Venkatesh, V. (2005). Model of adoption of technology in households: A baseline model test and extension incorporating household life cycle. MIS quarterly, 399-426.

Callari, T. C., & Puppione, L. (2025). Can generative artificial intelligence productivity tools support workplace learning? A qualitative study on employee perceptions in a multinational corporation. Journal of Workplace Learning, 37(3), 266-283.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: a 40-year meta-analysis. Psychological bulletin, 140(4), 980.

Coetzee, M. (2019). Organisational climate conditions of psychological safety as thriving mechanism in digital workspaces. In M. Coetzee (Ed.), Thriving in digital workspaces (pp. 311–327). Springer. https://doi. org/10.1007/978-3-030-24463-7_16

De Smet, A., Durth, S., Hancock, B., Mugayar-Baldocchi, M., & Reich, A. (2024). The human side of generative AI: Creating a path to productivity. McKinsey & Company.

Deci, E. L. (1976). The hidden costs of rewards. Organizational Dynamics, 4, 61–72. doi:10.1016/0090-2616(76)90036-X

Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-determination theory in work organizations: The state of a science. Annual review of organizational psychology and organizational behavior, 4, 19-43.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., … & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Technology & Operations Mgt. Unit Working Paper, (24-013).

Davis, Fred D. 1989. “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.” MIS Quarterly: Management Information Systems 13 (3): 319–39. https://doi.org/10.2307/249008.

Hessari, H., Bai, A., & Daneshmandi, F. (2024). Generative AI: boosting adaptability and reducing

workplace overload. Journal of Computer Information Systems, 1-14.

Hu, Y., & Min, H. K. (2023). The dark side of artificial intelligence in service: The "watching-eye" effect and privacy concerns. International Journal of Hospitality Management, 110, 103437.

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. Academy of Management Journal, 67(1), 5-32.

Kessler, S. R., Lucianetti, L., Pindek, S., Zhu, Z., & Spector, P. E. (2020). Job satisfaction and firm performance: Can employees' job satisfaction change the trajectory of a firm's performance?. Journal of Applied Social Psychology, 50(10), 563-572.

Keltanen, R. (2024). Generative AI in the Workplace: Effects on Employee Efficiency and Performance.

Krueger, R. A. (2014). Focus groups: A practical guide for applied research. Sage publications.

Lal, P. (2023). Artificial Intelligence and Employment: Issues and Challenges. AI and Emotional Intelligence for Modern Business Management, 213-225.

Li, J. J., Bonn, M. A., & Ye, B. H. (2019). Hotel employee's artificial intelligence and robotics awareness and its impact on turnover intention: The moderating roles of perceived organizational support and competitive psychological climate. Tourism Management, 73, 172-181.

Li, S., Wang, H., & Wang, S. (2021). Research on the factors that influence the labor structure of the manufacturing industry in the context of Artificial Intelligence. Management Review, 33(3), 307-314.

Lobuono, R., de Sevilha Gosling, M., Goncalves, C. A., & Medeiros, S. A. (2016). Relationship between dimensions of experience, satisfaction, word-of-mouth and intention to returning: The perception of cultural event participants. Podium Sport, Leisure and Tourism Review, 5(2), 15-37.

Mäkelä, E., & Stephany, F. (2024). Complement or substitute? How AI increases the demand for human skills. arXiv preprint arXiv:2412.19754.

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: guided by information power. Qualitative health research, 26(13), 1753-1760.

Marimon, F., Mas-Machuca, M., & Akhmedova, A. (2025). Trusting in generative AI: Catalyst for employee performance and engagement in the workplace. International Journal of Human–Computer Interaction, 41(11), 7076-7091.

Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. W. (2021). Algorithmic management of work on online labor platforms: when matching meets control. MIS quarterly, 45(4).

Nazareno, L., & Schiff, D. S. (2021). The impact of automation and artificial intelligence on worker well-being. Technology in Society, 67, 101679. https://doi.org/10.1016/j.techsoc.2021.101679

Nguyen, M., Trinh, N., Mehrotra, A., & Basahel, S. (2025). Generative AI in the workplace: how employee experiences influence work outcomes?. Journal of Enterprise Information Management.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187-192.

Przybylski, A. K., & Weinstein, N. (2013). Can you connect with me now? How the presence of mobile communication technology influences face-to-face conversation quality. Journal of Social and Personal Relationships, 30(2), 237–246. doi: 10.1177/0265407512453827

Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: classic definitions and new directions. Contemporary Educational Psychology, 25(1), 54–67. doi: 10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2000b). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. American Psychologist, 55(1), 68–78. doi: 10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. Annual Review of Psychology, 52(1), 141–166. doi: 10.1146/annurev.psych.52.1.141

Ryan, R. M., & Deci, E. L. (2017). Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York, NY: Guilford Press.

Sira, M. (2023) Generative AI Takes Centre Stage: Revolutionizing Productivity and Reshaping Industries. System Safety: Human-Technical Facility-Environment, 5(1), 57-65.

Söllner, M., Arnold, T., Benlian, A., Bretschneider, U., Knight, C., Ohly, S., ... & Wendt, D. (2025). ChatGPT and Beyond: Exploring the Responsible Use of Generative AI in the Workplace: M. Söllner et al. Business & Information Systems Engineering, 67(2), 289-303.

Tamilmani, K., Rana, N. P., Prakasam, N., & Dwivedi, Y. K. (2019). The battle of Brain vs. Heart: A literature review and meta-analysis of "hedonic motivation" use in UTAUT2. International Journal of Information Management, 46, 222-235.

Valeriya, G., John, V., Singla, A., Devi, J. Y., & Kumar, K. (2024). AI-Powered Super-Workers: An

Experiment in Workforce Productivity and Satisfaction. In BIO Web of Conferences (Vol. 86, p. 01065). EDP Sciences.

Van der Heijden, H. (2004). User acceptance of hedonic information systems. MIS quarterly, 695-704.

Vrinda, N. N., & Jacob, N. A. (2015). The impact of job satisfaction on job performance. International Journal in Commerce, IT & Social Sciences, 2(2), 27-37.

Wach, K., Duong, C. D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., ... & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. Entrepreneurial Business and Economics Review, 11(2), 7-30.

Wang, C., Zhou, W., & Zhao, S. (2019). Study on the relationship between massive-scale utilization of industrial robots and job insecurity: the moderating effects of employee's career ability. Business Management Journal, 41(04), 111-126.

Watchravesringkan, K., Nelson Hodges, N., & Kim, Y. H. (2010). Exploring consumers' adoption of highly technological fashion products: The role of extrinsic and intrinsic motivational factors. Journal of Fashion Marketing and Management: An International Journal, 14(2), 263-281.

Wisskirchen, G., Biacabe, B. T., Bormann, U., Muntz, A., Niehaus, G., Soler, G. J., & von Brauchitsch, B. (2017). Artificial intelligence and robotics and their impact on the workplace. IBA Global Employment Institute, 11(5), 49-67.

Xiaomei, Z., Sen, W., & Qin, H. (2021). Impact of skill requirements on employees' thriving at work: From the perspective of artificial intelligence embedding. Foreign Economics & Management, 43(11), 15-25.

Xie, M., Ding, L., Xia, Y., Guo, J., Pan, J., & Wang, H. (2021). Does artificial intelligence affect the pattern of skill demand? Evidence from Chinese manufacturing firms. Economic Modelling, 96, 295-309.

Xu, G., Xue, M., & Zhao, J. (2023). The Relationship of Artificial Intelligence Opportunity Perception and Employee Workplace Well-Being: A Moderated Mediation Model. International Journal of Environmental Research and Public Health, 20(3), 1974.

Zhou, W., & Wang, C. (2021). Impact of the Utilization of Robots on Job Performance—Study on Employees in Middle and Low Skilled Positions. China Soft Sci, 4, 106-119.

The Impact of Environmental Training on Employee Motivation and Corporate Sustainability in Spanish Companies

Marta Mas-Machuca, Alba Manresa
International University of Catalonia
Faculty of Economics and Social Science
Immaculada, 22 08017 Barcelona, Spain
{mmas, amanresa}@uic.es

Jasmina Berbegal-Mirabent
Universitat Politècnica de Catalunya
Polytechnic School of Engineering of Vilanova i la Geltrú (EPSEVG)
Av. de Víctor Balaguer, 1, 08800 Vilanova i la Geltrú, Barcelona
Department of Business Organization
jasmina.berbegal@upc.edu

Abstract*. This study examines how green training, employee motivation, and sustainability integration influence corporate environmental behavior. Specifically, it aims to understand how green training impacts employee engagement in eco-friendly actions and how this affects the overall sustainability efforts within the organization. Using qualitative insights from five focus groups, the findings reveal that green training significantly boosts employee participation in sustainable practices, with green motivation playing a critical mediating role. Most participants reported a stronger commitment to environmentally responsible behavior following the training, although individual efforts alone are insufficient. This research contributes to the understanding of green training's role in fostering sustainable organizational practices and emphasizes the need to integrate it into corporate sustainability strategies while aligning employee motivation with environmental goals.*

Keywords: Green training, green motivation, sustainable performance, company green behaviour, business behaviour, green education.

1 Introduction

Sustainability has become a priority for companies as they face increasing pressure to address environmental challenges and adopt more responsible practices (García-Cruz et al., 2024). Companies are integrating sustainability into their strategies to comply with regulations, enhance their competitive edge, and improve their long-term viability.

Green Human Resource Management (GHRM) has emerged as a key strategy to incorporate sustainable practices into workforce management (De Stefano et al., 2018). Among GHRM practices, green training stands out as an effective and cost-efficient method to foster environmental responsibility, aligning employees' skills with environmental management goals (Teixeira et al., 2016). Green training helps overcome barriers to environmental protection, promotes acceptance of sustainable practices, and integrates environmental objectives into daily operations (Yusoff et al., 2020). This practice significantly enhances awareness, boosts engagement in sustainability efforts, and shifts companies toward greener behavior (Yafi et al., 2021). However, many organizations focus on employees' capabilities without addressing the need to motivate them to sustain a green work culture.

Research suggests that combining green training with motivation can strengthen an organization's commitment to sustainability, yet the role among these variables (green training, green motivation, integration to sustainability and Company green behaviour) needs further exploration. Existing studies have examined the link between green training and company behavior in various countries (Daily et al., 2007; Mishra & Rath, 2024), but these have overlooked the context of Spain. This gap in the literature has been covered in the present research, which aims to explore which green training practices influences company green behavior in Spanish SMEs, with a focus on role of green motivation (GM). Grounded in the resource-based view (RBV) and the ability-motivation-opportunity (AMO) framework, this research addresses this area that have raised many interests for academics and practitioners.

2 Theoretical foundations

This research examines the impact of green training on employee green motivation, organizational sustainability, and environmental behavior. Green training programs aim to integrate environmental values and develop competencies that align employees with the organization's sustainability objectives (Pham

et al., 2019). Companies with a strong environmental focus implement such programs to foster pro-environmental behaviors that contribute to organizational performance (Pham et al., 2020). Through green training, employees acquire the knowledge and skills necessary to address environmental challenges and make more sustainable decisions (Memon et al., 2022).

The integration of sustainability within organizations goes beyond the adoption of responsible practices; it also involves embedding these principles into organizational culture and strategic decision-making processes. This integration is crucial to ensure that environmental goals are aligned with the company's overall objectives, which, in turn, strengthens the organization's commitment to sustainability.

From the perspective of the Resource-Based View (RBV) theory, green training acts as a strategic resource that enhances human capital and improves environmental performance. According to this theory, an organization's ability to achieve a sustainable competitive advantage relies on managing valuable, inimitable, and non-substitutable resources, such as the competencies developed through green training (Barney, 2001; Arulrajah & Opatha, 2014). Various studies have shown that green training not only increases pro-environmental behaviors but also improves productivity in key areas such as waste reduction and energy conservation (Singh et al., 2020; Seman et al., 2020). However, it is important to note that training alone is insufficient to generate significant behavioral changes without solid organizational support and adequate motivation (Bissing-Olson et al., 2013).

One of the key aspects of green training is its ability to significantly enhance employee green motivation by instilling environmental responsibility and a clear sense of purpose (Yafi et al., 2021; Pinzone et al., 2019). However, the long-term impact of training on motivation is not fully understood, and further research is needed to explore the mechanisms through which green training influences employee motivation (Mohamad et al., 2023). In this regard, green behavior within organizations reflects how companies manage their natural resources and internal processes sustainably, which not only helps reduce costs but also improves environmental performance.

Green motivation, defined as the drive for employees to adopt environmentally responsible behaviors, is essential for the success of sustainability initiatives. According to the Ability-Motivation-Opportunity (AMO) theory, performance in green initiatives depends on three key factors: employees' ability, motivation, and the opportunities provided by the organization (Renwick et al., 2013). Effective green motivation systems, such as incentives, play a crucial role in improving participation in sustainability efforts and achieving positive organizational outcomes (Anwar et al., 2020). Thus, creating a motivating and supportive environment is essential for ensuring that employees engage with the company's environmental goals, yielding both organizational and environmental benefits.

3 Method

This study consisted of five focus groups (3 participants in the all of them and 2 in the last one, as detailed in Table 1). Participants received the interview protocol with 11 questions and a detailed description of the aim of this current investigation. Participants were divided into five focus groups to ensure a rich, multi-sectoral perspective on sustainability. The first group included strategic leaders from regulated and infrastructure-intensive sectors, offering high-level insights into complex, large-scale sustainability challenges. The second brought together experts in environmental and social governance, focusing on stakeholder engagement and the integration of sustainability into corporate practices. The third group, composed of professionals in operational excellence and compliance, contributed practical knowledge from industrial environments, emphasizing quality, efficiency, and environmental standards. The fourth group consisted of specialists in data-driven sustainability, sharing experiences on embedding sustainability into business strategies and decision-making processes in large organizations. Finally, the fifth group included academic representatives, providing an educational perspective on preparing future professionals and promoting social responsibility. This composition enabled focused and diverse discussions aligned with the study's objectives.

The focus groups were recorded and a transcription software was used, and it produced a total of 41,866 words. The focus group were conducted online during June-July 2024. ATLASTI 14.0 was used to analyze the data.

The coding of empirical data refers to the process of symbolically grouping information to facilitate analysis, identify patterns, categorize, and construct theories, following the approach proposed by Saldaña (2021). In this study, most codes were generated through a deductive approach, derived from interview scripts and pre-existing categories, using the "IA Intentional Coding" tool in Atlas.ti. Although this function initially identified relevant segments, a

The analytical strategy and the coding strategy are considered two interrelated components of a unified process for identifying thematic patterns within qualitative data. This study adopts a thematic analysis approach, understood as the identification and analysis of thematic patterns in qualitative data (Braun & Clarke, 2006). The objective of this analysis is to uncover themes related to sustainability training within organizations. This approach is particularly suitable as it enables a comprehensive examination of how sustainability training is conceptualized and

implemented, as well as its broader implications, grouped under the umbrella of sustainable practices. Such an analytical framework provides a solid foundation for in-depth exploration of both sustainability training and its influence on organizational practices. In alignment with the research objectives, the use of multiple methods is warranted (Saldaña, 2021). Accordingly, thematic analysis is complemented by frequency analysis to enhance the depth and rigor of the findings.

4 Results

The results suggest that while sustainability training is a key driver in fostering individual motivation, its impact on structural changes within the organization depends on additional factors, such as internal policies, leadership, and corporate strategies that facilitate the large-scale implementation of sustainable practices.

First, during the coding process, a total of 9 inductive codes related to the category "Education" were identified. This category refers to aspects related to sustainability training programs and their implementation, including decision-making processes and the role of executive management in determining how the training is conducted and how it should be conducted. It also involves employees' knowledge of these training programs. The absolute frequency along with the percentages is presented in Table 2.

Table 2. Green training categories

Code	#	%
Education	21	37.50
Executive Teams	9	16.07
Awareness	6	10.71
Orientation to action	6	10.71
Decision-making	5	8.93
New forms of educations	4	7.14
Good practices	2	3.57
Realism and objective data	2	3.57
Positive messages	1	1.79
TOTAL	**56**	**100.00**

Fig. 1 identifies three main areas related to green training: decision-making processes around sustainability training, the role of executive teams, and the content of the training programs. A recurring theme is the responsibility of management to implement training, with workers acknowledging their role in promoting it but recognizing that the final decision lies with the company's leadership. As one participant stated, "The company's leadership must take responsibility for integrating sustainability into the agenda" (2024, F2, 3:70). Additionally, there is a focus on the need for training content to evolve toward more modern, action-oriented approaches.

In the coding process, 11 inductive codes were found referring to the category "motivation strategies". This category is based on the motivations that lead companies to conduct sustainability training, whether due to the influence of their environment, the company's own beliefs, or economic benefits. Table 3 contains the absolute coding figures and their respective percentages.

Table 3. Green motivations categories

Code	#	%
Employee incentives	20	19.80
Employees' motivations	12	11.88
Education	11	10.89
Interacting with the business environment	11	10.89
Legal and normative obligations	11	10.89
Sustainability practices	11	10.89
Economic benefits	10	9.90
Certifications & audits (C2)	8	7.92
Pride of belonging	3	2.97
Social responsibility	3	2.97
Public administration support	1	0.99
TOTAL	**101**	**100.00**

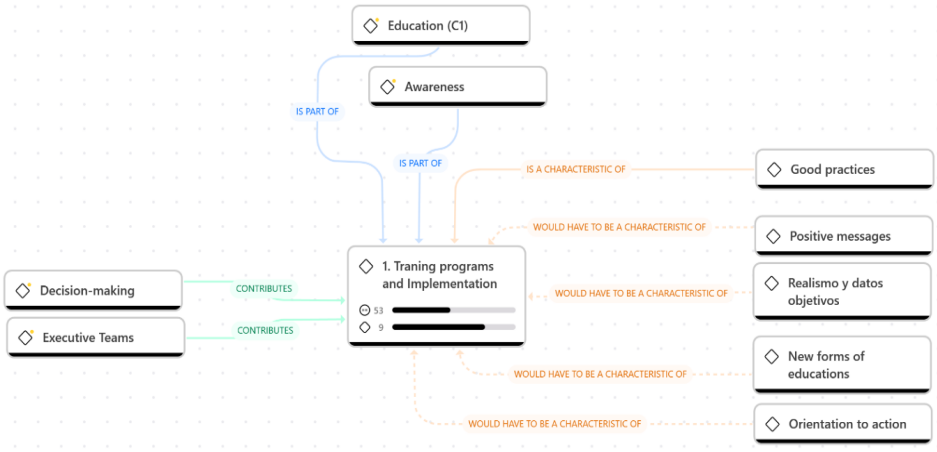


Figure 1. Green training relationships

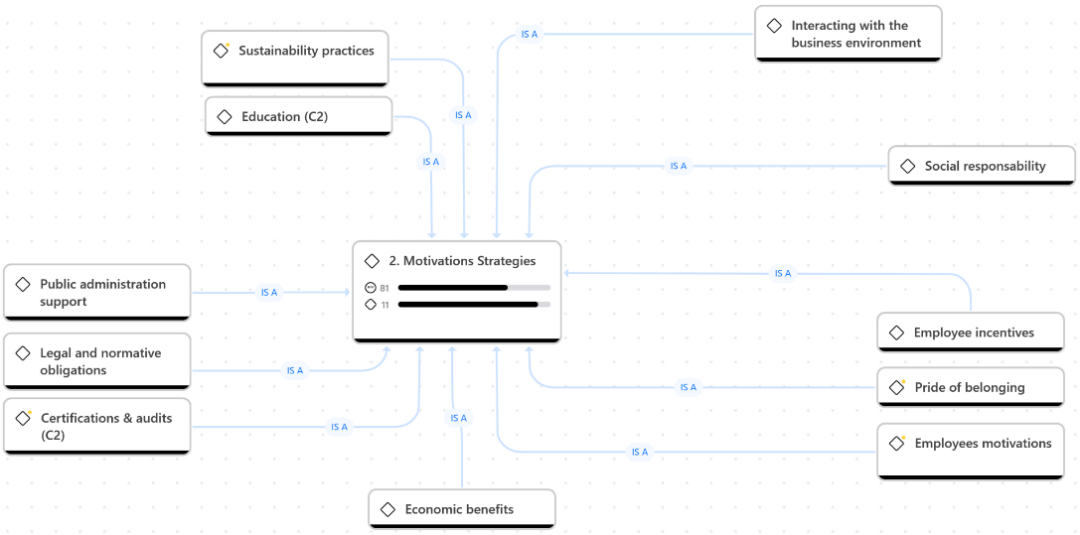


Figure 2. Company green behaviour relationships

This category (green motivation) identifies six main motivations areas related to motivation strategies (see Fig. 2): legal responsibilities, knowledge of sustainable practices, interaction with the business environment, social responsibility, employee sense of belonging, and economic benefits. Legal responsibilities are linked to public administration support, mandatory legal regulations, and certifications. As one participant stated, "Sustainability must be integrated into company objectives, whether motivated by regulations or market demands" (2024, F3, 4:16). Knowledge of sustainable practices enables companies to transform their operations, and interaction with the environment drives companies to differentiate from competitors. As noted, "Sustainability is collective; companies need to advance together" (2024, F3, 4:14). Additionally, sustainability is seen as a strategy for attracting talent, as young workers often align their values with sustainability (2024, F5, 7:28). Economic benefits are also a key motivator, as companies recognize that sustainability can lead to both environmental and financial gains, resulting in a "win-win" scenario (2024, F1, 2:89).

Finally, 7 codes were found inductively under the category "integration of sustainability" (Table 4). This category refers to how sustainability is materialized within the dynamics of the company. The most relevant topics include the business model, the alignment of sustainability within the company's dynamics, and the internal campaigns carried out as examples of the integration of sustainability in companies. The frequencies of occurrence in quotes and their percentages are shown below:

Table 4. Integration of sustainability categories

Code	#	%
Business model	31	22.63
Sustainability alignment	23	16.79
Internal campaigns	21	15.33
Integration of sustainability	17	12.41
Education	16	11.68
Strategy to integrate sustainability	16	11.68
Comprehension at all levels	13	9.49
TOTAL	**137**	**100.00**

This category highlights the strong connection between sustainable practices and sustainability training, which together represent the integration of sustainability into company operations (see Fig. 3). The business model is central to this integration, as sustainability adds both environmental value and business value: "Developing business through sustainability means not only respecting the environment, society, and workers, but also finding business value in it" (2024, F1, 2:76). Aligning company values with sustainability motivates employee retention, as one participant shared: "The company being aligned with your values and doing things for the environment is motivating for retention" (2024, F5, 7:30). Sustainability is also integrated into the company's goals, as seen in Nestlé's approach to embedding sustainability into its DNA (2024, F3, 4:31). Strategies for integrating sustainability, such as environmental practices and internal campaigns to promote sustainable values, are essential for creating a sustainable business culture (2024, F5, 7:23; 2024, F3, 4:17).

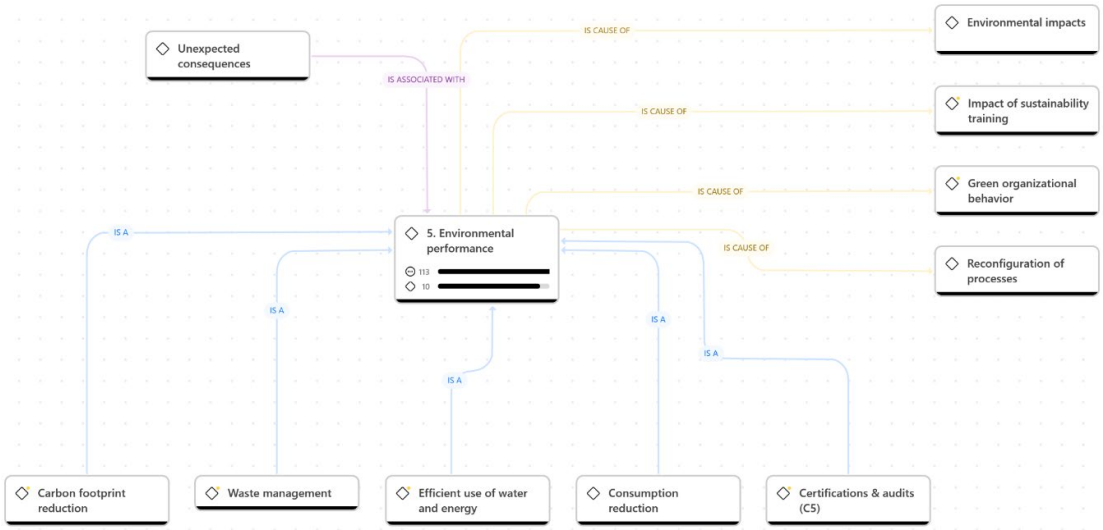


Figure 3. Integration to sustainability relationships

The last category refers to the sustainable practices carried out within the company, beyond training. As mentioned earlier, there is a very close link between sustainable practices and sustainability training. Therefore, this category reflects the practices implemented that allow for a sustainable transformation of companies, whether in their organizational structure or in their procedures. Table 5 contains the absolute frequency of the 10 codes found, along with their percentage representation within the category:

Table 5. Company green behaviour categories

Code	**#**	**%**
Efficient use of water and energy	33	20.50
Green organizational behaviour	29	18.01
Carbon footprint reduction	19	11.80
Reconfiguration of processes	19	11.80
Waste management	19	11.80
Consumption reduction	13	8.07
Impact of sustainability training	12	7.45
Certifications & audits	10	6.21
Environmental impacts	5	3.11
Unexpected consequences	2	1.24
TOTAL	**161**	**100**

A recurring theme has been the relationship between sustainable practices and sustainability training. In this regard as is shown in Fig. 4, the current category presents examples of sustainable practices implemented by companies. These include: reducing the carbon footprint, waste management, efficient use of water and energy, consumption reduction, and certifications and audits (understood as accreditations that differentiate companies). Sustainable practices, in turn, lead to favorable environmental impacts, influence sustainability training (creating a constant feedback loop), drive changes in organizational behaviour, and result in the reconfiguration of internal production processes.

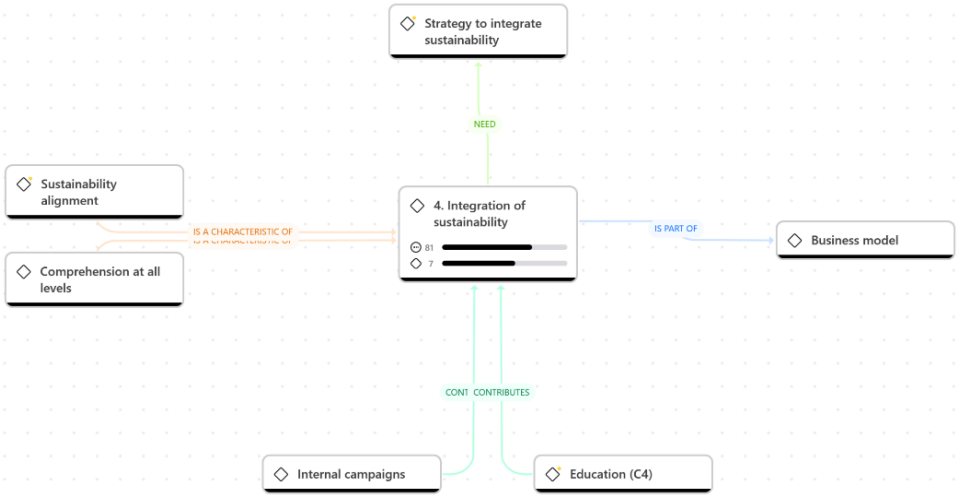


Figure 4. Integration to sustainability relationships

5 Discussion, contribution and limitations

The findings of this study underscore the critical role of sustainability training in driving sustainable practices within organizations. A strong relationship between training and sustainable practices emerges, with companies recognizing that integrating sustainability into both their operations and corporate culture is essential for long-term success. The integration of sustainability into the business model, as evidenced by the examples provided, reflects a growing commitment to sustainability that transcends environmental concerns and extends to economic and social dimensions.

One of the key insights is that sustainability training does not operate in isolation. Instead, it works as part of a broader system where sustainable practices, such as carbon footprint reduction, waste management, and resource efficiency, reinforce and drive the continuous improvement of training programs. The feedback loop between sustainable practices and training highlights the dynamic nature of sustainability efforts within organizations. As employees become more informed and motivated by sustainability training, they are more likely to implement environmentally responsible behaviors, which in turn, enhances the need for further training to keep up with evolving environmental standards.

Moreover, the study indicates that the alignment of sustainable values with organizational goals is a powerful motivator for employees. As organizations integrate sustainability into their strategic objectives, employees are more likely to see their own values reflected in the company's practices, leading to increased retention and job satisfaction. However, the integration of sustainability within organizational dynamics is not solely driven by internal factors; external pressures, such as legal regulations, market competition, and stakeholder expectations, also play a significant role in shaping the sustainability agenda.

The impact of sustainable practices on organizational behavior is noteworthy. Companies are not only making environmental strides but are also reconfiguring their internal processes, demonstrating that sustainability can lead to operational efficiencies. These changes are particularly evident in areas like waste management and energy use, where companies are implementing innovative solutions that align with broader sustainability goals.

Additionally, this study provides insights for managers to identify the most effective ways to promote sustainable environmental behave: for example, effective training strategies for sustainability include a focus on green skills development, such as energy efficiency practices, waste reduction, and environmental regulations. Motivational strategies like recognizing employees' contributions to sustainability efforts, offering incentives, and aligning sustainability goals with organizational values are also key. For SMEs with limited resources, implementing such programs can be achieved by utilizing cost-effective online training platforms, partnering with industry associations for shared resources, and encouraging employee-led initiatives that integrate sustainability into everyday practices without significant financial investment.

As show Fig. 5, the study suggests that the relationship between sustainability training and sustainable practices is integral to the success of corporate sustainability efforts. Companies that view sustainability as a strategic priority, align it with their core values, and continuously invest in training and internal processes are better positioned to achieve both environmental and business success.

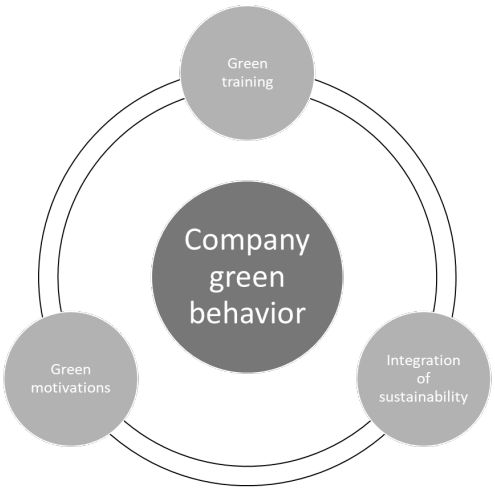


Figure 5. Relationships that emerge from the research

This study makes some theoretical. First, it strengthens the RBV perspective by showing how green training, as a firm-specific, valuable, and strategically important resource, enhances organizational human capital and develops environmental capabilities, positioning sustainability-oriented HR practices as key drivers of competitive advantage. Second, the study extends the AMO framework by empirically validating the mediating role of green motivation, introducing it as a distinct mechanism that complements employee competencies and drives behavioral outcomes, thus addressing a gap in existing models that have predominantly focused on ability.

A limitation of the study is its relatively small sample size and geographic focus, which may not represent the diversity of industries or regions. Expanding the sample to include more sectors and regions would provide a more comprehensive understanding of sustainability training's impact. Additionally, the study's cross-sectional nature limits its ability to capture long-term effects, and the reliance on self-reported data introduces potential biases. The limitations section should also address the generalizability of the findings, considering the possible response bias in the focus groups and the

specific contextual constraints of conducting the study in Spain. Future research should explore these limitations by comparing organizations with and without sustainability initiatives, investigating cultural and institutional differences, and focusing on sector-specific challenges to provide a more holistic view of sustainability practices across various contexts. Also, further research should explore the long-term impacts of sustainability training on organizational culture and the persistence of sustainable practices across different industries.

Acknowledgments

This project, Mesurar l'impacte positiu de les accions fetes per mitigar el canvi climàtic. Educació climàtica a la Universitat i des de la Universitat, was funded by AGAUR (Grant No. AGAUR 2023 CLIMA 00081) and coordinated by the Universitat Internacional de Catalunya.

References

Anwar, N., Mahmood, N. H., Yusliza, M. Y., Ramayah, T., Faezah, N. J., & Green, K.W. (2020). Human resource management for organisational citizenship behaviourtowards the environment and environmental performance on a university campus.Journal of Cleaner Production, 256, 120401

Arulrajah, A. A., Opatha, H. H. D. N. P., & Nawaratne, N. N. J. (2015). Green human resource management practices: A review. Sri Lankan Journal of human resource management, 5(1).

Barney, J., Wright, M., & Ketchen Jr, D. J. (2001). The resource-based view of the firm: Ten years after 1991. Journal of management, 27(6), 625-641.

Bissing-Olson, M. J., Iyer, A., Fielding, K. S., & Zacher, H. (2013). Relationships between daily affect and pro-environmental behavior at work: The moderating role of pro-environmental attitude. *Journal of Organizational Behavior*, *34*(2), 156-175.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

Daily, B. F., Bishop, J. W., & Steiner, R. (2007). The mediating role of EMS teamwork as it pertains to HR factors and perceived environmental performance. Journal of applied business research, 23(1), 95.

De Stefano, F., Bagdadli, S., & Camuffo, A. (2018). The HR role in corporate social responsibility and sustainability: A boundary-shifting literature review. Human Resource Management, 57(2), 549–566.

García-Cruz, J., Rincon-Roldan, F., Pasamar S. (2024). "When the stars align: The effect of institutional pressures on sustainable human resource management through organizational engagement" European Management Journal, *In Press.*

Memon, S., Rasli, A., Dahri, A., & Abas, I. H. (2022). Importance of Top Management Commitment to Organizational Citizenship Behaviour towards the Environment, Green Training and Environmental Performance in Pakistani Industries. Sustainability, 14(17), 11059

Mishra, S., & Rath, N. (2024). Impact of Green Human Resource Management Practices on Environmental Performance of Indian Banking Sector: An Empirical Study. International Journal of Sustainable Development & Planning, 19(4).

Mohamad, N. I., Abd Rahman, I., Hasan, H., & Sanusi, S. (2024). Unlocking proactivity at work: Exploring the mediating role of green motivation between managerial coaching in training programmes and employee green behaviour. Journal of Infrastructure, Policy and Development, 8(7).

Pham, N. T., Vo-Thanh, T., Shahbaz, M., Huynh, T. L. D., & Usman, M. (2020). Managing environmental challenges: Training as a solution to improve employee green performance. Journal of Environmental Management, 269, 110781.

Pinzone, M., Guerci, M., Lettieri, E., & Huisingh, D. (2019). Effects of 'green'training on pro-environmental behaviors and job satisfaction: Evidence from the Italian healthcare sector. *Journal of cleaner production*, *226*, 221-232.

Renwick, D. W. S., Redman, T., & Maguire, S. (2013). Green human resource management: A review and research agenda. International Journal of Management Review ,15(1), 1–14.

Saldaña, J. (2021). Coding techniques for quantitative and mixed data. *The Routledge reviewer's guide to mixed methods analysis*, 151-160.

Seman, N. A. A., Govindan, K., Mardani, A., Zakuan, N., Saman, M. Z. M., Hooker, R. E., & Ozkul, S. (2019). The mediating effect of green innovation on the relationship between green supply chain management and environmental performance. *Journal of cleaner production*, *229*, 115-127.

Singh, R. K., Berkvens, R., & Weyn, M. (2020, January). Energy efficient wireless communication for IoT enabled greenhouses. In *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)* (pp. 885-887). IEEE.

Teixeira, A. A., Jabbour, C. J. C., de Sousa Jabbour, A. B. L., Latan, H., & De Oliveira, J. H. C. (2016). Green training and green supply chain management: evidence from Brazilian firms. Journal of cleaner production, 116, 170-176.

Yafi, E., Tehseen, S., & Haider, S. A. (2021). Impact of green training on environmental performance through mediating role of competencies and motivation. Sustainability, 13(10), 5624.

Yusoff, Y. M., Nejati, M., Kee, D. M. H., & Amran, A. (2020). Linking green human resource management practices to environmental performance in hotel industry. Global Business Review, 21(3), 663-680.

Appendix

Appendix 1 1. Focus Group composition

Table 1. Focus Group composition

Focus group	Participant	Job position/Industry	Date	Duration	Nº words
1	JG	Director of the Pharmacy and Medicines at the Consortium of Health and Social Care of Catalonia/Pharma	10/6/2024	53:02	6,536
	JS	Consultant/ Energy consultancy (gas and electricity)			
	CP	Chief Business and Strategy Officer/Shipping and logistic industry			
2	RL	Environmental Products Production Specialist /Consumer Goods	11/6/2024	1:11:06	8,837
	ML	Consultant in Social Responsibility and Quality of Life for businesses, educators, and Public Administration/Consultancy			
	SG	CEO/ International project portfolio management & IT integration or separation during large M&A projects			
3	CC	Business Manager/ Testing, inspection, and certification industry	11/6/2024	1:11:12	9,204
	JL	CMO & CSO/Logistic			
	NM	Director of Occupational Risk Prevention and Environment/Manufacturing			
4	JS	Data & Information Director/Furniture company	17/7/2024	1:03:18	9,124
	AC	Project Manager/Facility Services,			
	PF	Sustainable Development & People Global Director | Promoting sustainability, social commitment and good governance/Paper Manufacturing (Large company)			
5	NG	Director of the Career Guidance and Employment Service and the Expert Degree in Sustainability Management and Social Responsibility /HEIs	17/7/2024	54:34	8,165
	MB	Architect lecturer/HEIs			
Total	14				41,866

Bridging Systems, Shaping Futures: A Theoretical Multilevel Framework for Curriculum Transformation in Higher Education

Predrag Trpeski, Violeta Cvetkoska
Ss. Cyril and Methodius University in Skopje, Faculty of Economic-Skopje
Blvd. Goce Delchev 9V, 1000 Skopje, North Macedonia
{predrag.trpeski,vcvetkoska}@eccf.ukim.edu.mk

Abstract. *Curricula are no longer just course maps, they are strategic blueprints for navigating uncertainty and driving societal progress. In an era defined by rapid digital change, climate volatility, and widening skills gaps, higher education institutions must evolve from static content providers into agile ecosystems of innovation and resilience. This paper proposes a theoretical, stakeholder-driven multilevel framework for curriculum transformation, grounded in complexity theory, systems thinking, and participatory design. Methodologically anchored in a design-based, mixed-methods approach, this conceptual framework emphasizes inclusive governance and adaptive learning but has not yet been implemented. The Faculty of Economics-Skopje at Ss. Cyril and Methodius University in Skopje, a pioneer candidate in Association to Advance Collegiate Schools of Business (AACSB) accreditation in North Macedonia, inspires this work and will serve as the initial institutional context for future empirical validation. The initial application phase will involve the development of tailored surveys and interviews to capture stakeholder perspectives. The study offers a strategic roadmap for universities, policymakers, and industry partners seeking to reimagine curricula as engines of sustainable, inclusive transformation.*

Keywords. curriculum transformation, innovation, resilience, higher education, systems thinking, multilevel strategy

1 Introduction

Higher education stands at a historic crossroads. Faced with intensifying global challenges, from pandemic-induced disruptions and digital acceleration to shifting labor markets and widening social inequities, universities must evolve beyond incremental reform to systemic reimagination. Traditional curriculum structures, often compartmentalized and discipline-centric, are increasingly inadequate for preparing graduates who must lead in an era of uncertainty, complexity, and constant change (Hassan et al., 2024; Nusche et al., 2024; Ross et al., 2022). This study responds to that urgency by proposing a theoretical, multilevel strategy for curriculum transformation rooted in innovation and resilience.

The impetus for this research arises from a growing consensus that higher education reform must transcend isolated pedagogical adjustments. Instead, transformation must be embedded across governance levels, driven by stakeholder collaboration, and informed by both qualitative insight and quantitative rigor (Law, 2022; Ritzen & Soete, 2011). Curriculum is no longer merely a vehicle for knowledge delivery; it is a strategic instrument for institutional renewal, talent development, and societal impact (Carayannis & Morawska-Jancelewicz, 2022).

This paper develops a conceptual framework for curriculum transformation, guided by a theoretical synthesis of complexity theory, systems thinking, socio-cultural and critical perspectives, and organizational change. Although the framework has not been piloted, it serves as a strategic design for future empirical testing. Methodologically, the study employs mixed-methods and participatory design. Structured surveys, semi-structured interviews, and stakeholder focus groups will be developed in a future phase to identify curriculum challenges, opportunities, and leverage points. Data visualization tools such as Power BI will support the modeling of institutional learning and engagement dynamics. It is important to emphasize that the research remains in a pre-implementation phase and is entirely conceptual at this stage.

The inspiration for this paper stems from the Faculty of Economics–Skopje's strategic momentum toward innovation, resilience, and international quality standards. Motivated by the institution's ongoing progress, particularly its commitment to curriculum modernization, digital integration, and global accreditation, the research team identified the need for a structured, evidence-based and scalable approach to curriculum transformation. The Faculty's proactive initiatives served as both a catalyst and a case model for developing the broader multilevel framework proposed in this study.

The framework is designed to be generalizable across higher education institutions but will undergo preliminary validation at the Faculty of Economics-Skopje at Ss. Cyril and Methodius University in Skopje. As a national leader in business and economics education, the faculty serves not as an isolated case, but as the emergent context where the proposed strategy was conceived and shaped. In the next phase, we plan to develop detailed instruments, including questionnaires and interview protocols tailored to capture stakeholder perspectives. These tools will first be applied within the Faculty to validate, refine, and test the assumptions embedded in the framework, before expanding its application to other institutions nationally and internationally.

The contribution of this study is fourfold. First, it introduces a robust, multilevel conceptual framework that integrates curricular innovation with governance alignment. Second, it demonstrates how a participatory, data-driven methodology can operationalize reform across institutional boundaries. Third, it highlights the role of HEIs as orchestrators of transformation, not just responders to change but proactive agents of social, economic, and digital progress Fourth, it provides practical guidance for institutional leaders, accreditation bodies, and policymakers aiming to embed systemic resilience, agility, and inclusiveness into the heart of curriculum reform.

In an era where curricula must prepare learners not only for existing jobs but for shaping the future of work and society, this paper positions curriculum transformation as both a theoretical imperative and a strategic opportunity.

The structure of the paper is organized as follows: Section 2 presents a synthesis of recent literature on curriculum innovation, digital transformation, institutional resilience, and multilevel governance. Section 3 outlines the theoretical framework that informs the study, integrating key concepts from educational systems theory and innovation studies. Section 4 details the methodological design and research phases, including stakeholder engagement mechanisms and analytics strategies. Section 5 provides an in-depth institutional profile of the Faculty of Economics–Skopje, highlighting its strategic orientation, leadership vision, and curricular innovations. Section 6 discusses the broader implications for higher education policy and practice. Section 7 concludes with key findings, research limitations, and directions for future studies.

2 Literature Review

The landscape of higher education has undergone significant disruption in recent decades, driven by technological advancement, labor market transformation, climate change, and global crises such as the COVID-19 pandemic. These forces have compelled universities worldwide to reassess the relevance and adaptability of their curricula. Scholars increasingly emphasize that curricula must evolve from static, discipline-bound structures to dynamic systems that integrate digital literacy, sustainability, experiential learning, and civic responsibility (Law, 2022; Paliwoda et al., 2025). Within this context, the concept of Curriculum 4.0 has emerged as a powerful framework for modernizing higher education. It promotes interdisciplinarity, problem-based learning, and the embedding of technological and soft skills, all of which are essential for graduates to thrive in the knowledge economy (Kunnari et al., 2018).

A growing body of literature has investigated how higher education institutions implement such transformative approaches. For example, Hasanefendic et al. (2017) introduce the notion of "individuals in action", faculty members who, equipped with intrinsic motivation, authority, and social networks, catalyze curricular reform from within their institutions. Their work underscores the importance of bottom-up innovation and points to the role of institutional culture in enabling or constraining change. Complementary studies by Laufer et al. (2025) explore trust-based leadership as a structural enabler, demonstrating how decentralized decision-making, transparency, and mutual trust can accelerate adoption of educational technology and participatory innovation.

At the governance level, Ritzen and Soete (2011) argue for the reconfiguration of educational governance structures to promote cross-level coordination among supranational entities, national ministries, and institutional actors. Their multilevel governance model frames HEIs as embedded within broader policy ecosystems and advocates for alignment between curriculum goals and societal needs. In parallel, Carayannis and Morawska-Jancelewicz (2022) propose the Quadruple Helix innovation model, which positions universities as central actors in collaborative networks that include industry, government, and civil society. These perspectives provide a theoretical foundation for systemic and stakeholder-inclusive curriculum reform.

Resilience is another key dimension addressed in literature. Dohaney et al. (2020) conceptualize resilience in higher education not only as the ability to recover from disruptions but also as a proactive stance of adaptability and preparedness. Ross et al. (2022) extend this idea by showing how organizational culture, leadership, and academic identity shape responses to systemic change. On the student side, Sahanowas and Halder (2024) emphasize that resilience can be cultivated through curricula that promote emotional intelligence, critical thinking, and metacognitive skills. Their findings advocate for integrated learning models that foster both academic performance and psychological adaptability.

In the digital domain, Eri et al. (2021) introduce the concept of digital resilience, which refers to learners' capacity to navigate and adapt to online learning

environments. Their comparative study across Asia and Australia reveals the critical role of institutional support structures, including IT infrastructure, pedagogical redesign, and professional development for instructors. These insights are especially relevant as universities increasingly adopt blended learning and digital tools such as learning analytics, AI-driven feedback systems, and immersive technologies.

Universities are increasingly seen as active agents in shaping equitable and sustainable futures. Spours and Grainger (2023) frame higher education institutions as mediators in a "just transition," emphasizing their role not only in responding to global challenges but in proactively guiding societal transformation. Through curricula that foster systems thinking, long-term impact assessment, and inclusive problem-solving, universities can help equip learners to navigate and lead change in complex, interdependent environments.

The literature points to three converging imperatives for curriculum reform: adaptability to digital and labor market change; integration of stakeholder input across governance levels; and cultivation of resilience at institutional, staff, and student levels. These imperatives underpin the theoretical and methodological choices of this study. They also justify the need for a general framework that not only guides curriculum reform conceptually but can be applied in diverse institutional contexts. The Faculty of Economics-Skopje, while serving as a case study in this research, exemplifies the global relevance of these themes. Its pursuit of innovation, resilience, and international accreditation provides fertile ground for examining how theory-informed reform can be both systemic and context-specific.

3 Theoretical Framework

The design of curriculum transformation must contend with the reality of complexity, institutional inertia, and shifting stakeholder demands. To address these challenges, the framework proposed in this study is built on a synthesis of four core theoretical perspectives: (1) complexity theory, (2) systems thinking, (3) multilevel governance, and (4) participatory engagement. Together, these lenses provide a holistic and actionable strategy for reimagining curriculum as a dynamic, evolving ecosystem embedded in broader institutional and societal systems.

3.1 Complexity Theory and Systems Thinking

Complexity theory provides a foundational understanding of curriculum transformation as a non-linear, adaptive process. Educational institutions are viewed as open systems that respond to continuous external pressures such as economic volatility, digital disruption, and socio-political change. Kolmos et al. (2015) argue that such systems must be flexible, reflexive, and capable of evolving in unpredictable environments. In this context, curriculum reform is not a one-off intervention but an ongoing learning process involving feedback loops, experimentation, and recalibration. Fenwick and Dahlgren (2015) emphasize that resilience emerges not from rigid stability but from the system's ability to reorganize and learn in response to crises. For the Faculty of Economics-Skopje at Ss. Cyril and Methodius University in Skopje this suggests the importance of embedding responsive structures into curriculum design, capable of adjusting to both anticipated and emergent challenges.

Systems thinking complements this by emphasizing the importance of structure, relationships, and purpose in shaping institutional change (Sterman, 2000). It provides a language for identifying leverage points and designing interventions that reflect the interconnected nature of academic governance, pedagogical design, and labor market relevance. Curricula are not isolated artifacts but subsystems within the larger university ecosystem, which itself is embedded in the regional, national, and global economy.

3.2 Multilevel Governance

Effective curriculum transformation requires alignment across governance levels, macro (national policy and accreditation standards), meso (university and faculty leadership), and micro (individual educators and students). Drawing from Ritzen and Soete's (2011) model, the framework integrates vertical and horizontal coordination mechanisms, ensuring that innovations are both top-down enabled and bottom-up informed. This multilevel logic is crucial in contexts such as North Macedonia, where European standards, national reforms, and local institutional dynamics converge.

3.3 Participatory Engagement

Transformative change cannot be imposed; it must be co-created. Participatory engagement draws from design thinking and stakeholder theory, advocating for the inclusion of diverse voices in shaping curriculum goals and delivery mechanisms (Carayannis & Morawska-Jancelewicz, 2022). Co-design with students, alumni, employers, faculty, and policymakers enhances ownership, relevance, and sustainability of reforms. This element ensures the framework is not only theoretically sound but contextually anchored and socially legitimate.

To enhance clarity and accessibility, a visual model of the framework has been developed (Fig. 1). The model situates curriculum at the center of intersecting systems, governance, stakeholder communities, and innovation pathways, demonstrating how systemic

inputs flow into curriculum redesign and how outcomes loop back into institutional learning. Each governance level (macro, meso, micro) is mapped with its respective roles, responsibilities, and stakeholder interactions.

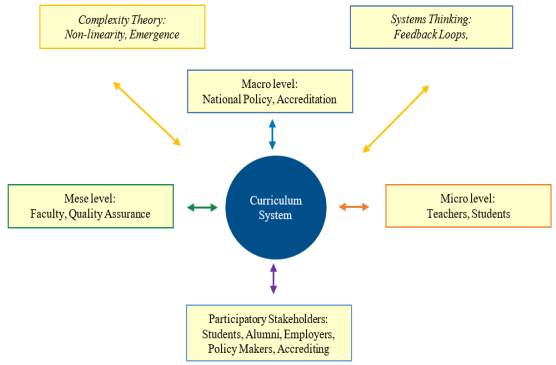


Figure 1. Theoretical Framework for Curriculum Transformation

4 Methodology and Data

This study adopts a mixed-methods research design that integrates both qualitative and quantitative approaches to explore curriculum transformation for innovation and resilience in higher education. The methodology is informed by the theoretical framework outlined in Section 3 and operationalizes its key constructs, complexity, systems integration, multilevel governance, and stakeholder engagement through a sequential and participatory research process. The research unfolds in four interconnected phases: (1) exploratory assessment, (2) co-design, (3) piloting and evaluation, and (4) strategy formulation. Each phase is strategically aligned with the principles of systems thinking and iterative learning and engages stakeholders from across governance levels, disciplines, and professional roles. We emphasize that the framework and methodological design proposed in this study are theoretical and have not yet been empirically implemented.

Phase 1: Exploratory Assessment

The first phase involves diagnosing the current state of curriculum innovation and institutional readiness. A structured questionnaire will be disseminated to faculty members, students, and alumni to assess perceptions of curriculum relevance, digital integration, practical orientation, and stakeholder responsiveness. The survey instrument draws on validated scales for digital resilience (Eri et al., 2021), critical thinking disposition (Sahanowas & Halder, 2024), and academic innovation (Hasanefendic et al., 2017). Simultaneously, semi-structured interviews will be conducted with ministry officials, university leadership, and employer partners to understand macro-level expectations, accreditation imperatives, and labor market alignment. The exploratory phase is grounded in the premise that innovation cannot be meaningfully pursued without understanding systemic constraints and stakeholder perceptions.

This phase will provide foundational data for constructing participatory mechanisms and will be further refined in collaboration with the Faculty's quality assurance and curriculum committees.

Phase 2: Co-Design

Building on the diagnostic findings, a Curriculum Transformation Working Group will be convened. This cross-functional team, comprising academic staff, students, alumni, employers, and policy representatives will engage in collaborative workshops to co-create proposed reforms. Techniques such as curriculum mapping, backward design, and persona-based planning will be employed to realign content, pedagogy, and outcomes. As Strachan et al. (2023) highlight, multidisciplinary collaboration is essential in designing curricula that embed sustainability, interdisciplinarity, and experiential learning.

This co-design effort will be supported by participatory design principles and will adopt an iterative validation model to ensure stakeholder buy-in and alignment with governance structures.

Phase 3: Piloting and Evaluation

Selected curriculum components developed in Phase 2 will be piloted in targeted undergraduate courses. For instance, an innovation lab course that brings students into direct collaboration with local firms or public institutions may be introduced. Evaluation of these pilots will be conducted through a combination of pre- and post-intervention surveys, focus groups, and classroom observations. Importantly, data visualization tools like Power BI will be used as a central analytical platform for real-time data visualization, tracking indicators such as engagement levels, satisfaction, digital tool usage, and perceived learning outcomes. Dashboards will display decomposed analytics across student demographics, academic departments, and instructional methods, offering granular insight into the effectiveness of interventions.

This phase incorporates dynamic data feedback loops and systems mapping, emphasizing complexity theory in identifying leverage points and emergent behaviors.

Phase 4: Strategy Formulation

Insights from the evaluation phase will be synthesized into a curriculum transformation strategy that includes design principles, governance recommendations, and a roadmap for scaling. This strategy will be documented in both policy and pedagogical formats enabling its adoption by faculty councils and academic program boards. At this stage, broader dissemination will occur through policy briefs, academic publications, and

stakeholder roundtables at the national level. The case of the Faculty of Economics-Skopje will serve as a benchmark model for other faculties in North Macedonia and beyond.
The strategy will be guided by institutional mission alignment, AACSB accreditation standards, and relevance to sustainable development goals (SDGs).
Figure 2 presents a conceptual diagram illustrating the four-phase methodology.

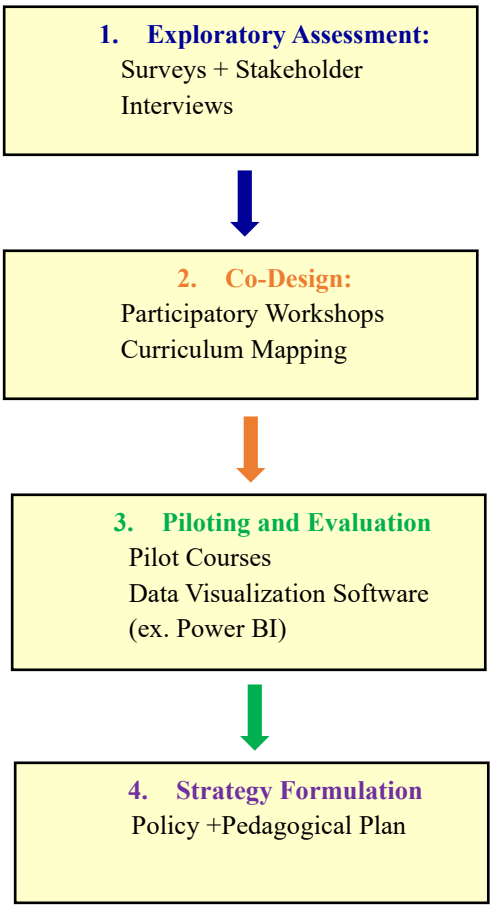


Figure 2. Visual representation of the four-phase methodology

The design of the questionnaire will be grounded in a comprehensive literature review of contemporary studies on curriculum transformation, innovation capacity, digital integration, resilience in education, and stakeholder engagement in higher education reform (e.g., Law, 2022; Kunnari et al., 2018; Hasanefendic et al., 2017). This analytical foundation ensures that the instrument captures not only generalizable constructs such as perceived curriculum relevance, pedagogical innovation, and institutional responsiveness, but also context-sensitive indicators tailored to the specific dynamics of educational systems in transition. The questionnaire will integrate validated scales and construct domains aligned with the study's theoretical framework, emphasizing complexity, systems thinking, and multi-level governance. Items will be carefully adapted and, where necessary, newly developed to reflect the distinctive priorities of digital transformation, sustainability, and assurance of learning. This approach ensures methodological rigor while enhancing the instrument's capacity to generate actionable insights for institutional decision-makers. Pre-testing will be conducted with faculty and students to ensure clarity, reliability, and contextual appropriateness before full-scale deployment.

All participants will be informed about the purpose of the study and their rights. Ethical approval will be sought from the institutional research board. To ensure the trustworthiness of findings, triangulation of methods and sources will be employed, and member checking will be applied during focus groups and interviews. A pilot study of the survey instrument will be conducted to refine clarity and reliability before full deployment.

To further enhance research integrity and alignment with contemporary standards, emphasis will be placed on transparency, inclusiveness, and responsiveness throughout the process.

This methodological design, rigorous, iterative, and stakeholder-driven ensures the integrity of the research and its alignment with contemporary expectations for responsive, inclusive, and evidence-based curriculum transformation.

5 Institutional Case: The Faculty of Economics-Skopje at Ss. Cyril and Methodius University in Skopje

The Faculty of Economics-Skopje, a constituent member of Ss. Cyril and Methodius University in Skopje, stands as the preeminent academic institution in North Macedonia for the study and advancement of economics and business disciplines. With over seven decades of tradition in higher education, the faculty has cultivated a reputation as a regional leader committed to academic excellence, internationalization, and social responsibility. Its graduates occupy leadership positions in public administration, private enterprises, and international organizations, reinforcing its identity as a cornerstone of the national knowledge economy.

The faculty has systematically strengthened its internationalization strategy by intensifying its participation in multilateral networks and embedding global standards within institutional governance. Its extensive portfolio of bilateral collaborations, Erasmus+ mobilities, and participation in European research frameworks (such as Horizon and COST projects) reflects a deliberate shift toward deeper international engagement. These partnerships not only elevate the faculty's visibility but also infuse global best practices into its teaching and research agendas. As such, the institution has positioned itself as a gateway between the domestic academic landscape and the broader European Higher Education Area.

A defining moment in this institutional trajectory has been the launch of a comprehensive curriculum reform initiative, one that places innovation, resilience, and digital fluency at the center of academic renewal. This reform effort is not episodic, but systemic, grounded in a strategic vision set forth by the Dean and its team. It responds both to the internal imperative for modernization and to the external expectations stemming from evolving student needs, employer expectations, and competitive accreditation environments.

The current initiative emerges under a visionary leadership agenda that positions innovation, resilience, and global integration at the heart of institutional development. This orientation is not merely rhetorical. Under the strategic guidance of the Dean and the team, the Faculty has undertaken bold steps toward internationalization and quality enhancement. One of the most emblematic of these efforts is the formal pursuit of AACSB accreditation, a globally recognized hallmark of academic excellence and continuous improvement in business education. This decision reflects not only a commitment to meet global standards but also a recognition that curriculum reform must be both evidence-driven and future-oriented.

In preparation for AACSB accreditation, the faculty has initiated several structural and pedagogical transformations. First, it has undertaken a comprehensive curriculum audit to map program learning goals against industry expectations, student competencies, and international trends in economic education. Second, faculty members have been mobilized into working groups aligned with AACSB's 9 accreditation standards, including Standard 4, which focuses explicitly on curriculum and embedding ethical reasoning, digital literacy, sustainability, and experiential learning into core and elective courses. Third, institutional governance processes have been streamlined to support Assurance of Learning (AoL) practices, linking student outcomes with feedback loops that inform teaching and assessment design. These internal shifts are aligned with global best practices promoted by AACSB but are also responsive to local socio-economic demands.

A pivotal moment in this process was a two-day participatory event initiated by the Dean during which the proposed structure of the redesigned curriculum was presented to internal staff. The new model reconsiders the number of credits allocated to specific modules, modernizes course sequencing, and aims to satisfy both national accreditation requirements and international quality expectations. The curriculum management process will incorporate structured participatory forums, exemplified by events where faculty, students, alumni, and industry partners collectively evaluate and refine the curriculum concept.

Leadership has played a decisive role in shaping the tone and direction of this transformation. The Dean's style, rooted in strategic foresight, inclusivity, and empowerment, has succeeded in mobilizing cross-functional teams, fostering trust, and building a unified institutional culture that views reform not as a burden but as an opportunity. This leadership model reflects an adaptive governance approach aligned with contemporary literature on organizational change in higher education. It has proven essential in aligning internal motivation with external accountability, making the Faculty of Economics-Skopje a national reference point for how strategic vision can translate into sustained academic excellence.

Through this ongoing transformation, the Faculty demonstrates that curriculum reform is not merely a technical or bureaucratic exercise, but a strategic process that redefines an institution's identity, values, and future trajectory. As the Faculty deepens its engagement with AACSB and other international quality assurance bodies, it continues to evolve as a dynamic, resilient, and forward-looking institution prepared to lead higher education in North Macedonia into the next generation.

6 Implications

The proposed framework for curriculum transformation holds far-reaching implications that extend beyond institutional boundaries and into the broader domains of national development, societal well-being, and sustainable innovation. At a time when education is called upon to serve as a pillar of societal resilience, this model offers a bold reconceptualization of how higher education can drive not only academic excellence but also economic adaptability, civic renewal, and ecological responsibility.

First, the framework contributes directly to national capacity-building. In countries like North Macedonia, where higher education institutions are increasingly navigating the dual demands of European integration and domestic reform, the need for cohesive, future-proof academic models has never been greater. By anchoring curriculum reform in a multilevel governance strategy, the framework enables alignment between government policy objectives, institutional mission, and community expectations. It fosters a model of higher education that supports national competitiveness, promotes equitable opportunity, and generates graduates capable of navigating and shaping the transformations of the 21st century.

At the societal level, the implications are equally profound. The framework shifts the focus of curriculum development from insular academic priorities to outward-facing, problem-oriented learning that connects students with real-world challenges. This reconnection between universities and their communities revitalizes the civic role of higher education, positioning institutions as mediators of social change and innovation. Through active engagement with industry, government, and civil society, the curriculum becomes a living interface

between knowledge creation and societal impact, training students not just to adapt, but to lead.

Crucially, the framework embeds sustainability not as an add-on, but as a core design principle. In doing so, it should align curriculum content and pedagogical practices with the Sustainable Development Goals (SDGs), particularly those related to quality education (SDG 4), decent work and economic growth (SDG 8), and climate action (SDG 13). By fostering interdisciplinary learning and ethical reasoning, students are equipped with tools to tackle complex, interlinked challenges such as climate change, digital inequality, and social fragmentation. This approach ensures that graduates are not only employable, but responsible, capable of contributing to both economic vitality and planetary stewardship.

From the perspective of innovation, the framework redefines how educational systems generate and sustain creative capacity. It operationalizes innovation through participatory curriculum design, digital transformation, and iterative improvement. Faculty are engaged as co-creators, empowered to design programs that reflect both disciplinary excellence and cross-sectoral relevance. Institutions benefit from dynamic feedback mechanisms through tools like Power BI and structured stakeholder consultations that allow continuous adaptation to new trends, technologies, and learner needs. This positions curriculum not as a static repository of knowledge but as a flexible, evolving platform for innovation.

Resilience, the ability to endure, adapt, and grow through disruption, is the defining characteristic of education systems in the post-pandemic era. The framework directly contributes to building institutional, pedagogical, and individual resilience. At the institutional level, it fosters agile governance structures and evidence-based decision-making. At the pedagogical level, it emphasizes learning-by-doing, scenario planning, and systems thinking. For students, the reformed curriculum cultivates emotional intelligence, critical reflection, and self-directed learning, all essential traits for personal and professional sustainability in an unpredictable world.

Finally, the implications for regional leadership are significant. By piloting this strategy at the Faculty of Economics-Skopje, North Macedonia has the opportunity to position itself as a frontrunner in educational transformation in Southeast Europe. The initiative serves not merely as an institutional innovation, but as a model that can inspire systemic change across other faculties, universities, and national systems. It signals that even in smaller or transitioning economies, world-class reform is possible when driven by strategic vision, inclusive governance, and a shared commitment to relevance, resilience, and responsibility.

The implications of this research are transformative. It offers a roadmap for embedding innovation, resilience, and sustainability into the heart of academic life, shaping not only how we teach and learn, but how education can serve the greater public good in a rapidly changing world.

7 Conclusion

This study introduces a novel and integrative framework for curriculum transformation that responds to the urgency of systemic reform in higher education. By embedding innovation, resilience, and multi-level alignment at its core, the proposed approach addresses both institutional challenges and wider socio-economic imperatives. Its added value lies in the synthesis of complexity theory, systems thinking, multilevel governance, and socio-cultural pedagogy, offering a unified lens through which curriculum reform can be approached holistically rather than piecemeal (Law, 2022; Kunnari et al., 2018; Hasanefendic et al., 2017).

Unlike traditional reform models that isolate curricular change from institutional governance or stakeholder input, this framework conceptualizes the curriculum as a living, adaptive system. In doing so, it moves beyond static program revisions to a dynamic, evidence-informed, and participatory model of academic transformation.

The methodology employed combines quantitative instruments (e.g., validated surveys, Power BI analytics) and qualitative mechanisms (interviews, focus groups, participatory workshops), ensuring both breadth and depth of analysis. This dual approach not only identifies existing curriculum gaps but also uncovers the socio-cultural and organizational levers needed for sustainable reform (Strachan et al., 2023).

The proposed approach carries implications that extend beyond institutional performance. By fostering interdisciplinary learning, digital fluency, and ethical responsibility, it supports the development of graduates who are equipped to lead societal, economic, and environmental transitions. Moreover, by embedding stakeholder co-creation and assurance of learning mechanisms, it strengthens governance structures and positions higher education as a resilient driver of national innovation systems (Eri et al., 2021; Sahanowas & Halder, 2024).

Future research will focus on the empirical validation of this framework through the development of tailored questionnaires and structured interview protocols for various stakeholder groups. These instruments will guide data collection on curriculum relevance, engagement, and readiness for innovation. The initial application of the framework will take place at the Faculty of Economics-Skopje, which has emerged as the institutional catalyst for this conceptual model. Insights from this pilot will inform further refinement, enabling broader implementation across other faculties and national education systems. In doing so, the study charts a pathway toward a more resilient, inclusive, and future-ready higher education landscape.

References

Carayannis, E. G., & Morawska-Jancelewicz, J. (2022). The futures of Europe: Society 5.0 and Industry 5.0 as driving forces of future universities. *Journal of the Knowledge Economy, 13*(4), 3445–3471. https://doi.org/10.1007/s13132-021-00854-2

Dohaney, J., de Róiste, M., Salmon, R. A., & Sutherland, K. (2020). Benefits, barriers, and incentives for improved resilience to disruption in university teaching. *International Journal of Disaster Risk Reduction, 50*, 101691. https://doi.org/10.1016/j.ijdrr.2020.101691

Eri, R., Gudimetla, P., Star, S., Rowlands, J., Girgla, A., To, L., Li, F., Sochea, N., & Bindal, U. (2021). Digital resilience in higher education in response to COVID-19 pandemic: Student perceptions from Asia and Australia. *Journal of University Teaching and Learning Practice, 18*(5), Article 7. https://doi.org/10.53761/1.18.5.7

Fenwick, T., & Dahlgren, M. A. (2015). Towards socio-material approaches in simulation-based education: Lessons from complexity theory. *Medical Education, 49*(4), 359–367. https://doi.org/10.1111/medu.12638

Hassan, M., Murtaza, A., & Rashid, K. (2024). *Redefining higher education institutions (HEIs) in the era of globalisation and global crises: A proposal for future sustainability*. European Journal of Education, 60(1), 92–105. https://doi.org/10.1111/ejed.12822

Hasanefendic, S., Birkholz, J. M., Horta, H., & van der Sijde, P. (2017). Individuals in action: Bringing about innovation in higher education. *European Journal of Higher Education, 7*(2), 101–119. https://doi.org/10.1080/21568235.2017.1296367

Kolmos, A., Hadgraft, R. G., & Holgaard, J. E. (2016). Response strategies for curriculum change in engineering. *International Journal of Technology and Design Education, 26*(3), 391–411. https://doi.org/10.1007/s10798-015-9319-y

Kunnari, I., Ilomäki, L., & Toom, A. (2018). Successful teacher teams in change: The role of collective efficacy and resilience. *International Journal of Teaching and Learning in Higher Education, 30*(1), 111–126.

Laufer, M., Deacon, B., Mende, M. A., & Schäfer, L. O. (2025). Leading with trust: How university leaders can foster innovation with educational technology through organizational trust. *Innovative Higher Education, 50*(1), 303–327. https://doi.org/10.1007/s10755-024-09733-5

Law, M. Y. (2022). A review of curriculum change and innovation for higher education. *Journal of Education and Training Studies, 10*(2), 16–23. https://doi.org/10.11114/jets.v10i2.5448

Nusche, D., Fuster Rabella, M., & Lauterbach, S. (2024), Rethinking education in the context of climate change: Leverage points for transformative change, *OECD Education Working Papers*, No. 307, OECD Publishing, Paris, https://doi.org/10.1787/f14c8a81-en.

Paliwoda, K., Strachan, S., & Rance, D. (2025, March). *Using experiential learning to engage students in interdisciplinary education for sustainable development: A vertically integrated project case study*. Paper presented at the Advance HE Annual Sustainability Symposium, University of Strathclyde.

Ritzen, J., & Soete, L. (2011). *Research, higher education and innovation: Redesigning multi-level governance within Europe in a period of crisis* (UNU-MERIT Working Paper No. 2011-056). United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology (UNU-MERIT). https://collections.unu.edu/view/UNU:71

Ross, P. M., Scanes, E., Poronnik, P., Coates, H., & Locke, W. (2022). Understanding STEM academics' responses and resilience to educational reform of academic roles in higher education. *International Journal of STEM Education, 9*(1), Article 11. https://doi.org/10.1186/s40594-022-00327-1

Sahanowas, S. K., & Halder, S. (2021). *Effect of emotional intelligence and critical thinking disposition on resilience of students in transition to higher education phase.* Journal of College Student Retention: Research, Theory & Practice, 23(4), 913–939. https://doi.org/10.1177/15210251211037996

Spours, K., & Grainger, P. (2023). The mediating role of further and higher education in a Just Transition social ecosystem. *Journal of Vocational Education & Training*, *76*(2), 307–330. https://doi.org/10.1080/13636820.2023.2258521

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Irwin/McGraw-Hill.

Using Large Language Models for Creating, Modifying and Interpreting Decision Tables in Multi-Criteria Method DEX

Marko Bohanec
Jožef Stefan Institute
Department of Knowledge Technologies
Jamova cesta 39, SI-1000 Ljubljana, Slovenia
marko.bohanec@ijs.si

***Abstract**. The potential of large language models (LLMs) was examined in the context of decision tables as used in the qualitative multi-criteria decision modelling method DEX (Decision EXpert). Interactive dialogues were conducted with two open-source LLM chatbots, DeepSeek and Llama, with their outputs evaluated from the perspective of an expert decision analyst. The interaction focused on the construction and modification of a single decision table, as well as its interpretation in terms of table properties, decision rules, attribute weights, and visualizations. Findings suggest that LLMs offer a convenient means for learning and executing simple tasks. However, at their current stage of development, they remain inconsistent and prone to errors, rendering them unsuitable for serious applications.*

Keywords. Large language models, multi-criteria decision making, method DEX (Decision EXpert), decision tables, decision rules

1 Introduction

Generative artificial intelligence (AI), particularly large language models (LLMs) (Ozdemir, 2023; Atkinson-Abutridy, 2024), has emerged as a transformative technology across various areas, revolutionizing tasks such as content creation, problem-solving, and decision-making. LLMs, trained on vast datasets, exhibit unparalleled capabilities in generating human-like text, answering complex queries, and even creating novel content.

Multi-Criteria Decision Making (MCDM) (Ishizaka & Nemery, 2013; Kulkarni, 2022) is a discipline focused on supporting decision-makers who face problems involving multiple, often conflicting criteria. It aims to identify the best or most preferred alternative(s) by explicitly considering these criteria and decision-maker's preferences. MCDM methods are widely applied across fields such as engineering, finance, and public policy, where decisions must balance trade-offs among diverse objectives. Among many MCDM approaches, DEX (Decision EXpert) (Bohanec, 2022) stands out as a qualitative, rule-based method that uses decision tables to represent and aggregate criteria hierarchically, facilitating transparent and interpretable decision models. Decision tables in DEX encode decision-maker's or expert's knowledge using if-then rules that map combinations of attribute values to evaluation outcomes.

This study explores the potential of LLMs for the creation, modification, and interpretation of DEX decision tables. The research was motivated by an event that surprised the author of this paper. In May 2025, he experimented with the DeepSeek chatbot (https://www.deepseek.com/) to test its file-upload capabilities. He uploaded a decision table (Table 1), formatted in a simple text-based layout, to see if the system would accept the data. After typing the prompt "*Uploaded is a DEX decision table"*, he accidentally pressed *Enter* before providing further instructions. To his surprise, the chatbot autonomously processed the file, correctly interpreting its structure and generating a concise set of decision rules (Response 1).

This one-time experience clearly indicated that LLMs are capable of representing and interpreting decision tables. This raised the follow-up research questions: what else can contemporary LLMs offer to a decision analyst working with DEX decision tables? To what extent and with what level of quality can they assist in the three core tasks: creation, modification, and interpretation of such tables? What functionality is provided out of the box, what requires user-supplied methodological guidance, and what types of errors or distractions may arise during LLM use?

The research method was deliberately kept light-weight and simple: we conducted several sessions with open-source chatbots running the models DeepSeek-r1:14b (14.8 billion parameters) and Llama3:3 (70.6 billion), both installed locally at the author's organization. We asked the same questions in the same order, except when responses of one or both LLMs

were unsatisfactory and required additional instructions. We conducted several sessions using decision tables of various sizes: two to three input attributes, up to 24 elementary rules. We observed the responses and assessed them from the viewpoint of an expert decision analyst.

Table 1. A DEX decision table in text format.

Injury History	Health/Nutrition Habits	Health & Injury Profile
frequent	poor	very poor
frequent	average	poor
frequent	good	poor
frequent	excellent	average
occasional	poor	poor
occasional	average	average
occasional	good	good
occasional	excellent	good
rare	poor	average
rare	average	good
rare	good	good
rare	excellent	excellent
none	poor	average
none	average	good
none	good	excellent
none	excellent	excellent

Response 1. DeepSeek's interpretation of Table 1.

Analysis of Health & Injury Profile Decision Table
This is a decision table that maps combinations of "Injury History" and "Health/Nutrition Habits" to a "Health & Injury Profile" assessment.
Key Observations:
1. **Injury History Impact**: The frequency of injuries has a clear correlation with the health profile:
   - Frequent injuries consistently lead to poorer health profiles
   - No injury history generally leads to better assessments
2. **Nutrition Habits Influence**: Better nutrition habits can mitigate some effects of injury history:
   - Even with frequent injuries, excellent nutrition can result in an "average" profile
   - Poor nutrition with no injury history only reaches "average"
3. **Combined Effects**:
   - The best profile ("excellent") requires either:
     - No injury history + good/excellent nutrition, OR
     - Rare injuries + excellent nutrition
   - The worst profile ("very poor") only occurs with frequent injuries + poor nutrition

Decision Rules Summary:
- **Very Poor**: Only when injury history is frequent AND nutrition is poor
- **Poor**: Frequent injuries with average/good nutrition OR occasional injuries with poor nutrition
- **Average**: Several combinations including frequent+excellent, occasional+average, rare+poor, or none+poor
- **Good**: Occasional+good/excellent or rare+average/good
- **Excellent**: Rare+excellent or none+good/excellent

In the next section, we briefly review the related work. In sections 3–5 we explore the performance of LLMs for the creation, modification and interpretation of decision tables, respectively. Basic concepts of DEX decision tables related to each task are explained along the way. To keep the text reasonably short and consistent, we illustrate the process using a small decision table with two input attributes and 12 elementary rules. Section 6 concludes the paper.

2 Related Work

While LLMs are receiving a lot of attention in recent scientific literature, their coverage in the contexts of MCDM and/or decision tables is still scarce. Several authors have suggested incorporating LLMs in the MCDM process. Wang et al. (2025) proposed a framework using LLMs in the model preparation and evaluation stages, comparing it with the MCDM methods AHP (Analytic Hierarchy Process) and FCE (Fuzzy Comprehensive Evaluation). Similarly, Svoboda & Lande (2024) proposed a decision analysis framework for cybersecurity that combines AHP with the GPT-4 LLM. The same LLM is used in software 1000minds (https://www.1000minds.com/), which incorporates an AI assistant to enhance user interaction by suggesting decision criteria and alternatives according to the MCDM method PAPRIKA (Potentially All Pairwise RanKings of all possible Alternatives). Radovanović et al. (2024) used LLMs to learn the structure and some components of DEX models from data.

While decision tables are an important component of information systems, they are still seldom addressed in connection with LLMs. In their study, Lu et al. (2024) provide a comprehensive overview of table-related tasks, covering tasks like table question-answering, spreadsheet manipulation and table data analysis, using both traditional and LLM-supported techniques. Goossens et al. (2023) experimentally evaluated an automated approach to generating decision tables from natural language, using GPT-3. Their findings indicate that GPT-3 can understand the decision context, identify the input and output variables, and provide template decision tables for problem-solving, but in general it cannot create complete and correct decision tables. The latter finding is important because it relates to completeness and consistency, two essential characteristics of DEX decision tables, as explained in the next section.

3 Creating DEX Decision Tables

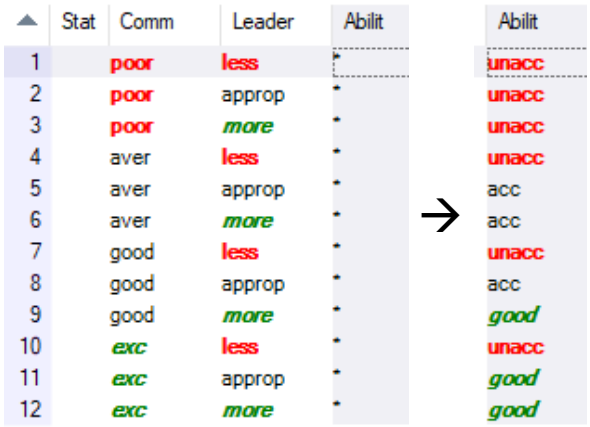
A DEX decision table is always bound to the context of one or more input attributes and one output attribute (outcome). *Attributes* are qualitative variables that can take their values from a discrete set of values, which are usually represented by words. Usually, value sets (called *scales*) are small (two to five values) and preferentially ordered from "bad" to "good" values. For example, Table 1 consists of two input attributes: (1) *Injury History*, having the scale <frequent, occasional, rare, none>, and (2) *Health/Nutrition Habits*: <poor, average, good, excellent>. The output attribute is *Health & Injury Profile*, using the scale <very poor, poor, average, good, excellent>. Each row in the table is called an *elementary decision rule* and maps some combination of input values to some outcome value. Generally, a DEX decision table is expected to have two important properties:

- *Completeness*: Contains all possible combinations of input attributes' values, without duplicates.

- *Preferential consistency*: When comparing two rules and the second one has all input values better than or equal to the first one, then the second outcome should be better or equal, too. We say that the second rule *dominates* the first. If this holds for all pairs of comparable rules, the table is *monotone* and is said to "obey the principle of dominance".

Traditionally, DEX decision tables are created interactively using software such as DEXiWin (Bohanec, 2024a; 2025). After defining input and output attributes together with their scales, DEXiWin constructs the decision table containing all possible combinations of input values, but with undefined outcomes (denoted by '*').

Fig. 1 shows an example decision table, which combines the input attributes *Comm* (communication skills of an employee candidate) and *Leader* (leadership skills) to *Ability* (aggregated assessment of the candidate's abilities). The example is extracted from the employee-selection model called *EmploySmallSuite*, which is distributed with DEXiWin. This example table is used hereafter.



	Stat	Comm	Leader	Abilit
1		poor	less	*
2		poor	approp	*
3		poor	more	*
4		aver	less	*
5		aver	approp	*
6		aver	more	*
7		good	less	*
8		good	approp	*
9		good	more	*
10		exc	less	*
11		exc	approp	*
12		exc	more	*

	Abilit
1	unacc
2	unacc
3	unacc
4	unacc
5	acc
6	acc
7	unacc
8	acc
9	good
10	unacc
11	good
12	good

Figure 1. Decision table creation in DEXiWin.

The right-hand side of Fig. 1 shows the *Abilit* column after it has been completely defined by the user. The process is usually quite dynamic, as it provides several techniques for defining outcomes (direct assignment, considering dominance of rules, considering weights of attributes, and considering attributes' symmetries), while continuously assessing the completeness and consistency of the table. A detailed description of those is beyond the scope of this paper; see Bohanec (2024a) for more information.

Can we construct (or reconstruct) such a decision table using LLMs? The dynamic DEXiWin-type interaction seems unfeasible in the text-oriented communication employed by chatbots. As an alternative, let us first define the bounding attributes and let the LLM to construct a decision table without further instructions:

P1: *Consider three qualitative attributes according to method DEX:*
(1) Comm (communication skills of an employee candidate), possible values: poor, average, good, excellent;
(2) Leader (leadership skills): less, appropriate, more;
(3) Ability (aggregated assessment of the candidate's abilities): unacceptable, acceptable, good.
Create a DEX decision table that maps Comm and Leader to the values of Ability. Call this table T0.

The two LLMs suggested different decision tables and presented them in the format similar to Fig. 1. The outcomes were not equal to Fig. 1, though. Importantly, however, both tables were complete (each contained 12 distinct input value combinations) and monotone, even if no instructions were given to this end. In order to save space in this paper and compactify responses, we formulated an additional request:

P2: *Display T0 using Comm as rows and Leader as columns.*

The response to P2, which was the same for both LLMs, is shown in Response 2. Notice that the required format would be unsuitable for decision tables with three or more input attributes.

Response 2. Decision table T0. Column headings correspond to the values of *Leader*, cells contain values of *Ability*.

Comm	Less	Appropriate	More
Poor	Unacceptable	Unacceptable	Acceptable
Average	Unacceptable	Acceptable	Acceptable
Good	Acceptable	Acceptable	Good
Excellent	Acceptable	Good	Good

While this result is satisfactory, we need to make a few remarks. In our previous experience with LLMs, we noticed that they tend to "forget" their suggestions from one question to another. This is why we explicitly requested to name the table T0. Such need of keeping track of results is annoying, distracts the user from the primary task and is easy to forget. Furthermore, the naming technique actually turned out ineffective with Llama, which proposed different table contents after P1 and P2. This is considered a serious flaw of LLMs, which seems to be improving with newer versions, but has still not been satisfactory resolved.

4 Modifying DEX Decision Tables

Even though the syntax of T0 is correct, we are not entirely satisfied with some of the rules. Thus, we requested the change:

P3: *Modify T0 so that whenever Comm=poor or Leader=less, Ability is unacceptable. Call the result TR.*

Both LLMs responded correctly (Response 3). In this way, we successfully reconstructed the decision table from Fig. 1.

Response 3. Decision table TR.

Comm	Less	Appropriate	More
Poor	Unacceptable	Unacceptable	Unacceptable
Average	Unacceptable	Acceptable	Acceptable
Good	Unacceptable	Acceptable	Good
Excellent	Unacceptable	Good	Good

While we successfully constructed TR in just two steps, this was typically not the case with larger decision tables (containing up to 24 rows). Usually, the initial decision tables suggested by LLMs (such as T0) were quite good and did make sense, however they only reflected "general" opinions, known to the LLM after consulting a vast amount of data. Human decisions are subjective, so decision tables should reflect decision maker's subjective preferences and should be formulated accordingly. For larger decision tables, this turned out a challenging task, which we tried to accomplish by formulating rules, similar to those in request P3. The process typically required several steps and was often distracted by LLM's forgetting previously formulated rules, which the user assumed they were already fixed. At the current level of LLM development, it seems that it is still more effective to create decision tables using specialized tools and importing them to LLMs, as illustrated in Table 1.

5 Interpreting DEX Decision Tables

Traditional tools for building decision tables, such as DEXiWin, guide the decision maker to focus on one elementary decision rule at a time. As the user defines individual outcomes, built-in algorithms monitor and ensure the overall consistency and completeness of the table. However, as the number of rules increases, particularly beyond 20, the resulting tables may become difficult to interpret as a whole. Therefore, more compact and readable representations are essential to support users in reviewing the content during model construction. Once complete, such representations also enhance the model's transparency and trustworthiness. To this end, DEXiWin offers multiple methods for interpreting decision tables in terms of aggregate (compact) decision rules, decision trees, attribute weights, and 3D visualizations (Bohanec, 2024a).

5.1 Properties of Decision Tables

To assess LLMs' abilities to interpret decision tables, let us continue with the TR table and first ask some questions about its properties.

P4: *Is TR complete?*

Both LLMs answered correctly that TR was complete and provided valid explanations referring to 12 unique non-empty combinations of input attributes' levels.

P5: *Is TR preferentially consistent?*

DeepSeek answered positively with a convincing explanation, summarizing that "*TR is preferentially consistent because the rankings and transitions between communication, leadership, and ability align logically without any contradictions*". Llama answered negatively, correctly quoting the definition of preferential consistency, but failing to detect that the whole table indeed adhered to the principle *"[when] one attribute improves (e.g., moving from "poor" to "average" Comm, or from "less" to "appropriate" Leader), the output (Ability) either stays the same or improves, but does not worsen.*"

P6: *Is TR symmetric?*

TR is clearly non-symmetric, because its input attributes use value scales of different cardinalities. Both LLM answered correctly and convincingly, justifying their decision:

- DeepSeek: "*TR lacks symmetry because the interactions between communication and leadership skills do not produce equivalent outcomes when the roles of these attributes are switched or varied independently*."
- Llama: *"[...] it's clear that Comm and Leader have distinct effects on Ability, and their influence is not symmetric.*"

5.2 Representations Using Decision Rules

The following requests were aimed to check the LLMs abilities to represent TR using different and possibly more compact representations.

P7: *Represent table TR using aggregate decision rules.*

When formulating this request, we did not define the concept of aggregate decision rules, assuming that the LLMs already knew it from DEX-related literature. This was not the case, and the answers reflected the LLMs own understanding of the concept. Nevertheless, the answers were logically correct. Response 4 shows the rules proposed by Llama. Notably, none of the LLMs identified the highly compact rule "Ability is unacceptable (unacc) if: Leader is less (regardless of Comm)." DEXiWin normally discovers such rules by allowing them to have mutually overlapping conditions.

Decision trees (Rokach & Maimon, 2014) are widely used in machine learning for representing decision logic. We asked:

P8: *Represent table TR using a decision tree.*

As expected, the LLMs had no issues with this representation. Among the two, DeepSeek's solution (Response 5) was considered slightly more readable.

Response 4. TR represented according to Llama's understanding of "aggregate rules".

- **Ability is unacceptable (unacc) if**:
  - Comm is poor (regardless of Leader),
  - or Comm is average and Leader is less,
  - or Comm is good/excellent and Leader is less.
- **Ability is acceptable (acc) if:**
  - Comm is average and Leader is appropriate or more,
  - or Comm is good and Leader is appropriate.
- **Ability is good if:**
  - Comm is good and Leader is more,
  - or Comm is excellent and Leader is appropriate or more.

Response 5. Decision tree representation of TR by DeepSeek.

- If Comm = poor: Directly conclude Ability = unacceptable (unacc).
- If Comm = average:
  - Next test: Leadership (Leader)
    - If Leader = less › Ability = unacc
    - If Leader = appropriate or more › Ability = acceptable (acc)
- If Comm = good:
  - Next test: Leader
    - If Leader = less › Ability = unacc
    - If Leader = appropriate › Ability = acc
    - If Leader = more › Ability = good
- If Comm = excellent:
  - Next test: Leader
    - If Leader = less › Ability = unacc
    - If Leader = appropriate or more › Ability = good

Response 6. "Ripple-down rules" representation of TR as proposed by Llama (incorrect).

Rule 1: If Comm is poor, then Ability is unacc.
Rule 2: If Comm is average and Leader is less, then Ability is unacc.
Rule 3: If Comm is average and Leader is appropriate or more, then Ability is acc.
Rule 4: If Comm is good and Leader is less, then Ability is unacc.
Rule 5: If Comm is good and Leader is appropriate, then Ability is acc.
Rule 6: If Comm is good and Leader is more, then Ability is good.
Rule 7: If Comm is excellent and Leader is less, then Ability is unacc.
Rule 8: If Comm is excellent and Leader is appropriate or more, then Ability is good.

Ripple-down rules (Compton & Jansen, 1990) are a somewhat less known representation of decision logic. After identifying the most common outcome (e.g., *Ability*=unacceptable) at the top level, the next levels consist of a hierarchy of context-dependent rules, where each new rule refines a parent rule by adding exceptions. Ripple-down rules are not implemented in DEXiWin and we wanted to know how are they handled by the LLMs:

P9: *Represent table TR using ripple-down rules.*

Both LLMs failed. Instead of making a general-to-detailed hierarchy of conditions, they both suggested *"[...] a set of rules that are applied in a specific order*", as shown in Response 6.

Observing the results of P7–9, we noticed that the LLMs did not take advantage of the fact that all attribute scales were preferentially ordered. To see if this affects LLMs in any way, we formulated the sequence of follow-up requests:

P10: *Consider that all attributes in table TR are preferentially ordered. This means that you can use the relations '>=' (better than or equal) and '<=' (worse than or equal) to compare values on these scales. Using these relations on input attributes Comm and Leader, represent table TR in terms of aggregate decision rules.*

P11: *Using the relations, represent table TR in terms of a decision tree.*

P12: *Using the relations, represent table TR in terms of ripple-down rules.*

Unfortunately, this made more harm than good. While the new interpretations indeed used the suggested relations, making more compact representations in some cases, many of them contained subtle and hard-to-detect errors. For instance, Llama's aggregate rules (Response 7) and ripple-down rules (Response 8) contain logical errors, which can be identified by following the respective logic and comparing the results with TR presented in Response 3. For instance, the rules for *Comm*=good and *Leader*=appropriate should result in *Ability*=acceptable, but do not. In contrast, DeepSeek's solutions remained largely unchanged: the aggregate and ripple-down rules were logically sound, but the decision tree grew more complex and introduced a minor logical flaw (not shown here).

Response 7. Representation of TR using aggregate rules with relational operators (Llama, incorrect).

1. If Comm <= average and Leader is any value, then Ability is unacc.
2. If Comm >= good and Leader >= appropriate, then Ability is good.
3. If Comm >= average and Leader >= appropriate, but not (Comm >= good and Leader >= appropriate), then Ability is acc.
4. In all other cases, Ability is unacc.

Response 8. Representation of TR using ripple down rules with relational operators (Llama, incorrect).

1. Rule 1: If Comm <= poor, then Ability is unacc.
2. Rule 2: If Comm <= average and Leader < appropriate, then Ability is unacc.
3. Rule 3: If Comm >= average and Leader >= appropriate, but Comm < good, then Ability is acc.
4. Rule 4: If Comm >= good and Leader >= appropriate, then Ability is good.

5.3 Representations Using Weights

In the next set of questions, we wanted to test LLMs' understanding of attribute weights (importances). This

is an overarching concept in MCDM (Kulkarni, 2022), where weights represent the relative importance assigned to each attribute and are fundamental in aggregating criteria to evaluate and rank alternatives. In DEX, weights are derived from decision rules, employing the least square regression.

We first checked the assessment of weights using information gain, a method commonly used in machine learning (Rokach & Maimon, 2014) from tabular data.

P13: *Assess the weights (importances) of input attributes in table TR using information gain. Explain the method and carry out the calculation. At the end, normalize the weights so that their sum equals 100.*

Both LLMs correctly explained the method and its main steps, but executed it inappropriately, giving incorrect results. The DeepSeek's explanation was insufficient to determine the cause, while Llama apparently made a mistake determining the table's size, using the number 16 instead of 12.

P14: *Now use ordinal attribute values. Assess attribute weights (importances) in table TR employing least squares regression. Explain the method and carry out the calculation. At the end, normalize the weights so that their sum equals 100.*

Again, both LLMs correctly explained the least squares method, which consists of approximating TR with

$$Ability = \beta_0 + \beta_{Comm} Comm + \beta_{Leader} Leader + \varepsilon$$

where β_0 is the intercept, β_{Comm} and β_{Leader} are the coefficients representing the effect of *Comm* and *Leader* on *Ability*, and ε is the error term. Both LLMs incorrectly assumed that $\beta_0 = 0$, and suggested to "*use a least squares regression calculator*" for the rest. Afterwards, they assumed their own values for β_{Comm} and β_{Leader} (not explaining how) and then carried out the normalization of weights correctly. Consequently, weights, calculated by the LLMs, were different from each other, and wrong: 56.67% and 43.33% by DeepSeek, and 55.56% and 44.44% by Llama. The correct answer, according to DEXiWin, is 50.98% and 49.02%.

All these results indicate that LLMs are indeed weak in mathematical computation and logical inference: they tend to "hallucinate" (Banerjee, 2024) and make up their answers. We knew this before and it was expected. However, the real problem is that LLMs' explanations of applied methods, calculations and inference procedures are exceptionally convincing to the level that can easily mislead the user. Errors are subtle and hard to detect, requesting the users to know and understand the applied methods really well, to meticulously check each and every response, and possibly even use external tools to find out the correct outcomes.

5.4 Visual Representations

In the last group of questions, we tested the LLMs' abilities to visualize data tables. Both DeepSeek and Llama do this by producing Python code, which, when run externally, draws the requested charts. Our first attempts were not really satisfactory, but after adding additional requests of how charts should look like, the results were becoming better and better. The following request resulted in computer code that, after some minor editing, gave a nice 3D chart of TR (Fig. 2), similar to those produced by DEXiWin.

P15: *Display table TR in a 3D chart, or provide code in one file. Use different colors for different levels of Ability. Use red and green color, respectively, for bad and good values of Ability. Connect displayed data points with dashed lines in each attribute direction.*

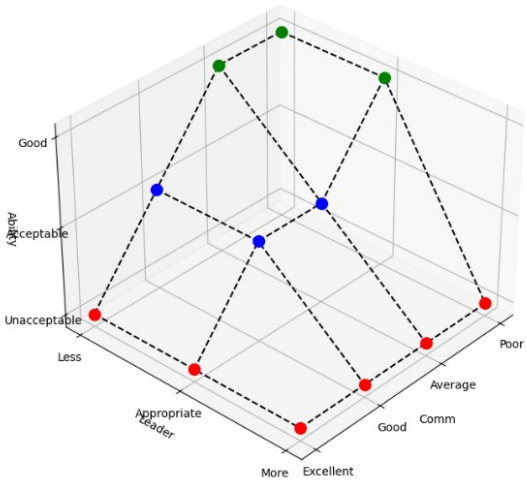


Figure 2. 3D chart of TR (DeepSeek + code edit).

2D charts, after several trial-and-error attempts and minor code editing, were also satisfactory (two variations produced by Llama are shown in Fig. 3 and Fig. 4).

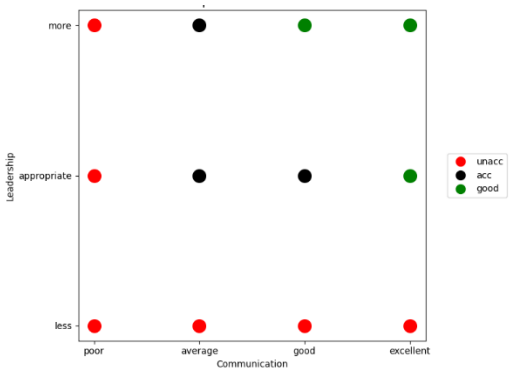


Figure 3. 2D chart of TR, using dots (Llama + code edit).



Figure 4. 2D chart of TR, filling rectangles (Llama + code edit).

6 Conclusion

We explored the abilities of large language models to operate with DEX decision tables that map two or more qualitative attributes to a qualitative outcome. The research was carried out by running two recent and popular open-source LLMs (DeepSeek and Llama) in parallel, asking them questions that addressed the most basic functionality necessary to handle an individual decision table: construction and modification, and interpretation in terms of table properties, decision rules, weights, and visualizations.

Specifically, we wanted to sense the operation of LLMs from the position of an ordinary user, who might want to apply LLMs on decision tables. However, the results themselves were assessed from the position of a skilled decision analyst.

Our findings yielded mixed results. On the one hand, LLM chatbots demonstrate an impressive ability to interpret user inputs and provide a substantial amount of knowledge "out of the box". LLM's responses are highly persuasive, and its engaging interface captures the user's attention, making the interaction both effective and enjoyable. Although the DEX method is not among the most widely known MCDM approaches, LLMs demonstrate a solid understanding of its core principles and key characteristics.

On the other hand, working with LLMs can often be uneven and, at times, frustrating. As highlighted in sections 5.2 (decision rules) and 5.3 (weights), the LLMs frequently provide convincing explanations that may conceal subtle errors that are difficult to detect and may mislead the user. The context (e.g., the current decision table), which the user may reasonably assume to be fixed, can change implicitly without notice or explanation (this behavior was particularly evident with Llama). Additionally, the programming code generated by LLMs occasionally contains errors that are nontrivial to diagnose and correct. After successfully resolving some issue, it may reappear several steps later. Consequently, rather than focusing on solving the problem at hand, the user needs to adopt the role of a meticulous supervisor, critically evaluating each step. This is both cognitively demanding and requires a high level of expertise in the applied methods, which is normally unnecessary while using specialized MCDM software.

In summary, at their current stage of development, LLMs are not yet suitable for reliably working with DEX decision tables. They are highly effective for learning the underlying concepts and performing basic tasks, such as summarizing decision tables (section 1) and analyzing their properties (section 5.1). LLMs may also prove useful for exploring less accessible algorithms, such as ripple-down rules (section 5.2), or generating customized visualizations (section 5.4). However, for more advanced operations, the limitations of LLMs remain too significant.

This study focused on working with individual decision tables. However, a typical DEX model comprises multiple decision tables, each associated with an internal node in a hierarchical attribute structure. Given the challenges we encountered in constructing and maintaining the consistency of even a single decision table within a session, it is reasonable to conclude that managing complete DEX models exceeds the capabilities of current LLM technology.

Work to date has been restricted to only two LLMs and small decision tables with only two or three input attributes, one output attribute, and at most 24 elementary decision rules. In practice, however, decision tables often contain 100 or more rules. Also, we did not investigate the ability of LLMs to restructure decision tables, for instance when adding or deleting an attribute or its value, or changing attribute weights; this typically requires quite advanced algorithms (Bohanec, 2024b). Moreover, the generalizability and replicability of our findings could have been enhanced through a more systematic experimental design, using APIs (Application Programming Interfaces). Future research may extend in these directions, particularly as LLMs continue to evolve. Given their rapid development, significant improvements in their applicability to decision table analysis can reasonably be expected.

Acknowledgments

The author acknowledges the financial support from the Slovenian Research and Innovation Agency for the programme Knowledge Technologies (research core funding No. P2-0103).

References

Atkinson-Abutridy, J. (2024). *Large Language Models: Concepts, Techniques and Applications.* Boca Raton: CRC Press. https://doi.org/10.1201/9781003517245.

Banerjee, S., Agarwal, A., & Singla, S. (2024). LLMs will always hallucinate, and we need to live with this. 10.48550/arXiv.2409.05746.

Bohanec, M. (2022). DEX (Decision EXpert): A qualitative hierarchical multi-criteria method. In: *Multiple Criteria Decision Making* (ed. Kulkarni, A.J.), Studies in Systems, Decision and Control 407, Singapore: Springer, doi: 10.1007/978-981-16-7414-3_3, 39–78.

Bohanec, M. (2024a). *DEXiWin: DEX Decision Modeling Software, User's Manual, Version 1.2*. Ljubljana: Institut Jožef Stefan, Delovno poročilo IJS DP-14747. Accessible from https://dex.ijs.si/dexisuite/dexiwin.html.

Bohanec, M. (2024b): *Transformations of decision tables in qualitative multi-criteria decision modeling method DEX*. Institut Jožef Stefan, Delovno poročilo IJS DP-14761, 2024. https://kt.ijs.si/MarkoBohanec/pub/2024_DP14761_Transformations.pdf

Bohanec, M. (2025): DEXi Suite: DEXi decision modelling software. *SoftwareX* 31, 102240. https://doi.org/10.1016/j.softx.2025.102240 .

Compton, P., & Jansen, R. (1990). A philosophical basis for knowledge acquisition. *Knowledge Acquisition* 2(3), 241–257. doi:10.1016/S1042-8143(05)80017-2.

Goossens, A., Vandevelde, S., Vanthienen, J., & Vannekens, J. (2023). GPT-3 for Decision Logic Modeling. *RuleML+RR'23: 17th International Rule Challenge and 7th Doctoral Consortium*, Oslo, Norway.

Ishizaka, A., & Nemery, P. (2013). *Multi-criteria decision analysis: Methods and software*. Chichester: Wiley. doi:10.1002/9781118644898.

Kulkarni, A.J. (Ed.) (2022). *Multiple Criteria Decision Making*. Studies in Systems, Decision and Control 407, Singapore: Springer, doi: 10.1007/978-981-16-7414-3_3.

Lu, W., Zhang, J., Fan, J., Fu, Z., Chen, Y., & Du, X. (2025). Large language model for table processing: A survey. *Frontiers of Computer Science*, *19*(2), 192350. https://doi.org/10.1007/s11704-024-40763-6.

Ozdemir, S. (2024*). Quick Start Guide to Large Language Models: Strategies and Best Practices for ChatGPT, Embeddings, Fine-Tuning, and Multimodal AI*. 2nd Edition. Addison-Wesley Professional. ISBN: 978-0135346563.

Radovanović, S., Delibašić, B., & Vukanović, S. (2024). Combining LLM and DIDEX method to predict Internal Migrations in Serbia. *Proc. 24th International Conference on Group Decision and Negotiation & 10th International Conference on Decision Support System Technology (GDN ICDSST 2024), Vol. 1: Technology as a support tool* (Eds. S.P. Duarte, P. Zaraté, A. Lobo, B. Delibašić, T. Wachowicz, M.C. Ferreira), University of Porto.

Rokach, L., & Maimon, O. (2014). *Data Mining with Decision Trees: Theory and Applications* (2nd. ed.). River Edge, USA: World Scientific Publishing Co., Inc. ISBN:978-981-4590-07-5.

Svoboda, I., & Lande, D. (2024). Enhancing multi-criteria decision analysis with AI: Integrating analytic hierarchy process and GPT-4 for automated decision support. https://arxiv.org/abs/2402.07404.

Wang, H., Zhang, F., Mu, C. (2025). One for all: A general framework of LLMs-based multi-criteria decision making on human expert level. arXiv e-prints, 2502.15778. https://arxiv.org/abs/2502.15778.

Students' Cognitive Engagement in Blended Learning: Observations from Event-contingent Self-report and Log Data

Sabina Rako, Nina Begičević Ređep, Barbara Šlibar, Diana Šimić
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
sabina.rako@student.foi.hr,{nina.begicevic,barbara.slibar,diana.simic}@foi.unizg.hr

Bart Rienties
Open University UK
Institute of Educational Technology
Walton Hall, Milton Keynes, UK
bart.rienties@open.ac.uk

Abstract. *Student engagement plays a crucial role in the learning process in higher education. This study aims to explore students' cognitive engagement in a blended learning environment by combining data collected through event-contingent self-report and Learning Management System (LMS) log data. Participants were undergraduate students enrolled in a course on business decision making, specifically adapted for information technology students. Responses from 140 students revealed that students felt more actively involved during in-person learning, while concentration and focus levels were comparable between in-person and online modes. Polychoric correlation indicated a moderate positive relationship between concentration and passive-active items, whereas principal component analysis highlighted latent constructs. Analysis of LMS log data provided additional behavioral indicators, including time spent in the system, the number of distinct learning components accessed, and the type of activity performed (distinguishing between 'just view' and 'more than view' actions). Two distinct student profiles emerged from the combined dataset. This approach shows that combining self-reported and LMS log data provides a clearer picture of how students engage in blended learning.*

Keywords. cognitive engagement, self-reports, logs, higher education, decision making

1 Introduction

Blended learning, which combines classroom-based learning with online learning, is nowadays widely used in higher education (Vo et al., 2024). Blended learning offers flexibility and access to various learning resources, but keeping students consistently and meaningfully engaged is still a challenge (Bergdahl et al., 2024).

Tight (2020) argues, based on a comprehensive literature review, that research on student engagement in higher education has been largely driven by institutional efforts to improve student retention. A systematic literature review and meta-analysis of 137 studies published between 2003 and 2019, involving 158,510 students, by Wong et al. (2024) found that student engagement has a strong correlation with academic achievement and subjective well-being, but also emphasized that the construct of engagement is often overgeneralized in current research. While this meta-analysis (Wong et al., 2024) focused on elementary, middle, and high school education, engagement is also a widely discussed construct in the context of higher education (Lowe & El Hakim, 2020; T. D. Nguyen et al., 2018; Prananto et al., 2025; Tani et al., 2021).

Advances in educational technology and the widespread adoption of Learning Management Systems (LMS) in higher education have enabled the collection of detailed data on student interactions (Henrie et al., 2015; Ifenthaler & Yau, 2020). Furthermore, an analysis of studies using Learning Analytics (LA) published between 2011 and 2023 shows that student engagement in higher education is still predominantly measured through observable behaviors like clicks and time on task (Bergdahl et al., 2024). A deeper understanding of student engagement in higher education, supported by LA, could also improve learning design and provide more effective support for students (Q. Nguyen et al., 2022). Additionally, combining system data with self-reports, such as surveys or reflections collected shortly during or after the learning activity, could provide a more complete picture by capturing

both actions and experiences, thereby contributing to Human-Centered Data Science (Aragon et al., 2022).

The variety of learning activities and the combination of in-person and online modes in blended learning create challenges for sustaining cognitive engagement (Halverson & Graham, 2019). Understanding variations in engagement is key to designing effective blended learning. Therefore, this study is driven by the following research questions:

- RQ1: In what ways does students' event-contingent self-reported cognitive engagement differ across learning modes (in-person and online) and activities in a blended learning environment?
- RQ2: How can LMS event logs help to explain or add to self-reported data about students' cognitive engagement in online learning?

To answer these questions, we collected and synchronized survey data based on a questionnaire grounded in the Experience Sampling Method (ESM) with LMS log data over a 22-day period in a higher education course, and then analyzed it. By integrating subjective and objective data sources (Tempelaar et al., 2020), this study contributes to a more granular understanding of cognitive engagement in higher education and blended learning environment.

2 Related Work

Skinner and Pitzer (2012) conceptualized the widely accepted model of engagement "grounded in self-determination theory, and organized around student engagement and disaffection with learning activities". In that model engagement is examined through three dimensions: behavioral, emotional, and cognitive. The behavioral aspect of engagement involves students' effort, persistence, and determination, especially when facing challenges. Emotional engagement refers to positive feelings such as enthusiasm, enjoyment, fun, and satisfaction, while cognitive engagement "encompasses attention, concentration, focus, absorption, "heads-on" participation, and a willingness to go beyond what is required "(Skinner & Pitzer, 2012).

Blended learning is considered to provide possibility for students to engage more in learning, and Halverson and Graham (2019) have recognized three main challenges regarding engagement in blended learning: (a) the concept of engagement is unclear, (b) blended learning is defined in many different ways, and (c) there is frequent confusion between indicators and facilitators of engagement. They proposed a conceptual framework that consists of indicators of cognitive and emotional engagement and takes into account the dynamics of learning in a blended learning environment.

Another frequently discussed aspect of engagement research is how it is measured. Ortega and Irala (2022) have reviewed various instruments that have been developed to assess student engagement in higher education. Concerning specifically cognitive engagement, Li (2022) provided an overview of various measurement methods of cognitive engagement such as self-report scales, observations, interviews, teacher ratings, experience sampling, eye-tracking, physiological sensors, trace analysis, and content analysis, as well as a summary of the advantages and disadvantages of each approach. The overview (Li, 2022) highlighted ESM as innovative method that measures students' cognitive engagement by asking them to report their thoughts during or closely after learning activities. Although it gives more accurate and timely data than traditional surveys, it can interrupt students, take time, and only allows short questionnaires.

3 Methodology

3.1 Setting and context

The participants in this study were first-year students enrolled in a university-level course, as part of an undergraduate study program conducted in Croatian language at the University of Zagreb. The topic of the course was related to business decision making adapted for information technology students. The course consists of 60 hours, divided equally between lectures (30 hours) and practice-oriented seminars (30 hours). Learning activities include both synchronous and asynchronous components and are supported through the e-course hosted on the institutional LMS.

After each learning activity, students were invited to complete a brief questionnaire. Participation was voluntary, and informed consent was obtained prior to questionnaire completion. Students could submit the questionnaire multiple times if they participated in multiple activities.

In the questionnaire, students were asked to specify the mode of learning (online or in-person), identify the learning activity they had participated in, and indicate the date on which the activity took place. The predefined learning activity options were:

- *lecture*
- *practice-oriented seminar*
- *interaction with colleagues* (e.g., student presentation, discussion, group work, collaborative project)
- *interaction with content* (e.g., reading course materials, reading articles, watching video content)
- *active learning* (e.g., taking knowledge tests, studying for assessment, working independently on a project)
- *other*.

The period for completing self-reports lasted from 7 April 2025 to 28 April 2025. Only complete responses were included in the final analysis, comprising 163 submissions from 140 unique students. After they participated in separate learning activities, responses were

submitted twice by 23 students. The course had a total enrollment of 264 students.

3.2 Cognitive engagement

Cognitive engagement was measured using an adapted version of the ESM survey instrument previously validated in blended learning environment (Manwaring et al., 2017). Linguistic validation of the instrument was carried out in accordance with guidelines for the process of cross-cultural adaptation of self-report measures (Beaton et al., 2000). The survey was submitted for translation to two translators. One translator was familiar with concepts in the field of e-learning and has been a researcher in the area of digital education for over 15 years. Second translator was a professional translator with no prior knowledge of the concepts mentioned in the questionnaire (a translator from a company providing translation service). The translators were asked to include comments during the translation process in case of any ambiguities or additional suggestions.

After receiving the translations, one of authors of this paper, and subsequently a Croatian language proofreader, prepared a proofread version. Based on the proofread version, a back translation into the original language was conducted to verify the accuracy of the translation. The back translations were independently carried out by two certified translators.

Before implementing of the survey instrument, the final version of the questions were selected and additionally reviewed by two authors of this paper. The survey was implemented using LimeSurvey survey software (LimeSurvey Project Team / C. Schmitz, 2024).

The final instrument, translated and adapted from the EMS survey (Manwaring et al., 2017), included the following items:

- learning activity
- date when learning activity occurred
- learning mode
- three items about cognitive engagement:
  - *concentration*
    "How well were you concentrating?"
    (5-point scale, 1 = not at all to 5 = very much)
  - *passive to active*
    "Your involvement in the learning activity:"
    (7-point bi-modal scale)
  - *focused to distracted*
    "Your involvement in the learning activity:"
    (7-point bi-modal scale, opposite-worded).

In addition, supplementary data were collected by the survey software, including the date and time when the questionnaire was completed and the time spent completing it.

Students were asked to complete the self-report either at their own convenience, during lectures, or in practice-oriented seminars. At the beginning of the self-report, there was an introductory note: "People learn in different ways and in different places. Please choose a learning activity you have engaged in recently (or just completed) and fill out this short questionnaire. You can submit reflections for more than one learning activity.".

To ensure comparability, 5- and 7-point scales were rescaled to a 0–100 scale prior to analysis. The oppositely worded item (*focused-distracted*) was reverse-coded (*distracted-focused* to align with the direction of other items.

3.3 Event log data

For each participant who reported an online activity on a specified date, corresponding log data were extracted from the Moodle log data, including: *datetime*, *event context*, *component*, *event name*, *description*, and *origin*. In total, extracted log data contained 707 records.

In the event logs dataset, the verb representing student activity was extracted from the *description* field. For example, in the entry: "The user with id 'X' has submitted the submission with id 'Y' for the assignment with course module id 'Z'," the verb "submitted" was extracted. Similarly, from the *description*: "The user with id 'X' viewed the 'quiz' activity with course module id 'Y','the verb "viewed" was obtained. The *component* field indicates the type of module involved in each activity (e.g., Forum, Quiz, Assignment), while *datetime* contains date and time of event.

A new dataset was created by aggregating event logs dataset by student ID and date. For each student and date, variables *activity*, *componentCount*, and *timeframe* were created.

If the only action student engaged in during a date was the "view" action, *activity* was set to "just view". If the student performed additional actions such as "submitted", "created", "uploaded", "saved", "updated", or "started" an activity, the value "more than view" was used.

For each student and date, the total number of distinct modules on which actions were performed was counted and stored in the variable *componentCount*.

The time spent on LMS was estimated by calculating the difference between the earliest and the latest time of event recorded for the student and date. This value was stored in variable named *timeframe*. It is important to note that this is an approximation, as it does not account for multiple login sessions throughout the day. This limitation arises from the structure of Moodle LMS logging system (Rotelli and Monreale, 2023).

The aggregated event log data and students' responses were merged using a composite key consisting of the variables *IDStudent* and *LADate* (learning activity date). The resulting dataset contained variables *concentration*, *passive-active*, *distracted-focused*, *activity*, *componentCount*, and *timeframe*.

3.4 Data analysis

Descriptive statistics were used to provide an overview of student responses. To compare mean scores between online and in-person learning modes, Welch's t-test was applied due to its robustness to unequal variances.

For correlation analysis, original responses on the 5- and 7-point scales were used, with the distracted–focused item reversed. Relationships among ordinal items were examined using polychoric correlations, and underlying constructs were identified through Principal Component Analysis (PCA) on the polychoric correlation matrix. Polychoric correlation is generally preferred for PCA when variables are ordinal, as it more accurately captures the relationships between the underlying continuous latent variables (Choi et al., 2010; Holgado–Tello et al., 2010). PCA was conducted using the `principal()` function from the *psych* package in R. Components with eigenvalue grater than 1 were retained and submitted to varimax rotation.

K-means cluster analysis was performed to identify distinct student profiles.

The analysis was conducted using R statistical software (R Core Team, 2024) and RStudio (Posit team, 2025), using packages *psych* (Revelle, 2025), *tidyverse* (Wickham et al., 2019), and *ggplot2* (Wickham, 2016).

4 Results

In total, 163 responses were collected from 140 unique students (78% male, 22% female). The number of responses by learning mode and activity is presented in Table 1.

Table 1. Number of responses (n = 163) by reported learning mode and learning activity

Reported learning mode	Reported learning activity
in-person (n=129; 79%)	lecture (n=35)
	interaction with content (n=29)
	active learning (n=26)
	seminar (n=17)
	other (n=16)
	interaction with colleagues (n=6)
online (n=34; 21%)	interaction with content (n=20)
	lecture (n=7)
	active learning (n=5)
	seminar (n=1)
	interaction with colleagues (n=1)

In this study, an event-contingent self-report approach was employed, where students were asked to reflect on their learning activity as soon as possible after its completion. The majority of responses (68.71%) were submitted within 24 hours of the learning activity. On average, students spent 1 minute and 13 seconds completing the questionnaire.

Student responses were aggregated by calculating the mean score per item when multiple responses were available for a single student. As shown in Fig. 1, the majority of students reported positive responses (scores above 50) across all three items (*concentration*, *passive-active*, *distracted-focused*). The *concentration* item had the highest number of scores above 50, suggesting that students generally felt confident in their ability to concentrate. A notable number of students reported the midpoint score of 50, ranging from 31 students (22%) on the *passive-active* item to 46 students (33%) on the *concentration* item. In contrast, scores below 50 were reported by 32 students (23%) on the *passive-active* item and 27 students (19%) on the *distracted-focused* item.

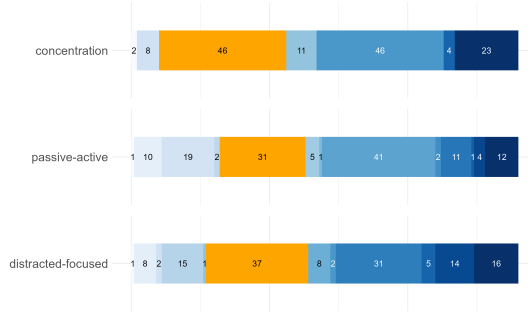


Figure 1. Frequencies of responses from 140 students for each cognitive engagement item, with score 50 highlighted

4.1 RQ1: Cognitive engagement across learning modes and activities

4.1.1 Learning modes

In terms of RQ1, when comparing scores across learning modes (Table 2), students have expressed noticeably higher activity during in-person sessions compared to online (*passive-active* item). Furthermore, Welch's t-test revealed a statistically significant difference in *passive-active* scores between in-person ($\bar{x}_{\text{in-person}} = 63.0$, $s = 23.1$) and online ($\bar{x}_{\text{online}} = 49.1$, $s = 20.1$) learning modes, $t(58.3) = 3.48$, $p = .001$, 95% CI $[5.89, 21.9]$. This suggests that students may feel significantly more actively involved during in-person sessions.

In contrast, scores on the *distracted–focused* item appear consistent across both modes, indicating similar levels of focus regardless of delivery format. Likewise, there was no significant difference in concentration scores between in-person and online modes.

As presented in Table 3, polychoric correlations revealed a moderate positive association between *con-*

Table 2. Mean and standard deviation (mean ± SD) by learning mode

Item	in-person	online
concentration	68.0 (±22.8)	63.2 (±19.7)
passive-active	63.0 (±23.1)	49.1 (±20.1)
distracted-focused	60.1 (±24.0)	60.4 (±22.4)

centration and *passive-active* items (r = 0.43). A weaker positive association was observed between *concentration* and *distracted-focused* items (r = 0.30), as well as between *passive-active* and *distracted-focused* items (r = 0.23). These results suggest that the underlying latent construct measured by these items are positively related, with the strongest relationship observed between *concentration* and *passive-active* items.

Table 3. Polychoric correlation matrix

	con	*pass-act*	*dist-foc*
con	1.00		
pass-act	0.43	1.00	
dist-foc	0.30	0.23	1.00

Fig. 2 shows PCA loadings plot as a result of PCA performed on polychoric correlation matrix. It can be seen that two principal components were retained, together explaining 81.3% of the total variance (RC1: 46.9%; RC2: 34.4%). This result shows that student cognitive engagement in not a unidimensional concept and that *concentration* and *passive-active* load on a different component than *distracted-focused*.

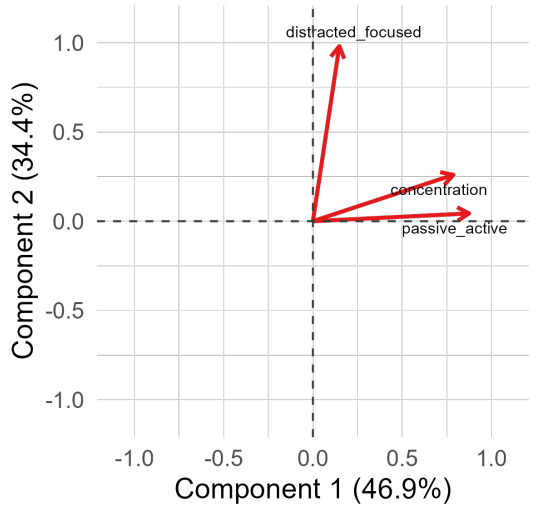


Figure 2. PCA loadings plot based on polychoric correlation matrix

4.1.2 Learning activities

Mean values and confidence intervals for various learning activities are displayed in the error bar plot (Fig. 3). "Interaction with content" (e.g., reading course materials, articles, or watching video content; n = 49) and "lecture" (n = 42) were perceived as the least engaging activities in terms of active involvement (*passive-active*). In contrast, "active learning" (e.g., taking knowledge tests, studying for assessment, working independently on a project, n = 31), was perceived as the most active learning activity. The highest mean concentration was reported for the "interaction with colleagues" activity, but this result is based on only 7 responses, which limits its comparability.

4.2 RQ2: Integration of event log data

Of the 34 reported online activities, 23 responses had corresponding LMS log data and those data were further processed and joined with self-report responses.

Two distinct student groups were identified through k-means cluster analysis (k = 2), as presented in Table 4. The optimal number of clusters was determined by inspecting the scree plot.

Table 4. Variables' means by cluster

Variable	Cluster 1	Cluster 2
activity	1.38	1.70
concentration	69.2	57.5
passive-active	51.2	38.4
distracted-focused	64.2	55.0
componentCount	3.00	4.10
timeframe	54.2	756.2

Cluster 1 represents students who are more cognitively engaged in their activities, but over a shorter period and with fewer components. Cluster 2 includes students with lower cognitive engagement, who participate in more activities and over a much longer timeframe. The distinction between the two clusters is further illustrated in the parallel coordinates plot (Fig. 4). Each line shows normalized variables for a single student. Cluster 1 is represented in black and cluster 2 in red.

5 Conclusions, limitations and future work

5.1 Conclusions

This study explored how students' self-reported cognitive engagement varies across different learning modes and activities in a blended learning environment (RQ1), and examined the extent to which LMS event logs can

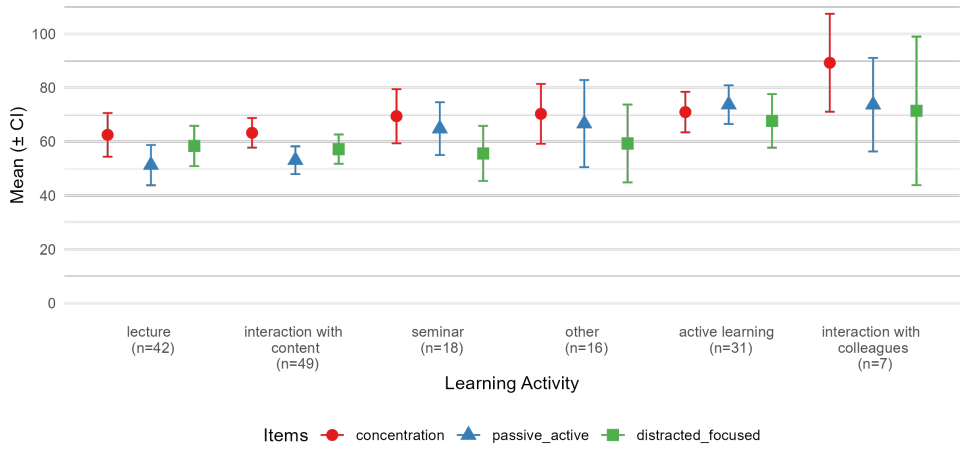


Figure 3. Mean and confidence interval (CI) by learning activity

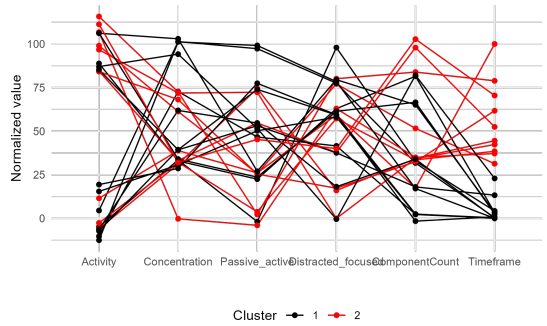


Figure 4. Parallel coordinates chart integrating self-report and event-log data by cluster (n=23)

complement and enrich self-reported cognitive engagement data in online learning settings (RQ2).

For RQ1, the results showed that students reported feeling more actively involved during in-person learning, while concentration and focus levels were similar across in-person and online modes. Among specific learning activities, "active learning" and "interaction with colleagues" were perceived as the most engaging, whereas "lectures" and "interaction with content" were rated lowest for active involvement. Previous studies have reported mixed results regarding levels of cognitive engagement in in-person and online activities (Huang et al., 2022; Manwaring et al., 2017). Our results also support findings that activities involving peer interaction positively influence cognitive engagement (Manwaring et al., 2017). In order to increase cognitive engagement instructional design in blended settings should offer content that requires active interaction.

For RQ2, integrating LMS log data with self-reports provided additional insights into student behavioral engagement. The combined data revealed two distinct student profiles: one group demonstrated higher cognitive engagement over shorter periods and with fewer learning components, while the other showed lower cognitive engagement but participated in more activities over a longer timeframe. This approach highlighted the value of merging behavioral log data with self-reported measures to capture a more comprehensive picture of student engagement in blended learning environments.

5.2 Limitations

The study primarily relied on one-time measurements of cognitive engagement. Although it adopted certain principles of ESM, data were mostly collected at a single time point. Consequently, the design does not support within-person or longitudinal analyses, limiting the findings to between-person comparisons.

Some bias may stem from the measurement instrument used to assess cognitive engagement. To address this, the instrument was translated using a standardized linguistic validation procedure. Another limitation of this study concerns the reliability of single-item measures. While such measures help reduce participant burden in experience sampling, they raise concerns about measurement accuracy. To address this, Dejonckheere et al. (2022) proposed two time-varying test–retest methods to estimate the measurement error of single-item assessments. In this study, the issue of reliability was only partially addressed.

The limited number of responses related to online learning, and consequently fewer LMS log data entries, may have contributed to the lack of observed differences between learning mode and learning activities. Although the number of responses was limited, the approach presented demonstrates the potential for combining self-reported data with LMS log data to capture complementary insights into learner engagement that would be difficult to obtain from a single data source alone.

5.3 Future work

Collecting data as close as possible to the learning activity offers new possibilities for capturing authentic and timely reflections. To enable longitudinal analysis, in future research self-reported data should be collected repeatedly over time, providing opportunities to explore inter-personal patterns and relationships. ESM offers a strong methodological foundation for this approach. Additionally, integrating event log data from various learning support systems could give deeper insights into learner behavior. Recently, Xie et al. (2024) have proposed framework of engagement where they mapped ESM methodologies, technologies, and statistical approaches to capture the characteristics of engagement in educational research. Another, more practical approach was published by Revol et al. (2024) on prepossessing of ESM data. Using these frameworks and tools could improve future studies.

Future research should also tackle Artificial Intelligence (AI) technologies. As identified by A. Nguyen et al. (2024), the use of AI tools in higher education may present new challenge for student engagement, particularly in terms of motivation, autonomy, and meaningful interaction with learning content. These emerging issues and their influence on engagement dimensions should be carefully examined and addressed in future work, highlighting the need for further theoretical conceptualization of engagement in blended learning environment.

Acknowledgments

The authors would like to thank Professor Nikola Kadoić for his support throughout this research.

References

Aragon, C. R., Guha, S., Kogan, M., Muller, M., & Neff, G. (2022). *Human-centered data science: An introduction*. The MIT Press.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures: *Spine*, *25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Bergdahl, N., Bond, M., Sjöberg, J., Dougherty, M., & Oxley, E. (2024). Unpacking student engagement in higher education learning analytics: A systematic review. *International Journal of Educational Technology in Higher Education*, *21*(1), 63. https://doi.org/10.1186/s41239-024-00493-y

Choi, J., Peters, M., & Mueller, R. O. (2010). Correlational analysis of ordinal data: From Pearson's r to Bayesian polychoric correlation. *Asia Pacific Education Review*, *11*(4), 459–466. https://doi.org/10.1007/s12564-010-9096-y

Dejonckheere, E., Demeyer, F., Geusens, B., Piot, M., Tuerlinckx, F., Verdonck, S., & Mestdagh, M. (2022). Assessing the reliability of single-item momentary affective measurements in experience sampling. *Psychological assessment*, *34*(12), 1138–1154. https://doi.org/10.1037/pas0001178

Halverson, L. R., & Graham, C. R. (2019). Learner Engagement in Blended Learning Environments: A Conceptual Framework. *Online Learning*, *23*(2). https://doi.org/10.24059/olj.v23i2.1481

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53. https://doi.org/https://doi.org/10.1016/j.compedu.2015.09.005

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., & Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, *44*(1), 153–166. https://doi.org/10.1007/s11135-008-9190-y

Huang, M., Kuang, F., & Ling, Y. (2022). EFL learners' engagement in different activities of blended learning environment. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(1), 9. https://doi.org/10.1186/s40862-022-00136-7

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, *68*(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Li, S. (2022). Measuring cognitive engagement: An overview of measurement instruments and techniques. *International Journal of Psychology and Educational Studies*, *8*(3), 63–76. https://doi.org/10.52380/ijpes.2021.8.3.239

LimeSurvey Project Team / C. Schmitz. (2024). *Limesurvey: An open source survey tool*. LimeSurvey Project. Hamburg, Germany. https://www.limesurvey.org

Lowe, T., & El Hakim, Y. (2020). *A handbook for student engagement in higher education: Theory into practice*. https://doi.org/10.4324/9780429023033

Manwaring, K. C., Larsen, R., Graham, C. R., Henrie, C. R., & Halverson, L. R. (2017). Investigating student engagement in blended learning settings using experience sampling and structural equation modeling. *The Internet and Higher Education*, *35*, 21–33. https://doi.org/https://doi.org/10.1016/j.iheduc.2017.06.002

Nguyen, A., Kremantzis, M., Essien, A., Petrounias, I., & Hosseini, S. (2024). Editorial: Enhancing Student Engagement Through Artificial Intelligence (AI): Understanding the Basics, Opportunities, and Challenges. *Journal of University Teaching and Learning Practice*, *21*(06). https://doi.org/10.53761/caraaq92

Nguyen, Q., Rienties, B., & Whitelock, D. (2022). Informing learning design in online education using learning analytics of student engagement. In B. Rienties, R. Hampel, E. Scanlon, & D. Whitelock (Eds.), *Open World Learning* (pp. 189–207). Routledge. https://doi.org/10.4324/9781003177098-17

Nguyen, T. D., Cannata, M., & Miller, J. (2018). Understanding student behavioral engagement: Importance of student interaction with peers and teachers. *The Journal of Educational Research*, *111*(2), 163–174. https://doi.org/10.1080/00220671.2016.1220359

Ortega, F. d. C., & Irala, V. B. (2022). The measurement of student online engagement of higher education: A scoping review in the international literature. *Texto Livre: Linguagem e Technologia*, *15*. https://doi.org/10.35699/1983-3652.2022.35747

Posit team. (2025). *Rstudio: Integrated development environment for R*. Posit Software, PBC. Boston, MA. http://www.posit.co/

Prananto, K., Cahyadi, S., Lubis, F. Y., & Hinduan, Z. R. (2025). Perceived teacher support and student engagement among higher education students – a systematic literature review. *BMC Psychology*, *13*(1). https://doi.org/10.1186/s40359-025-02412-w

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Revelle, W. (2025). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. Evanston, Illinois. https://CRAN.R-project.org/package=psych

Revol, J., Carlier, C., Lafit, G., Verhees, M., Sels, L., & Ceulemans, E. (2024). Preprocessing experience-sampling-method data: A step-by-step framework, tutorial website, R package, and reporting templates. *Advances in Methods and Practices in Psychological Science*, *7*(4). https://doi.org/10.1177/25152459241256609

Rotelli, D., & Monreale, A. (2023). Processing and Understanding Moodle Log Data and Their Temporal Dimension. *Journal of Learning Analytics*, *10*(2), 126–141. https://doi.org/10.18608/jla.2023.7867

Skinner, E. A., & Pitzer, J. R. (2012). *Developmental dynamics of student engagement, coping, and everyday resilience* (S. L. Christenson, A. L. Reschly, & C. Wylie, Eds.). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_2

Tani, M., Gheith, M. H., & Papaluca, O. (2021). Drivers of student engagement in higher education: A behavioral reasoning theory perspective. *Higher education*, *82*(3), 499–518. https://doi.org/10.1007/s10734-020-00647-7

Tempelaar, D., Rienties, B., & Nguyen, Q. (2020). Subjective data, objective data and the role of bias in predictive modelling: Lessons from a dispositional learning analytics application. *PLOS One*, *15*(6). https://doi.org/10.1371/journal.pone.0233977

Tight, M. (2020). Student retention and engagement in higher education. *Journal of Further and Higher Education*, *44*(5), 689–704. https://doi.org/10.1080/0309877X.2019.1576860

Vo, X. M., Do-Hong, C., Do, T. H. L., Ha, T. M. T., & Vu, C. T. (2024). An examination of blended learning in higher education over a two-decade period (2003-2022): Insights derived from scopus database. *Başlık*, *13*(4), 1821–1840. https://doi.org/10.12973/eu-jer.13.4.1821

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wong, Z. Y., Liem, G. A. D., Chan, M., & Datu, J. A. D. (2024). Student engagement and its association with academic achievement and subjective well-being: A systematic review and meta-analysis. *Journal of Educational Psychology*, *116*(1). https://doi.org/10.1037/edu0000833

Xie, K., Vongkulluksn, V. W., Heddy, B. C., & Jiang, Z. (2024). Experience sampling methodology and technology: An approach for examining situational, longitudinal, and multi-dimensional characteristics of engagement. *Educational technology research and development*, *72*(5), 2585–2615. https://doi.org/10.1007/s11423-023-10259-4

Exploring the Impact of Digital Infrastructure Challenges on Learning Outcomes and Student Satisfaction

Enrique Mu
Carlow University
Department of Business Management,
Accounting and Ethics
3333 Fifth Avenue, Pittsburgh, PA 151213
USA
emu@carlow.edu

Anna Florek-Paszkowska
CENTRUM/PUCP
Jr. Daniel Alonía Robles 125
Santiago de Surco 15023
PERU
aflorekpaszkowska@pucp.edu.pe

Santiago Luis Risco-Martinez
Texas A&M University
Department of Economics
2935 Research Pkwy Suite 200, College Station, TX 77843
USA
santiagoris93@tamu.edu

Abstract. *The present study explores the validity of two key dimensions proposed by the Framework for the Assessment of Challenges to Virtual Education (FACVE) in supporting the effective delivery of virtual instruction. Specifically, it examines: (1) C2 — Challenges to Digital Technical Infrastructure (i.e., Connectivity and Equipment), and (2) C5 — Challenges to Digital Learning Infrastructure (i.e., Learning Platforms and Access to Resources). Findings from this exploratory study indicate that C2 is negatively associated with Student Satisfaction (SS), while C5 exhibits a negative relationship with Learning Outcomes (LOs), which, in turn, significantly predicts SS. LOs themselves emerged as a strong predictor of higher levels of SS. Although both C2 and C5 were statistically significant predictors of SS and LOs, the effect sizes were very small. This study represents an initial step toward linking challenges identified by the FACVE framework with well-established constructs in educational research, such as LOs and SS.*

Keywords. FACVE, challenges to virtual education, connectivity and equipment, learning platform, access to resources, learning outcomes, student satisfaction, digital infrastructure, structural equation modelling.

1 Introduction

The framework for the assessment of challenges to virtual education (FACVE) is a new framework developed during the emergency remote teaching (ERT) context of the COVID-19 pandemic (Mu et al., 2022). During this period, most institutions worldwide had to switch their education practices to a virtual education model, which refers to a broad approach to education that uses digital environments for learning and instruction, primarily or entirely online (Allen et al., 2016). FACVE was developed to assess the extent of challenges faced by students and instructors to carry on the new approach in a successful way.

While still a novel technique, FACVE has started to become recognized as a valuable tool to assess virtual education challenges, in particular in developing countries; given that it encompasses aspects such as determined home facilities, difficulties in learning virtual platforms and financial problems as significant factors that challenge virtual instruction (Vargas-Hernández et al., 2024). Also, FACVE highlighted that the pandemic had negative consequences on student life, associated with their subjective well-being and allowed some extent of quantification (Vargas-Hernández et al., 2024). Overall, FACVE is a powerful tool to assess challenges to virtual education in ERT contexts (Mu et al., 2022).

As shown in Fig. 1, FACVE consists of the following dimensions C1 - Quality of Teaching Instruction, C2 - Connectivity and Equipment, C3 - Personal Issues (student), C4 - Home Infrastructure and Study Environment, C5 - Learning Platform and Access to Resources, C6 - Financial Issues (student), and C7 - University Administration and Costs. This framework was developed from the bottom up; that is, it was developed from the challenges expressed by the

students struggling to adapt to a sudden switch from physical instruction to a virtual education format. Also, the assessed challenges include some that are particularly suitable for the ERT situation in which the framework was originally developed. While the challenge dimensions reflect the practical reality of the students and have even been used to compare situations in different international settings (Gonzalez-Urango et al., 2025), these variables have not been formalized or tested within the context of normal virtual instruction (non-ERT) or connected to the virtual instruction research stream. This constitutes the research gap to be addressed in the present study.

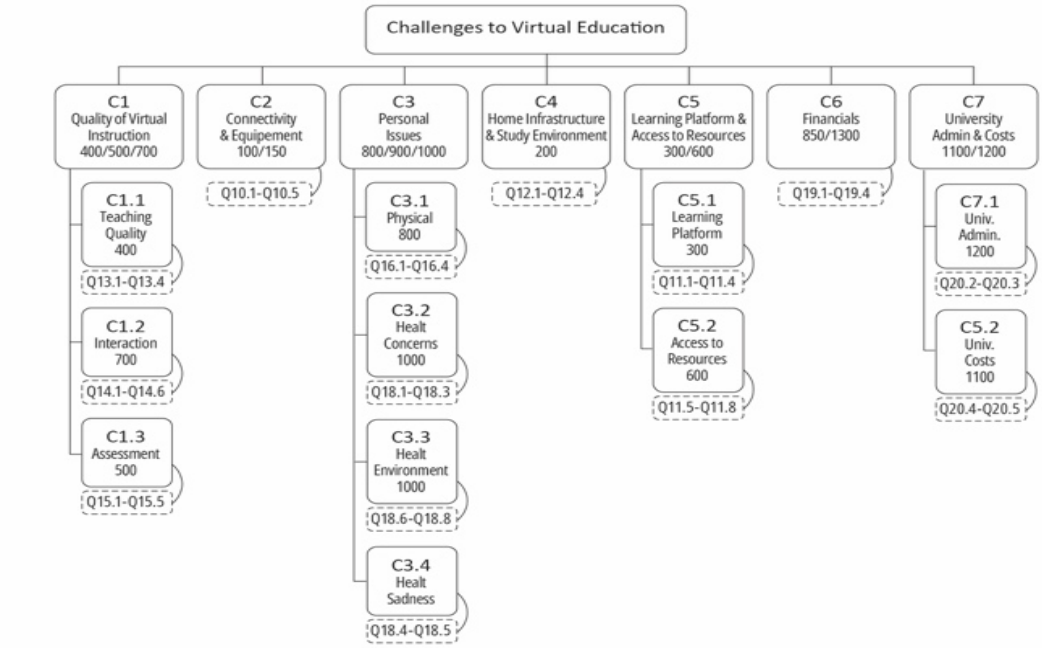


Figure 1. FACVE: Challenges to virtual education
Source: (Mu et al., 2022)

The above rationale suggests the following research objectives:

1. To determine the feasibility of using FACVE in a normal (beyond ERT) virtual instruction context.
2. To determine if FACVE challenge variables have an effect on well-known dependent outcomes of virtual instruction research.

To address the first objective, the focus should be on testing only FACVE dimensions more commonly related to normal (non-emergency) virtual instruction operations. To address the second objective, FACVE dimensions (variables) expected to have an effect on well-known virtual instruction variables should be selected.

For these reasons, this study has chosen to explore if the following FACVE variables: C2 — Connectivity and Equipment and C5 — Learning Platform and Access to Resources have an effect on the extent of Learning Outcomes (LO) and Student Satisfaction (SS), two well-known dependent variables in extant virtual education research (Eom & Ashill, 2016; Kuo et al., 2014).

Also, the present study posits that variables C2 and C5 constitute challenges to the virtual education digital infrastructure. Digital infrastructure includes everything that makes digital systems work (technical infrastructure) such as the internet, devices, and software (e.g., learning management systems). However, this doesn't just include technology, but also how people use these resources (learning infrastructure) (Bygstad et al., 2022; MagicEdTech, 2025). Based on this, we posit that variable C2 of FACVE constitutes challenges to the digital "technical" infrastructure while variable C5 of FACVE constitutes challenges to the digital "learning" infrastructure. For this reason, this study is proposed as an exploratory impact of digital infrastructure challenges on virtual education outcomes (Goodhue & Thompson, 1995; McGill & Klobas, 2008).

2 Literature Review

FACVE was proposed as a way to assess students' main challenges in switching to virtual instruction during the COVID-19 pandemic (Mu et al., 2022) as shown in Fig. 1.

Exploring the relation of the digital infrastructure with the proposed outcome variables requires the variables to be defined and the rationale for their expected effects to be explained.

Dependent Variables

SS is widely recognized in the educational literature as a key measure of academic quality, often reflecting the degree of instructional effectiveness and learner engagement (Eom & Ashill, 2016; Kuo et al., 2014). Higher satisfaction levels frequently correlate with stronger academic outcomes, including increased retention and performance. Likewise, LOs—defined as the cumulative knowledge, skills, and competencies developed by students—are essential benchmarks for evaluating educational success (Alqurashi, 2019).

SS is defined as the learner's perception of value and contentment with the educational experience, including its instructional, social, and technological aspects (Martin & Bolliger, 2022).

LOs are defined as the extent to which learners gain new knowledge and skills, and is often aligned with clear and measurable instructional goals. In summary, the success of the online learning depends on whether it has achieved the desired outcomes (Panigrahi et al., 2018).

Independent FACVE variables

The two target FACVE variables for the present study are: C2 — Connectivity and Equipment and C5 — Learning Platform and Access to Resources (Fig. 1).

Connectivity and Equipment (C2) challenges can be defined as the challenges related to the suitable availability of the proper *technical* digital infrastructure, in terms of connectivity (e.g. proper internet access) and the availability of suitable equipment for both the students and the educational institution participating in the virtual instruction.

Learning Platform and Access to Resources (C5) challenges is constituted by the challenges related to the *learning* digital infrastructure along two sub-dimensions: availability and usability of the learning platform (C5.1) and the challenges related to access to educational resources (C5.2); that is the learning digital material itself.

While the above variables refer to challenges related to the "technical" and "learning" digital infrastructure, the analysis will be conducted referring to the original FACVE names to maintain simplicity and consonance with the specific assessment tool used in this study. The purpose of the present research is to assess whether these two challenge variables (C2 and C5) also have an impact on the well-known LOs and SS from the extant virtual education research literature.

The rationale is grounded, theoretically, on the Technology-Task mediated theory which has shown that Task–Technology fit is one factor that has been shown to influence both the use of the digital infrastructure and its learning impact (Goodhue & Thompson, 1995; McGill & Klobas, 2008).

Based on the above, our research question will be as follows:

Research question: Do FACVE challenges C2 - Connectivity and Equipment and C5 - Learning Platform and Access to Resources on Learning Outcomes (LO) and Student Satisfaction (SS)?

This research question can be illustrated in the conceptual model proposed in Fig.2.

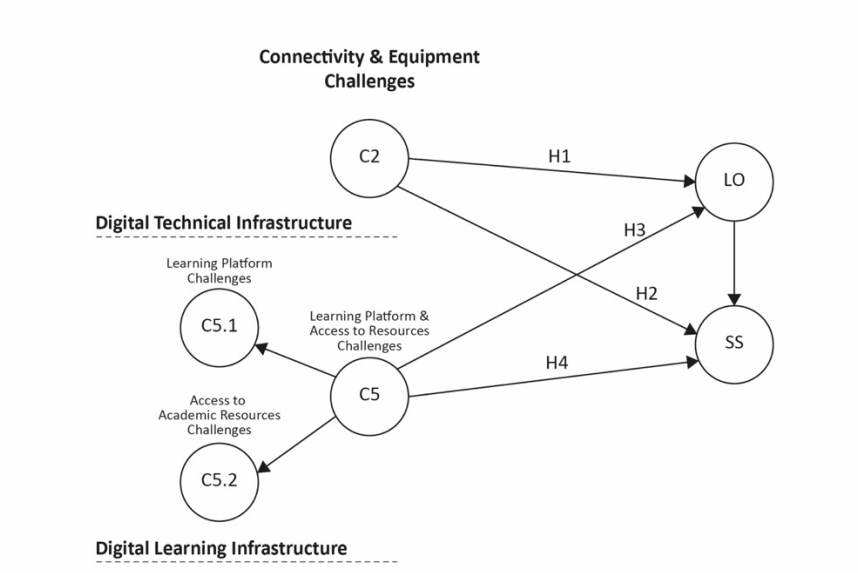


Figure 2. Conceptual model

Consistent with the research question and conceptual model (Fig. 2), the following hypotheses are proposed:

H1: C2 is negatively associated with LO.
H2: C2 is negatively associated with SS.
H3: C5 is negatively associated with LO.
H4: C5 is negatively associated with SS.
H5: LOs are positively associated with student SS.

3 Methodology

A FACVE survey was administered to graduate students enrolled in various Master of Business Administration (MBA) courses at a prominent private university in Peru. It was conducted through the university's online platform, during the period September-December, 2021.

3.1. Demographics

Participants were students enrolled in different synchronous online courses of the evening professional Master of Business Administration (MBA) study program. The survey was conducted by one of the authors, and the digital infrastructure consisted of Canvas as the learning platform system and Zoom as the video conferencing tool (Instructure, 2025; Zoom, 2025). Also, MBA students access to academic resources consisted fundamentally of access to the class presentations and assigned readings.

Participation was entirely voluntary and anonymous, aligning with ethical research standards requested by the home institution. Out of 177 initially submitted responses, 162 were deemed valid for analysis. The demographic data shows that 98.31% of respondents were pursuing graduate-level studies. The gender distribution was predominantly male (59.89%), followed by female participants (29.94%) and a smaller group who chose not to disclose their gender (10.17%). In terms of age, a significant portion of respondents were experienced professionals: 45.20% were between 30 and 34 years old, and 36.16% were older than 34. With respect to instructional formats, the majority (54.80%) attended synchronous online classes, while 28.81% were enrolled in hybrid modalities. A minority participated in asynchronous virtual (2.26%) or traditional face-to-face classes (1.69%). Furthermore, 79.10% of respondents reported active enrollment in at least one virtual course, highlighting the strong relevance of online education within this academic cohort.

3.2. Measures

Testing the nomological validity of the proposed research model (Fig. 1) involves the development of reliable scales. For this purpose, all the measurement items used in the present study were taken from previous studies, to ensure their validity and reliability (Eom & Ashill, 2016; Mu et al., 2022)

3.3. Analysis

Partial Least Squares Structural Equation Modeling (PLS-SEM) was used to analyze the proposed research model, PLS-SEM is particularly useful when the goal is to predict either key determinant (C2 & C5) or target (LO & SS) constructs (Hair et al., 2017). The analysis followed the recommended steps for models that include reflective-reflective HOC constructs, using the disjointed two-stage approach (Chua, 2024; Hair Jr. et al., 2018; Sarstedt et al., 2019)[1]. The analysis consisted of the following steps:

Step 1. Validity and reliability of reflective first-order constructs. The validity and reliability of the reflective first-order model constructs (C2, C5.1, C5.2, LO, and SS) were analyzed (Fig. 1). Because of high Variance Inflation Factors (VIF) values, two items were removed from SS. All indicators included in the model demonstrated loadings above 0.7 ($p < 0.005$), with the exception of the indicator Q10-3 for C2 (0.50) thus it was removed from the model. Average Variance Extracted (AVE) values for all constructs were above the threshold of 0.5, thus satisfying convergent validity for all constructs.

To meet the construct reliability acceptable threshold (>0.70), item Q11-1 was removed from C5.1. All latent constructs then exhibited both Cronbach's α and composite reliability (rho_c) values above the 0.70 threshold. Although the SS construct's rho_c just exceeds the 0.95 ceiling recommended by some authors, treating rho_c as the upper-bound and α as the lower-bound estimate suggests that its true reliability still remain within acceptable limits (Hair et al., 2017).

An initial HTMT assessment revealed discriminant-validity concerns between LOs and SS. For this reason, one indicator from LOs was removed to meet discriminant validity requirements according to best practices for this type of analysis (Hair et al., 2017). VIF values were below the upper threshold of 5, suggesting that, while possible, collinearity issues

[1] C5 is a Higher Order Construct (HOC) in the model, composed of C.5.1 and C.5.2. The software used for this study was SmartPLS 4 Ringle, C. M., Wende, S., & Becker, J.-M. (2024). *SmartPLS 4*. Bönningstedt: SmartPLS. . https://www.smartpls.com

are unlikely (Hair et al., 2019). Table 1 summarizes these results.

Table 1. First stage model validation results

LOC Constructs and Indicators[1]	Loadings	Cronbach's alpha	rho_c	AVE
C.2. Connectivity & Equipment		0.838	0.885	0.660
Q10-1	0.800			
Q10-2	0.841			
Q10-4	0.892			
Q10-5	0.705			
C5.1. Learning Platform		0.746	0.852	0.659
Q11-2	0.731			
Q11-3	0.886			
Q11-4	0.812			
C5.2. Access to Resources		0.828	0.880	0.647
Q11-5	0.752			
Q11-6	0.743			
Q11-7	0.858			
Q11-8	0.858			
Learning Outcomes (LO)		0.870	0.919	0.792
Q8-2	0.874			
Q8-3	0.916			
Q8-4	0.879			
Student Satisfaction (SS)		0.906	0.955	0.914
Q9-1	0.959			
Q9-3	0.954			

[1] These model validations apply only to Low Order Constructs (LOC) constructs.

Step 2: Validity of Formative Higher-Order Construct (C5). In the second stage, following Sarstedt et al. (2019), we assessed the higher-order model. Items C5.1 and C5.2 loaded significantly on the second-order construct C5, with loadings of 0.652 and 0.987 ($p < .01$). The construct attained an AVE of 0.686. It demonstrated acceptable internal consistency—Cronbach's $\alpha = 0.686$ and composite reliability (rho_c) = 0.817.

Step 3: Structural Model Evaluation. The structural model was evaluated to determine if the proposed hypotheses were supported. The results are presented in Fig. 3 and summarized below:

1. C2 negatively predicts SS ($\beta = -0.106$, $p < 0.05$, $f^2 = 0.026$).
2. C5 negatively predicts LO ($\beta = -0.258$, $p < 0.01$, $f^2 = 0.063$).
3. LO positively predicts SS ($\beta = 0.764$, $p < 0.01$, $f^2 = 1.453$).

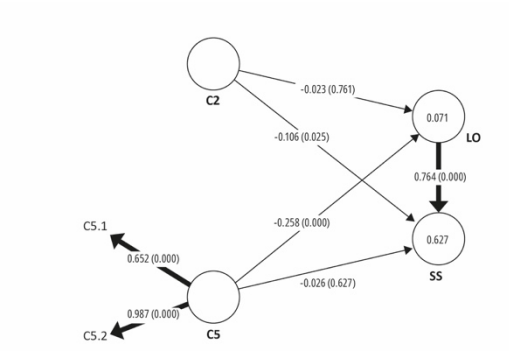


Figure 3. Model PLS-SEM results

The model explains 7.1% ($R^2 = 0.058$) of the variance in LO and 62.7% ($R^2 = 0.627$) of the variance in SS. Regarding the model fit, the standardized root mean square residual (SRMR) of the model was 0.086, which is above the conservative threshold of 0.08 but still within the acceptable level of below 0.1 (Hair, 2017). Positive Q^2 Predict values for both dependent variables further supported predictive power, indicating strong predictive relevance. The overall model predictive power is Q2 = 0.043 for LO and Q2=0.055 for SS.

Although H4 was not supported, the analysis did reveal a significant indirect effect of C5 on SS through LO ($\beta = -0.197$, $p < 0.01$), indicating that C5 might influences SS only via its impact on LOs.

4 Results

The results of the present study are shown in Table 2.

Table 2. PLS-SEM analysis results

H	Beta	f^2	R^2 (LO)	R^2 (SS)
H1	-0.023	0.000	0.071	
H2	-0.106*	0.026		0.627
H3	-0.258**	0.063	0.071	
H4	-0.026	0.001		0.627
H5	0.764**	1.453**		0.627

* $p < 0.05$.
**$p < 0.01$

For greater clarity, the results of hypothesis testing are summarized in Table 3 below.

Table 3. Hypothesis results

Hypothesis	Result
H1: C2 has a negative effect on LO	ns
H2: C2 has a negative effect on SS	**Supported**
H3: C5 has a negative effect on LO	**Supported**
H4: C5 has a negative effect on SS	ns*
H5: LO has a positive effect on SS	**Supported**

*However, it has significant indirect effect on SS through LO

The PLS-SEM results (Tables 2 and 3) show that two dimensions in the FACVE checklist—C2 and C5 have statistically detectable but small links with LO and SS. While only the C2-SS and C5-LO paths reach significance (although with no statistically significant effect sizes), this pattern still helps identify where technology-related issues may start to influence the student experience. In that sense, the framework offers a useful, if preliminary, guide for monitoring online-learning conditions; further work with larger samples will be needed to confirm and extend these findings.

4.1 Validation of FACVE-C2 and FACVE-C5 as Valid Multidimensional Constructs

The FACVE-C2 construct was found to be reliable and showed a statistically significant—but small—negative association with SS. Likewise, FACVE-C5 was validated as a second-order construct with two sub-dimensions (C5.1 learning-platform challenges; C5.2 access-to-resources challenges) and exhibited a statistically significant, though negligible, negative link with LO.

Taken together, these initial findings suggest that the FACVE checklist can flag technology-related issues that may begin to influence LO and SS, even if their direct effects are modest. The direct paths C2 → LO and C5 → SS were not supported; however, a significant indirect effect of C5 on SS via LO indicates full mediation. The findings align with the conceptual foundation of FACVE, which emphasizes a holistic, student-centered approach to addressing virtual education challenge barriers (Mu et al., 2022). Compared to other models aimed at assessing students' perceptions of digital tools and engagement only during emergency learning, the FACVE framework extends the focus to the quality of virtual instructional practices.

Thus, the framework offers a useful early lens for monitoring virtual-learning conditions, while further studies with larger samples are needed to confirm its explanatory power and practical impact.

4.2 Impact on Learning Outcomes

This study shows that C5 shows a statistically significant negative association with LO ($\beta = –0.258$, $p < 0.01$). The corresponding effect size is small ($f^2 = 0.063$), yet the link is robust enough to suggest that shortcomings in the platform or in access to digital resources can dampen students' academic performance. Also, because LO strongly predicts SS (described in the section below), this path also establishes an indirect route by which C5 affects satisfaction.

4.3 Impact on Student Satisfaction

This study shows that C2 has a small but significant negative effect on SS ($\beta = –0.106$, $p < 0.05$, $f^2 = 0.026$). This result suggests that the FACVE checklist can help draw attention to basic connectivity issues, making it worth monitoring the C2 dimension to ensure SS and, in turn, LO.

4.4 Relationship between Learning Outcomes and Student Satisfaction

LOs display a large, positive impact on SS ($\beta = 0.764$, $p < 0.01$, $f^2 = 1.453$). This sizeable effect underscores the centrality of academic achievement in shaping students' overall course evaluations. This finding aligns with the research conducted by Baber (2020) and Eom and Ashill (2016), which identifies academic achievement as a key driver of student satisfaction. This interconnected process contributes to overall virtual instruction success.

4.5 Contextual Implications

The study's findings contribute to the growing literature on virtual education by demonstrating FACVE's applicability beyond emergency remote teaching contexts. While frameworks like Rubtsova et al. (2023) and Kasperski et al. (2023) focus on crisis-specific environments, FACVE offers an assessment framework that is useful for standard virtual education. Its student-centered, bottom-up approach addresses socio-economic and infrastructural barriers, making it particularly relevant for developing countries.

4.6 Theoretical and Practical Contributions

Theoretically, this study confirms FACVE as a valid and robust virtual instruction construct consistent with the extant virtual instruction education literature. Also, this study shows the strong impact of LO on SS, consistent with the extant literature. It has been argued that when students feel they are acquiring meaningful knowledge and skills, their overall satisfaction with the course increases (Eom & Ashill, 2016; Kuo et al., 2014). However, the positive correlation can also be explained in the other direction, by arguing that satisfaction influences students' motivation to engage and thus achieve better learning outcomes (Alqurashi, 2019). Still, the current research results suggest that the first explanation may be the most plausible.

From a methodological point of view, this study reveals strong collinearity between LO and SS, at least when using the existing measures commonly used in educational research as in the classic study by Eom and Ashill (2016).

From a practical standpoint, it validates FACVE-C2 and FACVE-C5 dimension assessment as a powerful tool to foster better LO and SS. These findings align with and extend the practical implications noted by Baruth et al. (2021), Cleary et al. (2024), and Camacho-Zuñiga et al. (2023), providing a comprehensive lens for future interventions

5 Conclusions and Future Work

The FACVE framework was developed in the context of emergency remote teaching during the COVID-19 pandemic. The present study was conducted during the period September-December, 2021 and the studied FACVE challenges, related to digital infrastructure, were found to still have an impact -with different degrees of intensity- on educational outcome variables (SS and LO). More specifically, for this study the results obtained can be summarized as follows:

For digital *technical* infrastructure,

- Challenges related to digital technical infrastructure (C2) are negatively associated with SS.
- Digital technical infrastructure challenges (C2) consistently show a negative association with SS.

For digital *learning* infrastructure,

- Challenges related to digital learning infrastructure (C5) are negatively associated with LO, although the effect size is negligible.

Also, it was found that LO are positively associated with SS.

Limitations and Future Research: The results shown in this study constitute an exploratory study of the inter-relationship of the research variables. While the relations are meaningful and promising, the overall predictive power of the model is small (Q2 = 0.043 for LO and Q2=0.055 for SS), *suggesting the necessity of further studies with larger samples*. Also, the effect of the access to (academic) resources challenges (C5.2) may have been strongly attenuated by the fact that MBA students' academic resources were constituted mainly of reading material. The needed academic resources in other disciplines (e.g., chemistry or nursing) may be more complicated to have available (e.g., laboratory). *Testing the proposed relations with students in various disciplines* would constitute another logical future research step.

Acknowledgments

The authors would like to thank the anonymous reviewers for their meticulous work, which greatly helped improve this manuscript.

References

Allen, I. E., Seaman, J., Pulin, R., & Straut, T. T. (2016). *ONLINE REPORT CARD*: Tracking Online Education in the United States. B. S. R. Group.

Alqurashi, E. (2019). Predicting student satisfaction and perceived learning within online learning environments. *Distance Education*, 40(1), 133-148.

Baber, H. (2020). Determinants of students' perceived learning outcome and satisfaction in online learning during the pandemic of covid-19. *Journal of Education and E-Learning Research*, 7(3), 285-292. https://doi.org/https://doi.org/10.20448/journal.509.2020.73.285.292

Baruth, O., Gabbay, H., Cohen, A., Bronshtein, A., & Ezra, O. (2021). Distance learning perceptions during the coronavirus outbreak: Freshmen versus more advanced students. *Journal of Computer Assisted Learning, 37(*6), 1666-1681. https://doi.org/https://doi.org/10.1111/jcal.12612

Bygstad, B., O'vrelid, E., Ludvigsen, S., & Daehlen, M. (2022). From dual digitalization to digital learning space: Exploring the digital transformation

of higher education. *Computers & Education*, 182, Article 104463.

Camacho-Zuñiga, C., Peña-Becerril, M., Cuevas-Cancino, M. d. l. O., & Avilés-Rabanales, E. G. (2023). Gains from the transition from face-to-face to digital/online modality to improve education in emergency. *Frontiers in Education*, 8. https://doi.org/https://doi.org/10.3389/feduc.2023.1197396

Chua, Y. P. (2024). *A step-by-step Guide PLS-SEM Data Analysis using SMART PLS 4. KL*: Researchtree.

Cleary, K., Samarawickrema, G., Ambler, T., Loton, D., Krcho, T., & McCluskey, T. (2024). Transitioning to emergency remote teaching In A block model curriculum: A case study of academics' experiences in an Australian University. *British Journal of Educational Studies*, 72(1), 63-84. https://doi.org/https://doi.org/10.1080/00071005.2023.2248289

Eom, S., & Ashill, N. (2016). The Determinants of Students' Perceived Learning Outcomes and Satisfaction in University Online Education: An Update. *Decision Sciences Journal of Innovative Education,* 14(2).

Gonzalez-Urango, H., Mu, E., Florek-Pazszkowska, A., & Pereyra-Rojas, M. (2025). Beyond Boundaries: Validating a Comprehensive Framework for Assessing Challenges to Virtual Education (FACVE) - A Comparative Insight from Peru and Spain [Manuscript Under Review].

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, 213-236.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd Ed. ed.). SAGE.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to Use and How to Report the Results of PLS-SEM. *European Business Review*, 31(1), 2-24. https://doi.org/10.1108/EBR-11-2018-0203

Hair Jr., J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). *Advanced Issues in Partial Least Squares Structural Equation Modeling*. Sage Publications, Inc.

Instructure. (2025). *Instructure*. https://www.instructure.com/canvas

Kasperski, R., Porat, E., & Blau, I. (2023). Analysis of emergency remote teaching in formal education: Crosschecking three contemporary techno-pedagogical frameworks. *Research in Learning Technology*, 31. https://doi.org/https://doi.org/10.25304/rlt.v31.2982

Kuo, Y.-C., Walker, A. E., Schroder, K. E. E., & Belland, B. R. (2014). Interaction, Internet, self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *The Internet and Higher Education* (20), 35-50.

MagicEdTech. (2025). *Digital Learning for Everyone.* https://www.magicedtech.com/

Martin, F., & Bolliger, D. U. (2022). Developing an online learner satisfaction framework in higher education through a systematic review of research. *Int J Educ Technol High Educ*, 19(50). https://doi.org/https://doi.org/10.1186/s41239-022-00355-5

McGill, T. J., & Klobas, J. E. (2008). A task-technology fit view of learning management system impact. *Computers & Education*, 52, 496-508.

Mu, E., Florek-Pazkowzka, A., & Pereyra-Rojas, M. (2022). Development of a Framework to Assess Challenges to Virtual Education in an Emergency Remote Teaching Environment: A Developing Country Student Perspective-The Case of Peru. *Education Sciences*, 12(10). https://doi.org/https://doi.org/10.3390/educsci12100704

Panigrahi, R., Srivastava, P. R., & Sharma, D. (2018). Online learning: Adoption, continuance, and learning outcome—A review of literature. *International Journal of Information Management*, 43, 1-14.

Ringle, C. M., Wende, S., & Becker, J.-M. (2024). SmartPLS 4. Bönningstedt: *SmartPLS*. https://www.smartpls.com

Rubtsova, A., Semenova, N., Kats, N., & Zheleznyakova, O. (2023). Instructional framework for emergency remote teaching in higher education. *Education Sciences*, 13(6). https://doi.org/https://doi.org/10.3390/educsci13060613

Sarstedt, M., Hair, J. P., Cheah, J., & Becker, J. (2019). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australasian Marketing Journal*, 27, 197-211.

Vargas-Hernández, A., Robledo, S., & Rojas-Quiceno, G. (2024). Virtual Teaching for Online Learning from the Perspective of Higher Education: A Bibliometric Analysis. *Journal of Scientometric Research*, 13(2), 406-418. https://doi.org/10.5530/jscires.13.2.32

Zoom. (2025). *Zoom.* https://www.zoom.com/.

Appendix

I. DIGITAL TECHNICAL INFRASTRUCTURE

C2 – Connectivity & Equipment (Mu et al., 2022)
Q10.1 I do not have (or have limited) access to the internet where I live.
Q10.2 My internet speed is not adequate for my classes.
Q10.3 I do not have access (or it is rather limited) to a computer at home.
Q10.4 There are many technical problems while accessing classes or study material.
Q10.5 My educational institution does not have the appropriate computer equipment (e.g., servers) for virtual teaching.

II. DIGITAL LEARNING INFRASTRUCTURE

C5 – Learning Platform & Access to Resources (Mu et al., 2022)

C5.1 – Learning Platform

Q11.1 The educational platform in use is not suitable for virtual instruction.
Q11.2 Teachers do not know how to use the platform.
Q11.3 Students do not know how to use the platform.
Q11.4 There is no information about the use of the platform.

C5.2 – Access to (Academic) Resources

Q11.5 Lack of access to library books is a severe limitation
Q11.6 Lack of access to laboratories is a problem.
Q11.7 It is necessary to have access to more study material (e.g., PPTs) in addition to the recordings of the class.
Q11.8 Access to teaching resources is less in virtual instruction.

III. DEPENDENT VARIABLES

LO - Learning Outcomes (Eom & Ashill, 2016)

Q8.1 The academic quality of this online class is on par with the face-to-face classes I've taken.
Q8.2 I have learned as much from this online class as I might have from a face-to-face version of the course.
Q8.3 I learn more in online classes than in face-to-face classes.
Q8.4 The quality of the learning experience in online classes is better than in face-to-face classes.

SS – Student Satisfaction (Eom & Ashill, 2016)

Q9.1 I would recommend this instructor to other students.
Q9.2 I would recommend this online class to other students.
Q9.3 I would take an online class at this university again in the future.
Q9.4 I was very satisfied with this online class.

Application of Multi-Objective Optimization for Short-Term Operation Planning of an Energy Community

Sanja Kelemen, Nemanja Mišljenović, Goran Knežević
Josip Juraj Strossmayer University of Osijek
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Kneza Trpimira 2B, 31000 Osijek, Croatia
{sanja.kelemen, nemanja.misljenovic, goran.knezevic}@ferit.hr

Stjepan Babić
Croatian Transmission System Operator Plc., Transmission Area Osijek
Relay protection and control department
Vukovarska cesta 217, 31000 Osijek, Croatia
stjepan.babic@hops.hr

Abstract. *The transition to clean energy increases the importance of Energy Communities (ECs) in achieving sustainability objectives. Effective operational strategies are crucial for ECs to realize their full potential, striking a balance between economic viability and environmental responsibility. This paper presents a multi-objective optimization framework to determine an EC's optimal day-ahead operational strategy, minimizing both electricity costs and CO_2 emissions. The framework consists of an optimization model, developed as mixed-integer linear programming (MILP) in the General Algebraic Modeling System (GAMS) environment, and an algorithm in MATLAB, producing Pareto-optimal solutions. At a weighting factor of 0.05, the optimal solution reduces electricity costs by 7.99 € and CO_2 emissions by 1.03 kg.*

Keywords. CO_2 Emissions, Electricity Costs, Energy Communities, Multi-Objective Optimization

1 Introduction

Energy Communities (ECs) aim to promote local renewable energy generation and consumption, reduce greenhouse gas emissions, enhance energy security, lower energy costs, foster community engagement and social inclusion, support local economic development, and encourage energy efficiency.

ECs may be described as a way of organizing collective energy activities to provide benefits for all community members. There are two types of ECs: "Citizen Energy Communities", which are covered by the Directive on the Internal Electricity Market (EU) 2019/944, and "Renewable Energy Communities", which are included in the Directive on Renewable Energy (EU) 2018/2001. Although ECs participate in economic activities, their primary objective is to ensure ecological and social benefits for the community (Caramizaru and Uihlein, 2020).

Given the complexity of planning and operating ECs, where technical, economic, and environmental objectives often conflict, multi-objective optimization (MOO) methods have become essential tools for supporting decision-making and ensuring balanced development. In contrast to single-objective optimization, which targets the improvement of a single performance criterion, MOO simultaneously addresses multiple objectives. Rather than producing a single optimal outcome, MOO algorithms generate a set of trade-off solutions known as the Pareto-optimal set or Pareto front. Each solution within this set is non-dominated, implying that no other solution is superior across all considered objectives.

Although the Pareto front enables a comprehensive evaluation of alternative system configurations, the selection of a final solution necessitates the application of additional decision-making criteria or constraints, aligned with the specific priorities of the system or stakeholders. MOO methods are generally classified into decomposition-based approaches and direct solution approaches. In decomposition-based methods, the original multi-objective problem is transformed into a single-objective problem. Common techniques in this category include the Weighted Sum, Weighted Metric Sum, and the ϵ-constraint method. The main disadvantage of decomposition approaches is that the Pareto front set will be established after multiple iterations. In contrast, direct approaches aim to identify the Pareto-optimal set in a single optimization run by simultaneously evaluating all objective functions. Some of the MOO methods based on direct approaches are: Non-Dominated Sorting Genetic Algorithm II (NSGA-II), Multi-Objective Particle Swarm

Optimization (MOPSO), Strength Pareto Evolutionary Algorithm II (SPEA II), Multi-Objective Evolutionary Algorithm Based on Decomposition (MOEA/D), and Pareto Envelope-based Selection Algorithm II (PESA-II) (Salehi et al., 2022).

2 Related work

Authors in the literature have adopted different approaches to solving multi-objective optimization problems. In (Shaker et al., 2021), the authors proposed a new multi-objective algorithm, Multi-Objective Hunger Game Search Optimizer (MOHGS), to determine the optimal charging or discharging decisions for the energy storage units within the microgrid system. Four objective functions are proposed: minimizing the operating cost, minimizing the emissions, minimizing the loss of power supply probability, and maximizing the renewable factor (RF). Additionally, the authors modified the first objective function by adding the battery degradation cost. The objective is achieved through the Pareto function. Compared to the multi-objective versions of the latest techniques, including Marine Predators Algorithm (MOMPA), Slime Mould Algorithm (MOSMA), Multi-Verse Optimizer (MOMVO), Golden-Eagle Optimizer (MOGEO), Grasshopper Optimization Algorithm (MOGOA), Antlion Optimizer (MOALO), and Grey Wolf Optimizer (MOGWO), the authors presented the results demonstrating the superiority of the proposed MOHGS-based approach in providing the optimal charging/discharging cycles.

The authors in (Kong et al., 2023) presented a bi-level optimization model for enhancing the independence of distributed generation systems by integrating a shared energy storage system (ESS) and an internal energy market within an EC. At the upper level, a multi-objective optimization approach is adopted. The first objective function is to minimize the total cost associated with the ESS, including investment, replacement, and operating costs, while the second objective function is to maximize the self-sufficiency rate, i.e., the proportion of demand that the community meets. The objective function of the lower-level model is to maximize the community's social welfare, that is, to minimize the operating cost of the community. To solve two optimization problems, the authors applied a nested algorithm: the upper-level multi-objective problem is solved using the NSGA-II to obtain a Pareto front of optimal trade-off solutions, while the lower-level single-objective problem is solved using the Gurobi solver. Furthermore, to obtain an optimal solution from the Pareto solution set, the Utopian point method is applied. Although due to the trade-off relationship between objective functions, no Utopian point could be achieved. In (Wang et al., 2023), the authors proposed a multi-objective optimization problem applied to a combined power system, including photovoltaics, diesel generators, and fuel cells. The optimization aimed to minimize two objective functions: the loss of power supply probability (LPSP) and the total net present cost (TNPC). To solve the multi-objective problem, the authors improved the Sea Lion Optimization Algorithm (SLOA), which simulates the social behavior of sea lions during hunting. Furthermore, to improve SLOA to be used for MOO purposes, the authors utilized the Particle Swarm algorithm, the optimizer which looks for solutions to multi-objective problems by applying the Pareto dominance principle (using an external archive or secondary population to accumulate non-dominated answers and influence the direction of the search for the next descendant). The multi-objective optimization was conducted using the Multi-Objective Sea Lion Optimization Algorithm (MOSLO) and applying the Pareto ranking technique.

In (Mariuzzo et al., 2023), the authors presented an optimal sizing and management MILP-based methodology for ECs, which incorporates the flexibility of appliances in the form of demand side management (DSM) and demand response (DR). The A-AUGMENCON2 algorithm, as an improved version of the augmented ϵ-constraint method, is applied. The proposed objective functions are: minimizing total annualized costs, minimizing CO_2 emissions, maximizing social comfort, and minimizing energy exchange with the grid. The solution is obtained by applying the Pareto front. The authors in (Liu et al., 2023) proposed a distributed mixed-integer conic programming (MICP)-based energy management framework for actively interfaced participants, such as distributed energy resources (DERs), flexible loads, and microgrids, embedded within a modern distribution system. The energy management framework is formulated as an MOO problem, including the network operational objectives such as bus voltage deviations, feeder losses, and power factor of the distribution substation, while minimizing the utility grid's total operational cost. These objectives are aggregated into a single objective function using corresponding weighting coefficients, which were determined by the Analytic Hierarchy Process (AHP). To solve the optimization problem, the authors applied the Alternating Direction Method of Multipliers (ADMM).

The authors in (Anuradha et al., 2024) proposed a multi-objective stochastic optimization framework to size and place a community energy storage system (CESS) under three different energy pricing schemes (EPSs) for the CESS provider. The objective functions are to minimize the CESS provider's investment and operating costs, as well as the prosumers' operating costs. The uncertainties in PV generation and real/reactive energy consumption are considered. To solve this multi-objective problem, the authors used the ϵ-constraint method, which generates a Pareto front of non-dominated solutions. Furthermore, the authors in (De Leon et al., 2024) presented a multi-objective lin-

ear programming (MOLP) model to optimize the technical and business design of the novel local energy community DC microgrid scheme. The two objective functions are proposed: minimizing the total lifetime costs of the participating households and of the sponsoring local government unit (LGU). The Pareto optimal front is obtained through the ϵ-constrained algorithm to identify optimal solutions.

In (Di Somma et al., 2022), the authors proposed stochastic MOO model, based on mixed integer linear programming (MILP), of a real case study, which is treated as an integrated local energy community (ILEC) with two electrically interconnected multi-energy hubs involving technologies such as PV, solar thermal, combined heat and power systems, electric and geothermal heat pumps, absorption chillers, electric and thermal storage. The optimization model has two objective functions: the minimization of the net daily energy costs and carbon emissions. The optimization problem is solved by applying the Weighted Sum Method, while the optimal solution is established through a Pareto-optimal set of solutions. To account for the uncertainties associated with renewable energy sources (RES), the roulette wheel method (RWM) is applied to generate an initial set of solar irradiance scenarios. the fast forward selection algorithm is applied to retain the most representative scenarios, thereby reducing the computational burden in the subsequent optimization stage.

The literature review reveals that greater emphasis is placed on economic objectives than on environmental ones. To fill this research gap, this paper proposes an MOO framework to determine an EC's optimal day-ahead operational strategy, which simultaneously minimizes electricity costs and CO_2 emissions by applying the Weighted Sum Method. The optimal solution is identified through sets of Pareto-optimal solutions.

This paper is organized as follows. In Chapter 2, the proposed optimization model is presented. In Chapter 3, the input data and the applied optimization technique are described. In Chapter 4, the results are presented, and in Chapter 5, the conclusion is given.

3 The proposed multi-objective optimization framework

3.1 The proposed optimization model

This paper proposed a multi-objective optimization model, which consists of two objective functions: the minimization of electricity costs and the minimization of CO_2 emissions. The optimization problem is solved by applying the Weighted Sum Method, which aggregates two objective functions into one single objective function. The objective is achieved through the Pareto front, which balances electricity costs and CO_2 emissions.

The objective function is defined according to eq. 1.

$$opt = \omega \cdot E_{cost} + (1 - \omega) \cdot CO_{2,total} \tag{1}$$

where: opt represents the function to be minimized; ω represents a weighting factor indicating the importance of each of the two objective functions; E_{cost} represents total electricity costs; $CO_{2,total}$ represents the total amount of CO_2 emissions.

The upper limit of ω is 1, meaning that setting ω = 1 produces a solution that minimizes total expected electricity costs. The lower limit of ω is 0, meaning that setting ω = 0 produces a solution that minimizes the expected $CO_{2,total}$ emissions. Adjusting ω within the range of 0 to 1, the Pareto front can be identified, including possible trade-off solutions between economic and environmental objectives. This method is simple to implement and has been shown to identify all Pareto-optimal solutions in convex problems with only two objective functions (Di Somma et al., 2022).

The total electricity costs of all prosumers in EC are defined according to eq. 2.

$$E_{cost} = c \cdot \sum_{p=1}^{P} \sum_{t=1}^{T} \left(\frac{P_{k(p,t)}}{6} \cdot C_{kt(p,t)} - \frac{P_{p(p,t)}}{6} \cdot C_{pt(p,t)} + \frac{P_{lk(p,t)}}{6} \cdot C_{c(p,t)} - \frac{P_{lp(p,t)}}{6} \cdot C_{c(p,t)} \right) \tag{2}$$

where: c represents a scaling factor; $P_{k(p,t)}$ and $P_{p(p,t)}$ represent amounts of power imported and exported to the grid by prosumer p at a time step t, respectively; $P_{lk(p,t)}$ and $P_{lp(p,t)}$ represent amounts of power imported and exported to the local electricity market (LEM) by prosumer p at time step t, respectively; $C_{kt(p,t)}$ and $C_{pt(p,t)}$ represent the price of electricity imported and exported from/to the grid, respectively; $C_{c(p,t)}$ represents the price of electricity imported/exported to the LEM.

The total amount of CO_2 emissions caused by the import of electricity from the grid is defined according to eq. 3.

$$CO_{2,total} = \sum_{p} CO_2 \cdot E_{imp,uk(p)} \tag{3}$$

where: CO_2 represents the carbon intensity of imported electricity from the grid; $E_{imp,uk(p)}$ represents total electricity imported from the grid by prosumer p.

The state of charge (SOC) of the electric vehicle (EV) battery is defined according to equations 4 and 5. Eq. 4 applies to the first time step (t = 1), while eq. 5 is used for each subsequent time step.

$$SOC_{a(p,ev,t)} = SOC_{a(p,ev,start)} + \eta_c \cdot \frac{P_{cha(p,ev,t)}}{6} - \frac{1}{\eta_d} \cdot \frac{P_{dcha(p,ev,t)}}{6} \tag{4}$$

$$SOC_{a(p,ev,t)} = SOC_{a(p,ev,t-1)} + \eta_c \cdot \frac{P_{cha(p,ev,t)}}{6} - \frac{1}{\eta_d} \cdot \frac{P_{dcha(p,ev,t)}}{6} \quad (5)$$

where: $SOC_{a(p,ev,t)}$ represents the SOC of the EV battery at time step t; $SOC_{a(p,ev,start)}$ represents the SOC of the EV battery at the beginning of the observed scheduling horizon; η_c and η_d represent the charging and discharging efficiencies of EV battery; $P_{cha(p,ev,t)}$ and $P_{dcha(p,ev,t)}$ represent amounts of power required to charge and discharge the EV battery at time step t, respectively.

Constraints related to the import/export of electricity from/to the grid and LEM are defined according to equations from 6 to 9:

$$0 \leq P_{k(p.t)} \leq P_{kmax(p.t)} \cdot P_{kbin(p.t)} \quad (6)$$

$$0 \leq P_{p(p.t)} \leq P_{pmax(p.t)} \cdot P_{pbin(p.t)} \quad (7)$$

$$0 \leq P_{lk(p.t)} \leq P_{lkmax(p.t)} \cdot P_{lkbin(p.t)} \quad (8)$$

$$0 \leq P_{lp(p.t)} \leq P_{lpmax(p.t)} \cdot P_{lpbin(p.t)} \quad (9)$$

where: $P_{kmax}(p,t)$ and $P_{pmax(p,t)}$ represent the maximum amounts of power that prosumer p can import and export to the grid at time t, respectively; $P_{lkmax(p,t)}$ and $P_{lpmax(p,t)}$ represent the maximum amounts of power that prosumer p can import and export to the LEM at time step t, respectively; $P_{kbin(p,t)}$, $P_{pbin(p,t)}$, $P_{lkbin(p,t)}$ and $P_{lpbin(p,t)}$ represent binary decision variables determining the import or export of electricity from the prosumer p at time step t.

Equations 10 and 11 define constraints related to charging and discharging an EV battery, respectively.

$$0 \leq P_{cha(p.ev,t)} \leq P_{cha,max} \quad (10)$$

$$0 \leq P_{dcha(p,ev,t)} \leq P_{cha,max} \quad (11)$$

where: $P_{cha,max}$ represents the maximum amount of charging, i.e., discharging power of the EV battery.

The constraint related to the SOC of the EV battery is defined as:

$$SOC_{min} \leq SOC_{a(p,ev,t)} \leq SOC_{max} \quad (12)$$

where: SOC_{min} and SOC_{max} represent the minimum and maximum SOC of the EV.

The following constraint ensures the balance between electricity import and export to the LEM:

$$\sum_{p=1}^{P} P_{lk(p,t)} = \sum_{p=1}^{P} P_{lp(p,t)} \quad (13)$$

The constraint on the simultaneous import and export of electricity to/from the grid and LEM is defined by equations 14 and 15.

$$P_{kbin(p,t)} + P_{pbin(p,t)} \leq 1 \quad (14)$$

$$P_{lkbin(p,t)} + P_{lpbin(p,t)} \leq 1 \quad (15)$$

3.1.1 The proposed algorithm

To ensure that both objective functions are treated equally in the optimization model, a weighting factor (ω) and a constant (c), which presents a scaling factor, are applied. The scaling factor c is introduced into the model to enable the comparison of different units of measurement - electricity costs, expressed in [€], and CO_2 emissions, expressed in [kg]. By varying ω from 0 to 1 in steps of 0.05, the objective is to achieve a balance between minimizing electricity costs and minimizing CO_2 emissions. At ω = 1, the primary focus is on reducing total electricity costs, while at ω = 0, the optimization focuses on reducing CO_2 emissions. To automate the process of multi-objective optimization, an algorithm was developed in MATLAB. The algorithm begins by defining the upper and lower limits of ω. Furthermore, MATLAB passes the necessary input data to the optimization model developed in GAMS, where the optimization problem is solved for each value of ω in the range from 0 to 1. After each iteration, the results are returned to MATLAB and stored, and it is checked whether the upper limit of ω has been reached. If the upper limit has not been reached, the process is repeated. The objective of this algorithm is to find the optimal value of ω. After the upper limit of ω is reached, the optimization results are processed and graphically displayed. Based on the results, the optimal solution is selected, meaning the value of ω that provides the best balance between minimizing electricity costs and minimizing CO_2 emissions. The process of the above algorithm is shown in Fig. 1.

4 Case study

The EC analyzed in the simulation consists of six prosumers. Prosumer P2 represents a business, while prosumers P1, P3, P4, P5 and P6 represent households. The electricity consumption and generation profiles were measured at the real prosumer households, while the business measurements were conducted at a faculty building in Osijek, Croatia.

Fig. 2 shows the electricity consumption profile of each prosumer in the community over the observed day-ahead. As a business, P2 has significantly higher electricity consumption compared to households, which is expected given the greater electricity demands of a business. Household consumption (P1, P3, P4, P5, P6) exhibits a lower and more evenly distributed load.

Fig. 3 shows electricity generation from photovoltaic (PV) power plants within the EC. All prosumers contribute to electricity generation, with P2, as a business, achieving the highest generation output. Household generation is also significant, but aligns with their lower electricity demands. Furthermore, all participants reach peak generation during similar time intervals.

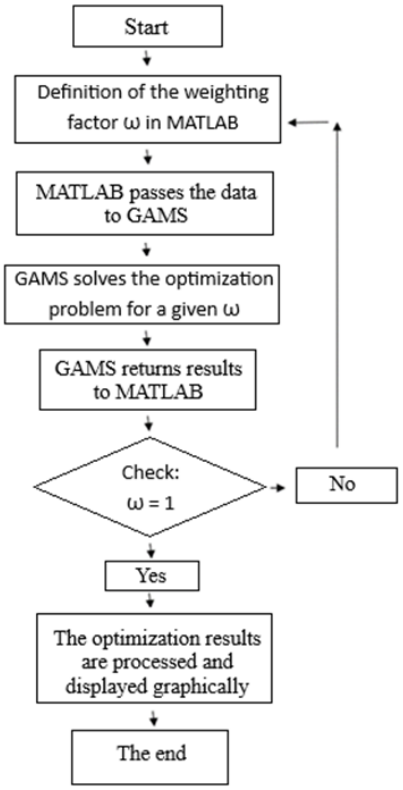


Figure 1. Graphical representation of the algorithm for automating the multi-objective optimization process in MATLAB

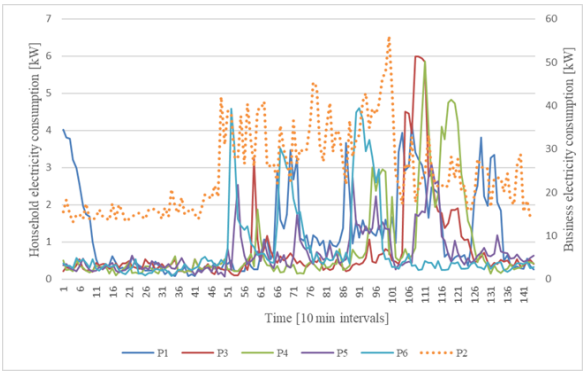


Figure 2. Electricity consumption of prosumers

In this and the following cases, time step t represents an interval of 10 minutes. The power for every time step t is divided by 6, since an hour contains 60 minutes. This approach allows the model to accurately monitor electricity consumption and generation in 10-minute intervals, while simultaneously displaying aggregated values on an hourly basis, which contributes to better analysis and management of electricity flows within the system.

Electricity prices are determined according to the "White Tariff Model", which includes two rates: a lower and a higher tariff. The pricing data was obtained from the official HEP website ("HEP - Tarifne

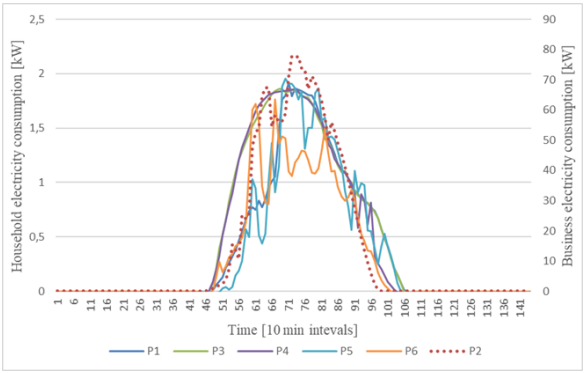


Figure 3. Electricity generation of prosumers

stavke (cijene)", 2024). For households, the prices are 0.083 €/kWh for the lower tariff and 0.159 €/kWh for the higher tariff. For businesses, prices are 0.12 €/kWh and 0.201 €/kWh, respectively. The price of electricity exported to the grid by households and businesses is determined according to the market model "Customer with Own Production." According to this model, households receive 0.05 €/kWh, while businesses receive 0.0374 €/kWh for the exported electricity. The electricity price on the LEM is 0.025 €/kWh, representing the average distribution grid usage fee for households. Other input parameters are given in Table 1, based on real data.

Table 1. Input parameters of the optimization model

Input parameter	Amount
P_{kmax}, P_{lkmax}	13.8/120 kW
P_{pmax}, P_{lpmax}	5/120 kW
η_c, η_d	0.985
P_{chamax}	7.4 kW
CO_2	0.246 kg/kWh
Electricity consumption of EV	0.175 kWh/km

The driving time of the EV and the distance it covers within that time are also considered. On average, EV owners travel between 40 and 90 km per day. In the simulation, each prosumer with an EV is assumed to travel within these limits. The travel time varies for each prosumer, ranging from 30 to 90 minutes (Mišljenović et al., 2024).

5 Results

The impact of ω on electricity costs and CO_2 emissions is shown in Figs. 4 and 5. The costs presented in the figures and tables represent actual electricity costs multiplied by the scaling factor c of 2.1, which was determined before the optimization process, ensuring equal order of magnitude of the two objective functions. According to Fig. 4, electricity costs are highest at $\omega = 0$, amounting to 130.67 €. When the scaling fac-

tor c is considered, the actual electricity cost is 62.22 €. This occurs because, at $\omega = 0$, the objective is to minimize CO_2 emissions. Fig. 5 shows an opposite trend compared to Fig. 4. At $\omega = 0$, CO_2 emissions are at their lowest value of 104.55 kg. However, as ω increases, CO_2 emissions also increase. The emissions stabilize at $\omega \geq 0.4$, reaching a value of 105.96 kg, with a maximum recorded value of 106.04 kg.

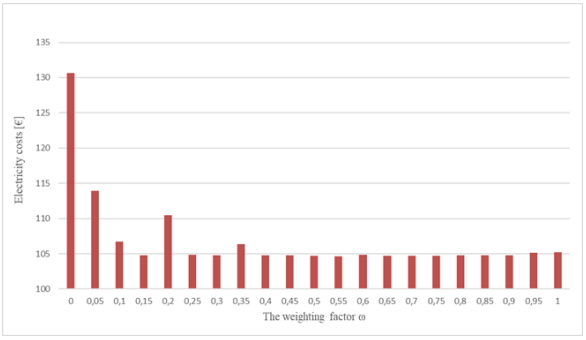


Figure 4. Total electricity costs considering ω

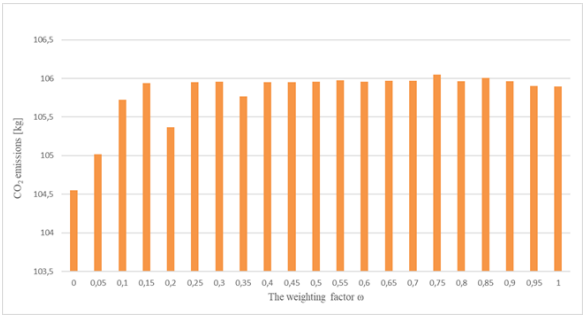


Figure 5. CO_2 emissions considering ω

Fig. 6 shows a comparative representation of scaled electricity costs and CO_2 emissions concerning different values of ω. As ω increases, total electricity costs decrease and stabilize at higher values of ω, while CO_2 emissions increase. This trade-off between economic and environmental objectives is crucial for system decision-making.

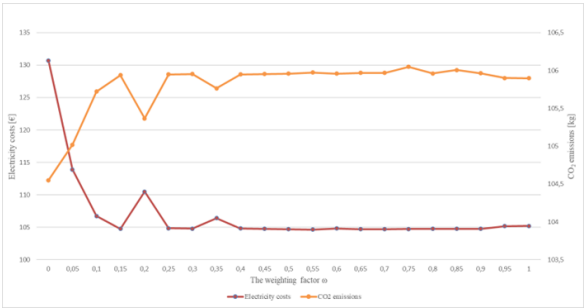


Figure 6. Combined representation of electricity costs and total CO_2 emissions

Fig. 7 shows the Pareto front, which shows that electricity costs and CO_2 emissions cannot be minimized simultaneously. The Pareto front helps identify the most efficient, i.e., optimal solution, where neither total electricity costs nor total CO_2 emissions are significantly aggravated. In this simulation, the Pareto front provides insight into how much the EC must compromise on electricity costs to reduce CO_2 emissions, and vice versa.

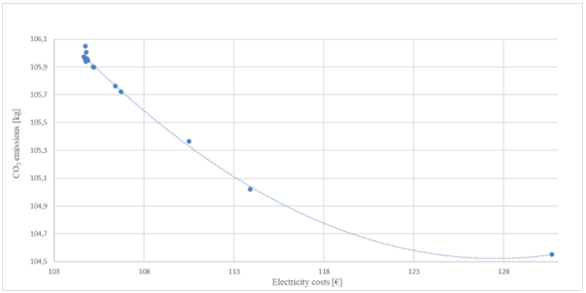


Figure 7. Pareto front

The results presented in the figures correspond to the optimized solution obtained for $\omega = 0.05$.

Fig. 8 shows the amounts of imported and exported electricity throughout the day. These values represent electricity exchange with the grid, as well as LEM. Prosumer P2 imports significantly more electricity than other prosumers, which is expected, as businesses generally have higher electricity consumption than households, making P2 the key electricity importer within the EC. On the other hand, prosumers P1, P3, P4, P5, and P6 (households) require considerably less imported electricity than P2. Additionally, these households contribute to electricity exports within the LEM, reducing the EC's reliance on the grid.

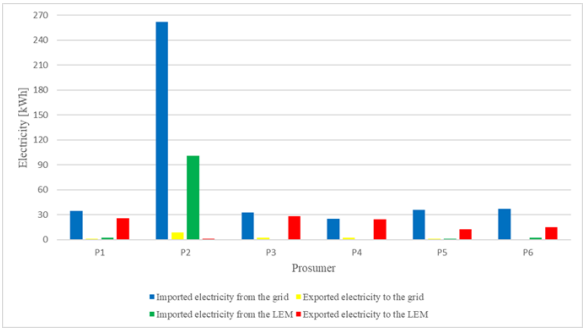


Figure 8. Exchange of electricity of prosumer with the grid and LEM

Fig. 9 shows the SOC of EV batteries throughout the day. The battery state varies depending on the time of day, indicating whether the batteries are charging or discharging. Three intervals can be identified in which the SOC changes. In the first interval, there is an increase in the SOC of the EV batteries when they are connected to home chargers in households. This suggests that households charge their EVs during the night and morning hours when electricity prices are lower.

In the second time interval of the day, the EV batteries are charged at the company during working hours. The EV batteries are charged with the surplus electricity, which is justified by the price of exporting electricity to the grid. EVs require a significant amount of electricity to charge, especially if the company has multiple EVs being charged simultaneously. In the final third of the day, there is a decrease in the SOC, suggesting that EVs in households are being used or the batteries are discharging to cover consumption. The change in SOC during the first and third intervals indicates that households are using electricity from their batteries to meet their daily electricity needs. Instead of charging EV batteries in the evening when electricity prices are higher, prosumers discharge batteries. In contrast, at night and in the morning, when electricity prices are lower, prosumers charge EV batteries. This shows how electricity is used and stored within the EC and provides insight into the efficiency of EV battery management throughout the day. Optimal management allows a balance between consumption, storage, and electricity generation, reducing the need for electricity import from the grid and thereby increasing the electricity independence of the EC.

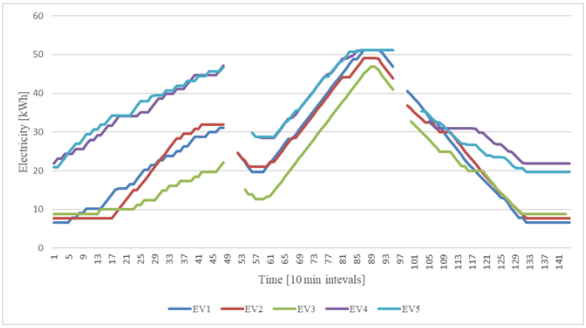


Figure 9. SOC of EV batteries

Correlations between ω and total electricity costs, as well as total CO_2 emissions, are shown in Table 2. The optimal solution is achieved with a weight factor ω of 0.05. At this value of ω, the total scaled electricity costs amount to 113.89 €, or actual electricity costs of 54.23 €, with the scaling factor c of 2.1, while the total CO_2 emissions are 105.01 kg. Through multi-objective optimization of the EC's operations, total electricity costs were reduced by 7.99 €, compared to the initial electricity cost of 62.22 €, and CO_2 emissions were reduced by 1.03 kg, compared to the highest amount of 106.04 kg.

The main objective is to reduce CO_2 emissions while keeping electricity costs within acceptable limits. Although electricity costs are significantly lower at higher values of ω, it is important to consider the objectives of EC. Since ECs in Croatia are non-profit organizations, their primary focus is to reduce CO_2 emissions, i.e., reduce any negative environmental impacts. Therefore, more importance is given to reducing CO_2 emissions than to minimizing electricity costs.

Table 2. Electricity costs and total CO_2 emissions in correlation with ω

ω	Electricity costs [€]	Scaled electricity costs [€]	Total CO_2 emissions [kg]
0.00	62.22	130.67	104.55
0.05	54.23	113.89	105.01
0.10	50.82	106.72	105.72
0.15	49.89	104.76	105.93
0.20	52.61	110.49	105.36
0.25	49.93	104.86	105.94
0.30	49.89	104.78	105.95
0.35	50.66	106.39	105.76
0.40	49.91	104.82	105.94
0.45	49.88	104.76	105.95
0.50	49.86	104.70	105.95
0.55	49.83	104.64	105.97
0.60	49.92	104.83	105.95
0.65	49.86	104.70	105.96
0.70	49.85	104.69	105.96
0.75	49.87	104.73	106.04
0.80	49.88	104.76	105.96
0.85	49.89	104.77	106.00
0.90	49.88	104.76	105.96
0.95	50.08	105.16	105.90
1.00	50.10	105.21	105.89

6 Conclusion

This paper proposes a multi-objective optimization framework to determine an EC's optimal day-ahead operational strategy, simultaneously minimizing electricity costs and CO_2 emissions. The proposed framework consists of an optimization model, developed as an MILP in GAMS, and an algorithm, developed in MATLAB. The optimal solution is determined through sets of Pareto-optimal solutions.

The results show how effective optimization of EC's operations can impact greenhouse gas emissions and the overall costs of electricity generation, consumption, and distribution. Through multi-objective optimization, a balance between ecological and economic objectives is achieved with a weight factor ω of 0.05, where the EC's electricity costs were reduced from 62.22 € to 54.23 €, representing a decrease of 7.99 €, while CO_2 emissions were reduced from 106.04 kg to 105.01 kg, representing a decrease of 1.03 kg. This solution emphasizes that a higher priority was placed on reducing CO_2 emissions, while electricity costs are within acceptable limits for the community members. Although there is potential for further cost reduction, this would result in higher CO_2 emissions, which is not in line with the objectives of ECs, which, as non-profit organizations, aim to reduce their negative environmental impact.

Future research may explore the application of multi-objective optimization to ECs providing flexibility services, simultaneously minimizing electricity costs and CO_2 emissions, while considering stochastic behavior.

7 Acknowledgments

This work was supported by the Croatian Science Foundation under the project "Prosumer-rich distribution power network" (project number: UIP-2020-02-5796). The work of doctoral student Sanja Kelemen has been supported by the "Young Researchers' Career Development Project - training of Doctoral Students" (DOK-NPOO-2023-10-4888) of the Croatian Science Foundation.

References

Anuradha, K. B., Iria, J., & Mediwaththe, C. P. (2024). Multi-objective planning of community energy storage systems under uncertainty. *Electric Power Systems Research*, *230*(December 2023), 110286. https://doi.org/10.1016/j.epsr.2024.110286

Caramizaru, A., & Uihlein, A. (2020). *Energy communities: an overview of energy and social innovation Social aspects of the energy transition*. https://doi.org/10.2760/180576

De Leon, J. A., Tan, R. R., & Billones, R. K. (2024). Multi-Objective Linear Programming for Optimizing a Local Energy Community DC Microgrid System and Business Model. *IEEE Access*, *12*(September), 171513–171526. https://doi.org/10.1109/ACCESS.2024.3493774

Di Somma, M., Buonanno, A., Caliano, M., Graditi, G., Piazza, G., Bracco, S., & Delfino, F. (2022). Stochastic Operation Optimization of the Smart Savona Campus as an Integrated Local Energy Community Considering Energy Costs and Carbon Emissions. *Energies*, *15*(22), 1–27. https://doi.org/10.3390/en15228418

Hep - tarifne stavke (cijene). (2024). https://www.hep.hr/elektra/kucanstvo/tarifne-stavke-cijene/1547 (Accessed: 2024.)

Kong, X., Mu, H., Wang, H., & Li, N. (2023). Independence enhancement of distributed generation systems by integrating shared energy storage system and energy community with internal market. *International Journal of Electrical Power and Energy Systems*, *153*(July), 109361. https://doi.org/10.1016/j.ijepes.2023.109361

Liu, G., Ollis, T. B., Ferrari, M. F., Sundararajan, A., & Chen, Y. (2023). Distributed Energy Management for Networked Microgrids Embedded Modern Distribution System Using ADMM Algorithm. *IEEE Access*, *11*, 102589–102604. https://doi.org/10.1109/ACCESS.2023.3316513

Mariuzzo, I., Fioriti, D., Guerrazzi, E., Thomopulos, D., & Raugi, M. (2023). Multi-objective planning method for renewable energy communities with economic, environmental and social goals. *International Journal of Electrical Power and Energy Systems*, *153*(March), 109331. https://doi.org/10.1016/j.ijepes.2023.109331

Mišljenović, N., Knežević, G., Žnidarec, M., & Topić, D. (2024). Optimal State of Charge Control of EV Batteries within Energy Community Considering Cost Minimization and Environmental Impact. *International Conference on the European Energy Market, EEM*, 1–6. https://doi.org/10.1109/EEM60825.2024.10608901

Salehi, N., Martinez-Garcia, H., Velasco-Quesada, G., & Guerrero, J. M. (2022). A Comprehensive Review of Control Strategies and Optimization Methods for Individual and Community Microgrids. *IEEE Access*, *10*, 15935–15955. https://doi.org/10.1109/ACCESS.2022.3142810

Shaker, Y. O., Yousri, D., Osama, A., Al Gindy, A., Tag Eldin, E., & Allam, D. (2021). Optimal Charging/Discharging Decision of Energy Storage Community in Grid-Connected Microgrid Using Multi-Objective Hunger Game Search Optimizer. *IEEE Access*, *9*, 120774–120794. https://doi.org/10.1109/ACCESS.2021.3101839

Wang, H., Mansor, N. N. B., Mokhlis, H. B., & Mashhadi, A. (2023). Optimal size selection of combined diesel generator/fuel cell/photovoltaic system components using a multi-objective strategy and sea lion optimization algorithm. *IET Renewable Power Generation*, (September 2022), 1–16. https://doi.org/10.1049/rpg2.12795

Reactive Scheduling under Due-Date and Intermediate Storage Constraints: A Conceptual Multi-Objective Approach

Krisztián Attila Bakon, Tibor Holczinger
Faculty of Information Technology, University Center for Circular Economy Nagykanizsa, University of Pannonia
Department of Applied Informatics
Zrinyi u. 18., Nagykanizsa, 8800, Hungary
{bakon.krisztian, holczinger.tibor}@pen.uni-pannon.hu

Abstract. *This paper presents a conceptual reactive scheduling framework for Flexible Job Shops with dynamic order arrivals, due-date targets, and intermediate storage constraints (NIS/UIS). Unlike existing models that treat these aspects separately, our approach integrates them into a unified multi-objective optimization method. We introduce a storage-aware branch-and-bound algorithm with four policy-based flexibility levels and prove a theorem enabling intermediate storage minimization through local slack redistribution. This framework offers a novel way to balance due-date penalties and storage costs, setting the foundation for future real-time applications in adaptive manufacturing environments.*

Keywords. Reactive Scheduling, Flexible Job Shop Problem, Due-Date Optimization, Intermediate Storage Policies, Branch-and-Bound Algorithm, Earliness/Tardiness Minimization, Storage-Gap Adjustment, Policy-Driven Rescheduling

1 Introduction

Modern production systems face challenges from dynamic order arrivals, stringent due dates, and complex resource constraints. This paper presents a comparative framework for reactive (Bakon and Holczinger, 2024) and deterministic scheduling (Bakon and Holczinger, 2025) in an extended Flexible Job Shop Problem (FJSP), integrating real-time order processing, due-date optimization, and intermediate storage policies (NIS/UIS).

Although reactive scheduling and due-date optimization with intermediate storage constraints have each been addressed individually in prior research, they are rarely combined within a single, unified framework—particularly under dynamic order arrivals and rescheduling scenarios. Existing methods typically either overlook the impact of storage behavior on reactive scheduling or focus solely on makespan or tardiness, without incorporating storage-related penalties.

The proposed framework targets a dual objective: minimizing total earliness/tardiness (E/T) penalties and reducing intermediate storage time (IST). Success is defined as achieving a schedule that balances these competing goals under different levels of rescheduling flexibility and realistic storage constraints.

This paper addresses this gap by proposing a conceptual scheduling framework that integrates reactive rescheduling with due-date and storage-aware objectives in an extended FJSP environment. The core contributions are: (1) a policy-aware branch-and-bound (B&B) algorithm supporting four levels of rescheduling flexibility, and (2) the Storage-Gap Adjustment Theorem, which formally enables IST minimization through local slack redistribution. Together, these contributions bridge previously disjoint research threads and lay the foundation for scalable, multi-objective scheduling in adaptive manufacturing systems.

2 Literature Review

Scheduling research spans reactive rescheduling, due-date control, and intermediate storage modeling, yet these dimensions are rarely treated together in a unified framework. However, most contributions treat these dimensions in isolation. This section reviews five thematic strands and clarifies how they inform our unified, policy-driven framework.

Reactive Scheduling Models

Reactive scheduling literature emphasizes balancing responsiveness and stability. Early frameworks distinguish dispatching, predictive–reactive, and proactive-reactive strategies (Bahroun et al., 2024; Vieira et al., 2003). Event-driven policies outperform periodic rescheduling in turbulent environments (Ouelhadj and Petrovic, 2009). More recent contributions incorporate robustness, decision trees (Portoleau et al., 2020), rolling-horizon control laws (Kopanos and Pistikopou-

los, 2014), and adaptive intervals (Wang et al., 2017). Despite this evolution, cognitive constraints and integration with domain-specific policies (e.g., storage constraints) remain under-addressed (Herroelen and Leus, 2005)—a gap this work targets explicitly.

Due-Date Assignment and Flexibility

Due-date strategies range from static job-level rules (e.g., TWK, DPPW) to dynamic updates based on real-time shop load (Cheng and Gupta, 1989; Cheng and Jiang, 1998; T et al., 2017). Extensions incorporate weighted tardiness (Koulamas, 2011), job-shop adaptation (Ojstersek et al., 2020), and project-based economic criteria (Vanhoucke, 2002). While several models address tightness vs. service trade-offs, interactions between due-date pressure and storage-aware rescheduling are rarely modeled. Our framework explicitly incorporates this interaction.

Storage Constraints in Scheduling

Storage modeling has matured in batch and job-shop contexts, formalizing UIS, NIS, and ZW policies (Ha et al., 2000; Kim et al., 1996; Kopanos and Puigjaner, 2009; Romero et al., 2004). S-graph methods and MILP formulations support comparative makespan studies. However, most models are static and ignore rescheduling flexibility. Our policy-aware approach embeds UIS/NIS behavior directly into dynamic rescheduling.

Multi-Objective Scheduling Trends

Recent literature emphasizes Pareto-optimal trade-offs among cost, stability, and service (Holguin Jimenez et al., 2024; Wang et al., 2017). Metaheuristics (e.g., MOSM (Mihály and Kulcsár, 2023), PRIMP (Mansini et al., 2023)) enhance solution diversity, though few frameworks jointly optimize E/T and storage. We address this by integrating both into a reactive branch-and-bound strategy.

Computational Considerations

Scheduling complexity is well established: even simplified job-shop problems are NP-hard with no constant-factor approximations (Mastrolilli and Svensson, 2011). Instance features like dispersion influence heuristic selection (Ruiz-Vanoye et al., 2011). Robust project scheduling models address both solution quality and deviation limits (Herroelen and Leus, 2004, 2005), motivating scalable yet adaptable algorithms. Our model responds with policy-aware pruning and tractable bounding.

In summary, while each domain shows strong development, their integration remains limited. Our contribution lies in combining reactive rescheduling, due-date compliance, and intermediate storage optimization into a single multi-objective scheduling framework for dynamic job-shop environments.

3 Problem Definition

The production environment consists of a finite set of equipment units J, a finite set of products P, and a finite set of tasks I. Each product $p \in P$ is defined by a set of tasks $I_p \subseteq I$, with precedence constraints captured by $I_i^- \subseteq I$, the set of immediate prerequisites for task $i \in I$. Each task $i \in I$ may be performed by a subset of equipment $J_i \subseteq J$, with processing time $pt_{i,j}$ depending on the assigned unit $j \in J_i$.

The production system operates under a dynamic arrival process, where orders $O = \{o_1, o_2, \ldots, o_n\}$ aare revealed sequentially over time. Each order $o_k = (I_k, t_k)$ consists of a product-specific task set $I_k \subseteq I$ and an associated arrival time t_k. All arrival times are assumed distinct, with $t_1 = 0$ without loss of generality. This model does not assume any probabilistic distribution over arrivals; instead, it responds deterministically to each new order as it occurs. Thus, while future arrivals are unknown, the system handles them online in a non-anticipative fashion. Let $I_k^* = \bigcup_{k' \leq k} I_{k'}$ denote the cumulative set of tasks from the first k orders.

Each product $p \in P$ is associated with a due date d_p, either explicitly specified or computed based on the task processing time lower bound:

$$d_p = \left\lfloor f \sum_{i \in I_p} \min_{j \in J_i} pt_{i,j} \right\rfloor .$$

At time t_{k+1}, when a new order o_{k+1} arrives, the current schedule S_k is partitioned into:

- $\underline{S}_k$: the set of assignments that have started before t_{k+1}, and are preserved as fixed assignments in the revised schedule;
- $\overline{S}_k$: the set of assignments that have not yet started and may need to be rescheduled.

Let $\underline{I}_k^*$ and $\overline{I}_k^*$ be the corresponding task sets, such that $I_k^* = \underline{I}_k^* \cup \overline{I}_k^*$.

The objective is to construct a new feasible schedule S_{k+1} satisfying the following:

- All tasks from $I_{k+1} \cup \overline{I}_k^*$ are rescheduled, maintaining precedence and resource constraints;
- Assignments in $\underline{S}_k$ remain unchanged, i.e., $\underline{S}_k \subseteq S_{k+1}$;
- No task begins before its order's arrival time: $t^s \geq t_{k+1}$ for all new assignments;
- The total weighted penalty for E/T across all products up to o_{k+1} is minimized:

$$\min \sum_{p \in P} (w_p^E E_p + w_p^T T_p),$$

where $E_p = \max(0, d_p - C_p)$, $T_p = \max(0, C_p - d_p)$, and C_p is the completion time of the last task $\hat{i}^p$ of product p.

The scheduling environment adheres to one of two intermediate storage policies, each imposing distinct temporal constraints on task execution and inter-operation coordination:

- **NIS**: $\forall i \in I, \quad t^s_{i^+} = t^f_i$ (if equipment available)
- **UIS**: $\forall i \in I, \quad t^s_{i^+} \geq t^f_i$

This problem formulation integrates dynamic order arrivals, task-equipment flexibility, precedence constraints, E/T penalties, and realistic storage behavior into a unified scheduling model. The aim is to develop a robust, scalable, and efficient solution method capable of addressing both the reactivity and optimization challenges in modern production systems.

4 Methodology: Reactive Scheduling with Due-Date and Storage-Aware Optimization

4.1 Reactive Scheduling Framework

Upon receipt of the first order $o_1 = (I_1, t_1 = 0)$, an initial schedule S_1 is constructed. This schedule minimizes due-date deviation (E/T) and respects either the NIS or UIS storage policy. It is derived using the traditional S-graph approach—either equipment-based or task-based branching—extended to compute start and finish times that align as closely as possible with each job's due date d_p.

When a new order $o_{k+1} = (I_{k+1}, t_{k+1})$ arrives, the current schedule S_k is partitioned into:

$$\underline{S}_k = \{(i, j, t^s, t^f) \in S_k \mid t^s < t_{k+1}\}, \quad \overline{S}_k = S_k \backslash \underline{S}_k,$$

with corresponding task sets $\underline{I}^*_k$ (started) and $\overline{I}^*_k$ (not yet started).

Any revised schedule S_{k+1} must ensure that the updated schedule S_{k+1} adheres to the following criteria:

1. Retain all assignments in $\underline{S}_k$.
2. Ensure that no task of I_{k+1} begins before time t_{k+1}.
3. Preserve precedence, storage, and machine-task feasibility.
4. Continue minimizing total $\sum(w^E_p E_p + w^T_p T_p)$, evaluating completion times C_p under either NIS or UIS storage constraints.

4.2 Reactive Policies under Due-Date Minimization

Depending on the desired trade-off between computational efficiency and schedule quality, the rescheduling policy for $\overline{I}^*_k$ can be selected from the following levels of flexibility:

- **Policy 1:** $\overline{I}^*_k$ remains unaltered. New tasks I_{k+1} are appended. While simple and fast, delays incurred by new orders may increase tardiness penalties.
- **Policy 2.1:** Sequencing of $\overline{I}^*_k$ is fixed, and start times are unchanged. New tasks can only be inserted into idle time windows. This strategy safeguards due-date alignment but may limit flexibility if idle time is scarce.
- **Policy 2.2:** Sequence is fixed, but start times of $\overline{I}^*_k$ may shift, facilitating insertions without reassignments. This improves due-date optimization at the cost of modest computation.
- **Policy 3:** Full rescheduling of $\overline{I}^*_k \cup I_{k+1}$ is permitted. This offers the greatest flexibility for due-date equality and storage time reduction but is the most computationally intensive.

4.3 S-Graph Augmentation for Reactive Due-Date Scheduling

The S-graph for rescheduling must encode both the due-date objective and the constraints imposed by reactive policies:

- An artificial root node Z is introduced to connect to each task via arcs weighted by their respective order arrival times, thereby enforcing earliest start time constraints.
- Zero-wait arcs fix the start times of tasks in $\underline{I}^*_k$.
- According to the chosen policy:
  - **Policy 1 & 2.1:** Add zero-wait arcs to all tasks in $\overline{I}^*_k$, freezing their start times.
  - **Policy 2.2:** Only add sequence-preserving schedule-arcs among $\overline{I}^*_k$, allowing temporal flexibility.
  - **Policy 3:** Do not add any arcs among $\overline{I}^*_k$; full rescheduling is allowed.

The graph structure supports integration of task precedence, equipment assignment, and temporal restrictions. B&B search traverses assignments and sequencing steps to build schedule S_{k+1}:

- Each node corresponds to a partial S-graph including current assignments and arcs.
- The algorithm evaluates completion times C_p for all affected products and computes new E_p, T_p, along with storage delays if applicable.

- Lower bounds on the total weighted due-date penalty are computed to prune dominated branches.
- Feasible extensions—such as task insertions or reassignment—are performed in accordance with the constraints imposed by the selected reactive policy.

4.4 Multi-Objective Optimization: Earliness/Tardiness and Intermediate Storage

The reactive scheduling problem involves determining an updated schedule S_{k+1} upon the arrival of a new order o_{k+1} at time t_{k+1}, such that E/T and IST across all jobs are minimized. The multi-objective function is:

$$\min\left(\alpha \cdot IS_{\text{total}} + \beta \cdot (E_{\text{total}} + T_{\text{total}})\right),$$

where $\alpha, \beta \geq 0$ and $\alpha + \beta = 1$. The E/T penalties and IST are computed as:

$$E_{\text{total}} = \sum_{p \in P} \max(0, d_p - t^f_{i^p}),$$
$$T_{\text{total}} = \sum_{p \in P} \max(0, t^f_{i^p} - d_p),$$
$$IS_{\text{total}} = \sum_{i \in I} is_i.$$

4.4.1 Storage-Aware Schedule Adjustments

At each branching step within the S-graph framework, tasks are assigned with respect to both machine availability and storage restrictions. The system checks:

- Recipe sequence constraints for immediate downstream operations
- Machine sequence feasibility without conflicts
- Potential for intermediate storage reduction while maintaining feasibility

If a task i incurs positive IST $is_i > 0$, the scheduler investigates if $is_i = 0$ can yield a valid schedule. If not, a search procedure identifies the minimal feasible is_i, applied recursively to neighboring tasks to balance cumulative IST.

4.5 Algorithmic Framework: Reactive B&B with Storage Optimization

This subsection introduces a storage-aware branch-and-bound (B&B) algorithm tailored to reactive scheduling with due-date and intermediate storage constraints. The algorithm dynamically adjusts its branching strategy based on the chosen rescheduling policy $\pi \in \{1, 2.1, 2.2, 3\}$, balancing computational complexity and scheduling flexibility.

Algorithm 1: Reactive B&B with Due-Date and Storage-Aware Objectives

Input: Current schedule S_k; new order o_{k+1}; policy $\pi \in \{1, 2.1, 2.2, 3\}$
Output: Revised schedule S_{k+1}

Initialize queue Q with root node based on S_k and policy π
Best objective $Z^\star \leftarrow \infty$, $S_{k+1} \leftarrow \emptyset$
while $Q \neq \emptyset$ **do**
  node $u \leftarrow$ pop_best(Q)
  if *LB*$(u) \geq Z^\star$ **then**
    continue
  if u *is complete* **then**
    $Z^\star \leftarrow Z(u)$, $S_{k+1} \leftarrow$ schedule(u)
  else
    Choose unscheduled task i according to branching policy π
    foreach $j \in J_i$ **do**
      $(t^s_i, t^f_i) \leftarrow$ ComputeTentativeInterval(i, j, π, u)
      if (t^s_i, t^f_i) *feasible* **then**
        $is_i \leftarrow$ EstimateStorageTime(i, π)
        if *equipment conflict occurs* **then**
          StorageGapAdjust(i, j)
        Create child node v with updated schedule and costs
        Compute partial objective: $Z(v) \leftarrow \alpha \cdot$ IST$(v) + \beta \cdot$ DueDatePenalty(v)
        Compute lower bound: LB(v)
        if *LB*$(v) < Z^\star$ **then**
          push node v into Q
return S_{k+1}

The following pseudocode (Algorithm 1) provides a structured overview of the algorithm's logic, illustrating the interactions between branching decisions, storage policies, and multi-objective optimization.

The algorithm integrates three core concerns: (i) reactivity to incoming orders by updating the schedule in real time, (ii) due-date adherence via E/T penalties in the objective, and (iii) intermediate storage control through IST estimation and slack redistribution. The function `StorageGapAdjust` ensures local feasibility and cost reduction under UIS policies, while the policy-driven branching logic ensures alignment with practical flexibility constraints.

Complexity and Convergence. The theoretical worst-case complexity of the B&B algorithm grows factorially with the number of reschedulable tasks ($|\overline{I}_k^*|$). Under Policy 1, complexity simplifies to equipment assignments only, yielding $\mathcal{O}((m)^{|\overline{I}_k^*|})$. Policy 3, permitting complete resequencing and reassignment, expands complexity significantly to $\mathcal{O}(|\overline{I}_k^*|!\, m^{|\overline{I}_k^*|})$. However, in practice, empirical convergence is commonly observed within acceptable computational timeframes for realistic scenarios (10–50 tasks, 5–10 machines), aided by tight bounding, heuristic guidance, and effective pruning.

4.5.1 Node Expansion and Temporal Assignment

Each node in the search tree extends the current partial schedule by tentatively assigning a start time t_i^s to an unscheduled task $i \in \overline{I}_k^* \cup I_{k+1}$, respecting precedence ($t_i^s \geq \max(t_{i'}^f)$ for all $i' \in I_i^-$), equipment availability (unit $j \in J_i$ must be idle), order arrival constraints ($t_i^s \geq t_{k+1}$ for new orders), and the immutability of previously fixed assignments in $\underline{S}_k$. The finish time is then set as $t_i^f = t_i^s + pt_{i,j}$. IST is estimated as $is_i = t_{i^+}^s - t_i^f$, with the algorithm attempting to enforce $is_i = 0$ under NIS where possible, and minimizing is_i under UIS via trade-offs between early starts and storage overhead. If conflicts arise on equipment j, the algorithm invokes a storage-gap adjustment: it identifies temporal neighbors i^- and i^+, redistributes slack ($\delta > 0$) from their respective storage times, and verifies that precedence remains valid ($t_{i^+}^s \geq \max(t_{i'}^f)$). If adjustment fails, t_i^s is incremented and reassessed. After placement, the algorithm evaluates the partial objective function,

$$Z = \alpha \sum_{i \in I_{\text{sched}}} is_i + \beta \sum_{p \in P_{\text{sched}}} (E_p + T_p),$$

where I_{sched} and P_{sched} are the sets of scheduled tasks and completed products, respectively, and $E_p = \max(0, d_p - t_{i_p}^f)$, $T_p = \max(0, t_{i_p}^f - d_p)$ are early/tardy penalties. Finally, nodes are pruned if their lower bound $Z_{\text{lower bound}}$, estimated using heuristics and policy-specific constraints, exceeds the current best objective value.

4.5.2 Storage-Gap Adjustment Proposition

Proposition 1. *For any feasible UIS schedule, local slack redistribution can strictly reduce the total intermediate storage time IS_{total} without violating precedence or resource constraints.*

Proof. Let $\Gamma = \{i_1, \ldots, i_q\}$ be the tasks on a unit j ordered by start time. Define $\Delta_k = t_{i_{k+1}}^s - t_{i_k}^f \geq 0$. If $\sum_k \Delta_k = 0$ the schedule is no-wait and thus minimal. Otherwise select $k^* = \arg\max_k \Delta_k$ and shift tasks $i_{k^*+1}, \ldots, i_q$ left by $\varepsilon = \min\{\Delta_{k^*}, \min_{r>k^*}(\Delta_r + pt_{i_r})\}$. All predecessor finish times remain $\leq t_{i_{k^*+1}}^s - \varepsilon$, so precedence holds. Each iteration reduces IS_{total} by ε; after at most $q-1$ iterations $\sum_k \Delta_k = 0$. □

This proposition guarantees a local improvement in intermediate storage under the UIS assumption. However, it does not constitute a proof of global optimality or convergence, which remains an open challenge for future work.

4.5.3 Lower Bounding Techniques

Accelerate search via:

- **Zero-Lag Bound**: Optimistic assumption:

  $$C_p = d_p \quad \text{and} \quad is_i = 0 \quad \forall \text{ unscheduled } i, p$$

  Provides aggressive pruning for early-stage nodes
- **Critical-Path Bound**: Computes longest delay path through precedence arcs:

  $$\text{CP}_{\text{delay}} = \max_{\text{paths}} \left(\sum_{(i',i)} (t_i^s - t_{i'}^f) \right)$$

  Incorporates temporal and storage latency.
- **IST Heuristic**: For unscheduled task i, estimate minimal feasible is_i considering:
  - Temporal proximity to immediate predecessors
  - Projected equipment availability
  - Policy constraints (NIS/UIS)

4.5.4 Multi-Objective Trade-Off and Dominance Consideration.

Although Pareto frontiers are not empirically constructed here, the scalarized cost function

$$Z = \alpha \sum_{i \in I_{\text{sched}}} is_i + \beta \sum_{p \in P_{\text{sched}}} (E_p + T_p)$$

implies an underlying trade-off surface between IST and due-date penalties. Varying α and β across runs could approximate a frontier. Since both objectives may conflict, future implementations could apply dominance checks to improve pruning. The non-convexity of this space—due to discrete task sequencing—also suggests that metaheuristic or adaptive scalarization techniques may better approximate trade-offs in practice.

4.6 Policy-Aware Search Adaptation and Computational Complexity Analysis

The proposed B&B algorithm dynamically adjusts its search strategy based on the selected reactive rescheduling policy, directly influencing branching complexity and computational requirements. Let $n = |\overline{I}_k^* \cup I_{k+1}|$ be the number of unscheduled tasks and $m = |J|$ the number of machines. Each policy imposes distinct constraints:

- **Policy 1 (Append-only):** Tasks in $\overline{I}_k^*$ remain fixed in sequence and timing. New tasks I_{k+1} are simply appended. This results in low branching complexity $O(nm)$, due to the linear number of new task-machine assignment decisions. However, limited flexibility often results in suboptimal solutions concerning due-date and intermediate storage objectives.
- **Policy 2.1 (Idle-Gap Insertion):** The sequence and start times of tasks in $\overline{I}_k^*$ remain fixed, but new tasks can be inserted into available idle time windows. This moderately increases flexibility and computational complexity to $O(n^2m)$ due to the polynomial number of task-slot-machine checks required. The policy balances due-date adherence but may restrict task placement significantly when idle time is scarce.
- **Policy 2.2 (Fixed Sequence, Shifted Starts):** Tasks in $\overline{I}_k^*$ retain their sequence but can shift their start times within constraints. This additional temporal flexibility significantly improves the potential for optimal due-date alignment, increasing branching complexity to approximately $O(n^3m)$, influenced by temporal discretization and resource availability.
- **Policy 3 (Full Rescheduling):** Offers maximum flexibility, allowing full resequencing and reassignment of all unscheduled tasks $\overline{I}_k^* \cup I_{k+1}$. This combinatorial freedom leads to factorial complexity $O(n! \cdot m^n)$, making it computationally intensive but optimal in solution quality.

The selected policy fundamentally shapes algorithm performance and suitability for real-time implementation:

- **Branching Complexity:** Ranges from linear (Policy 1) to factorial (Policy 3).
- **Constraint Handling:** Policies 1 and 2 enforce hard constraints, significantly reducing flexibility.
- **Solution Quality:** Flexibility directly correlates with the potential to optimize E/T penalties and IST.
- **Real-Time Viability:** Lower-complexity policies (1, 2.1) offer rapid solutions suitable for real-time scenarios but sacrifice optimality, whereas higher-complexity policies (2.2, 3) achieve better optimization at increased computational cost.

Although the underlying combinatorial complexity remains fundamentally invariant to the chosen objective functions, incorporating multi-objective criteria—such as E/T and IST—affects the algorithm's practical efficiency. Specifically, multi-objective considerations influence the effectiveness of lower-bounding techniques and branch pruning, resulting in increased computational overhead under highly flexible policies (e.g., Policy 3) despite identical theoretical complexity.

5 Discussion on Trade-offs

The proposed reactive scheduling framework inherently balances multiple competing factors, such as flexibility, computational complexity, due-date adherence (E/T penalties), and storage costs. Table 1 succinctly illustrates these trade-offs across the defined rescheduling policies.

Table 1. Qualitative Trade-off Comparison Across Rescheduling Policies

Policy	Flexibility	Comp. Time	E/T Penalty	Storage Penalty
1	Low	Low	High	High
2.1	Medium	Medium	Moderate	Moderate
2.2	High	Med–High	Low	Moderate
3	Very High	High	Lowest	Lowest

Policy 1 is suitable for environments prioritizing computational efficiency and scheduling stability, yet its limited flexibility typically results in higher penalties for both due-date adherence and intermediate storage. Policies 2.1 and 2.2 incrementally enhance flexibility, progressively improving the schedule's alignment with due-date targets while moderately controlling storage costs at the expense of increased computational complexity. Policy 3 offers maximal flexibility, achieving the lowest penalties for both due-date violations and intermediate storage. However, this optimality comes at the highest computational cost, making it best suited for scenarios where solution quality outweighs computational constraints.

Future modelling extensions should explore more nuanced scenarios, including hybrid or adaptive storage policies, dynamic task prioritization, and robustness analysis under stochastic disruptions, to deepen insights into these complex trade-offs.

6 Conclusion and Future Work

This paper establishes a conceptual foundation for integrating reactive scheduling with due-date management and intermediate storage constraints in flexible job shop environments. The proposed framework introduces a multi-objective branch-and-bound algorithm that balances E/T penalties and IST, structured around

four reactive rescheduling policies that progressively increase scheduling flexibility.

A key theoretical contribution is the Storage-Gap Adjustment Proposition, which provides a constructive approach for local IST minimization under UIS constraints via targeted slack redistribution. While this result offers promising structure-preserving improvement, it remains conceptual and lacks empirical implementation at this stage.

Importantly, the framework evaluates candidate schedules using a weighted cost function combining storage and due-date objectives. While scalarized in this study, the underlying trade-offs are inherently multi-objective. Varying the weighting coefficients would yield a Pareto frontier of non-dominated schedules, with each representing a distinct balance of E/T performance and storage cost. Future work will explore these frontiers explicitly, assessing convexity and dominance relationships to guide solution selection. As the objective space may be non-convex due to combinatorial constraints, such analysis can inform adaptive weight adjustment or hybrid optimization strategies.

Future research should also focus on computational benchmarking using simulated or industrial instances to evaluate policy-specific scalability and solution quality. Additionally, extending the model to incorporate uncertainty—such as stochastic order arrivals, equipment failures, and variable processing times—would improve robustness. Finally, integrating real-time decision support elements such as digital twins, reinforcement learning for policy selection, and cost-aware storage models would further enhance the framework's practical applicability in dynamic manufacturing environments.

Nomenclature

α, β	Multi-objective weights ($\alpha + \beta = 1$)
$\overline{I}_k^*$	Tasks in $\overline{S}_k$
$\overline{S}_k$	Unstarted assignments (reschedulable)
$\underline{I}_k^*$	Tasks in $\underline{S}_k$
$\underline{S}_k$	Started assignments (fixed) before t_{k+1}
C_p	Completion time of product p's last task
d_p	Due date for product p
$E_p = \max(0, d_p - C_p)$	Earliness of product p
I	Set of tasks
$I_i^- \subseteq I$	Immediate prerequisites for task i
$I_p \subseteq I$	Tasks for product p
is_i	Intermediate storage time for task i
J	Set of equipment units
$J_i \subseteq J$	Equipment capable of performing task i
O	Set of orders $\{o_1, o_2, \ldots, o_n\}$
P	Set of products
$pt_{i,j}$	Processing time of task i on unit j
S_k	Schedule at order k's arrival
t_i^s, t_i^f	Start/finish time of task i
t_k	Arrival time of order o_k
$T_p = \max(0, C_p - d_p)$	Tardiness of product p
w_p^E, w_p^T	Earliness/tardiness weights for product p

References

Bahroun, Z., Shamayleh, A., As'ad, R., & Zakaria, R. (2024). Integrated proactive-reactive tool for dynamic scheduling of parallel machine operations. *International Journal of Engineering Business Management*, *16*, 18479790241301164. https://doi.org/10.1177/18479790241301164

Bakon, K. A., & Holczinger, T. (2024). S-graph-based reactive scheduling with unexpected arrivals of new orders. *Machines*, *12*(7). https://doi.org/10.3390/machines12070446

Bakon, K. A., & Holczinger, T. (2025). Addressing due date and storage restrictions in the s-graph scheduling framework. *Machines*, *13*(2). https://doi.org/10.3390/machines13020131

Cheng, T., & Gupta, M. (1989). Survey of scheduling research involving due date determination decisions. *European Journal of Operational Research*, *38*(2), 156–166. https://doi.org/https://doi.org/10.1016/0377-2217(89)90100-8

Cheng, T., & Jiang, J. (1998). Job shop scheduling for missed due-date performance. *Computers & Industrial Engineering*, *34*(2), 297–307. https://doi.org/https://doi.org/10.1016/S0360-8352(97)00317-3

Ha, J.-K., Chang, H.-K., Lee, E. S., Lee, I.-B., Lee, B. S., & Yi, G. (2000). Intermediate storage tank operation strategies in the production scheduling of multi-product batch processes. *Computers & Chemical Engineering*, *24*(2), 1633–1640. https://doi.org/https://doi.org/10.1016/S0098-1354(00)00438-5

Herroelen, W., & Leus, R. (2004). Robust and reactive project scheduling: A review and classification of procedures. *International Journal of Production Research*, *42*(8), 1599–1620.

Herroelen, W., & Leus, R. (2005). Project scheduling under uncertainty: Survey and research potentials [Project Management and Scheduling]. *European Journal of Operational Research*, *165*(2), 289–306. https://doi.org/https://doi.org/10.1016/j.ejor.2004.04.002

Holguin Jimenez, S., Trabelsi, W., & Sauvey, C. (2024). Multi-objective production rescheduling: A systematic literature review. *Mathematics*, *12*(20), 3176.

Kim, M., Jung, J. H., & Lee, I.-B. (1996). Optimal scheduling of multiproduct batch processes for various intermediate storage policies. *Industrial & Engineering Chemistry Research*, *35*, 4058–4066. https : / / api . semanticscholar . org / CorpusID : 96817294

Kopanos, G. M., & Pistikopoulos, E. N. (2014). Reactive scheduling by a multiparametric programming rolling horizon framework: A case of a network of combined heat and power units. *Industrial & Engineering Chemistry Research*, *53*(11), 4366–4386. https://doi.org/10.1021/ie402393s

Kopanos, G. M., & Puigjaner, L. (2009). A milp scheduling model for multi-stage batch plants. In *Computer aided chemical engineering* (pp. 369–374, Vol. 26). Elsevier.

Koulamas, C. (2011). A unified solution approach for the due date assignment problem with tardy jobs. *International Journal of Production Economics*, *132*(2), 292–295.

Mansini, R., Zanella, M., & Zanotti, R. (2023). Optimizing a complex multi-objective personnel scheduling problem jointly complying with requests from customers and staff. *Omega*, *114*, 102722.

Mastrolilli, M., & Svensson, O. (2011). Hardness of approximating flow and job shop scheduling problems. *Journal of the ACM (JACM)*, *58*(5), 1–32.

Mihály, K., & Kulcsár, G. (2023). A new many-objective hybrid method to solve scheduling problems. *International Journal of Industrial Engineering and Management*, *14*(4), 326–335.

Ojstersek, R., Tang, M., & Buchmeister, B. (2020). Due date optimization in multi-objective scheduling of flexible job shop production. *Advances in Production Engineering & Management*, *15*(4), 481–492.

Ouelhadj, D., & Petrovic, S. (2009). A survey of dynamic scheduling in manufacturing systems. *Journal of scheduling*, *12*, 417–431.

Portoleau, T., Artigues, C., & Guillaume, R. (2020). Robust predictive-reactive scheduling: An information-based decision tree model. *Information Processing and Management of Uncertainty in Knowledge-Based Systems: 18th International Conference, IPMU 2020, Lisbon, Portugal, June 15–19, 2020, Proceedings, Part III*, *1239*, 479–492. https://doi.org/10.1007/978-3-030-50153-2_36

Romero, J., Puigjaner, L., Holczinger, T., & Friedler, F. (2004). Scheduling intermediate storage multipurpose batch plants using the s-graph. *AIChE Journal*, *50*(2), 403–417.

Ruiz-Vanoye, J. A., Díaz-Parra, O., & Zavala-Díaz, J. C. (2011). Complexity indicators applied to the job shop scheduling problem to discriminate the best algorithm. *International Journal of Combinatorial Optimization Problems and Informatics*, *2*(3), 25–31.

T, V. K., A, J. O., & Setupathi, R. (2017). Multi-objective comparison of due-date assignment methods in a dynamic job shop with sequence dependent setup time. *International Journal of Engineering Research & Technology (IJERT)*, *6*(04).

Vanhoucke, M. (2002, December). *Optimal Due Date Assignment In Project Scheduling* (Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium No. 02/159). Ghent University, Faculty of Economics and Business Administration.

Vieira, G. E., Herrmann, J. W., & Lin, E. (2003). Rescheduling manufacturing systems: A framework of strategies, policies, and methods. *Journal of scheduling*, *6*, 39–62.

Wang, L., Luo, C., & Cai, J. (2017). A variable interval rescheduling strategy for dynamic flexible job shop scheduling problem by improved genetic algorithm. *Journal of Advanced Transportation*, *2017*(1), 1527858.

A conceptual model to assess and develop digitalisation in the social economy's people care sector: the case of Prato (Italy), Mataro (Spain) and Varaždin (Croatia)

Teresa Maria Monllau Jacques
Universitat Pompeu Fabra, Department of Economic and Business, Ramon Trias Fargas, 25-27
08005 Barcelona, Spain
teresa.monllau@upf.edu

Jaime Lacueva Rueda
Tandem Social SCCL, Premià 15, 2a Planta
08014 Barcelona, Spain
jaimelacueva@tandemsocial.coop

Enrico Testi, Silvia Sordi
University Center City of Prato, Piazza dell'Università, 1
59100 Prato PO, Italy
enrico.testi@pin.unifi.it

Nikola Kadoić, Ivan Malbašić
University of Zagreb Faculty of organization and informatics, Pavlinska 2
42000 Varaždin, Croatia
{nkadoic, imalbasic}@foi.hr

Abstract. *This paper aims to provide researchers and policymakers with a model to assess and develop digitalisation in the social economy's people care sector. The model consists of internal and external factors. After the model had been theoretically conceptualised, it was evaluated using the survey that included 58 organisations: 49 participated in the questionnaire, and nine were visited and interviewed. The organisations came from Prato (Italy), Mataro (Spain), and Varaždin (Croatia). In all three, there are good examples regarding digitalisation, and also examples of organisations still in the beginning phase of considering digitalisation. Interview results confirmed this. Additionally, the authors presented digital tools used in the field and gave some recommendations as a starting point to foster the digitalisation process in social organisations dedicated to personal care services.*

Keywords. social economy; digitalization; people care

1 Introduction

We are currently navigating through what is termed "The Fourth Industrial Revolution," a period distinguished by the rapid pace of change and its far-reaching impacts. According to Schwab (2017), this era is characterised by two principal factors: Velocity, as the global landscape is increasingly interconnected, leading to accelerated transformations; Convergence of Spheres, as there is a growing blur between technological, physical, and biological realms, influencing how individuals live, work, and the structuring of societies. Both the European Commission (EC) and the Organisation for Economic Co-operation and Development (OECD) prioritise digital transformation on their political agendas. The EC emphasises the merging of technologies and the amalgamation of physical and digital systems to foster innovation environments within Europe. Conversely, the OECD highlights the synergy between data and technology as pivotal for societal transformation. In the realms of economics and business, the COVID-19 pandemic has expedited the digital transformation process, condensing years of change into mere months. The resilience of economic activities during this crisis has been notably higher in sectors with advanced digital integration. The "Recovery Plan for Europe," proposed by the EC, underscores the significance of digitalisation within the business sector. In response to the economic and social repercussions of the pandemic, the EC has introduced a series of measures focused on economic restoration. These include strategies aimed at equitable climate and digital transitions, facilitated through the Just Transition Fund and the Digital Europe Programme. These initiatives ensure that transitions toward climate neutrality, digitalisation, and demographic shifts are equitable, supporting the success of the European Green Deal and the 2030 Digital Decade for all Europeans.

In this research, the people care sector refers to industries, services, and sectors that focus on the well-being, support, and care of individuals, particularly those in need of assistance due to age, illness, disability, or other personal circumstances. It encompasses a broad range of services designed to ensure people's physical, emotional, social, and mental well-being. These services include health care, social care, elderly care, disability support, mental health services, child care and personal care. The sector of

people care is not immune to the imperatives of digitalisation. It faces the challenge of adapting to new operational and relational dynamics. Within this context, the people care ecosystem must identify tailored digital solutions to navigate its transition, adhering to principles of inclusivity and equality. Digital Missions for Care Social Economy's Resilience (DIMCARE) is a project financed by the European Union to catalyse digitalisation within the social economy's people care sector. It advocates for adopting new technologies and digital social innovations, recognising the sector's unique digitisation challenges due to its reliance on on-site services. DIMCARE promotes the integration of digital tools into core service activities, highlighting the role of EU-supported local and regional administrations in facilitating this transition.

This paper aims to offer research and policy-making bodies a framework for evaluating digitalisation within social economy organisations in the people care sector, fostering the development of digitalisation initiatives. It aspires to be a benchmark for promoting digital transformation within organisations or industry sectors. By providing insights into the implementation of digitalisation across different European territories, the study aims to enhance understanding and facilitate the efficient adaptation of organisations to digital imperatives. Additionally, it seeks to lay the groundwork for transnational and inter-regional collaboration among social economy stakeholders. Related to digital maturity, this concept in the people care sector refers to the extent to which organizations have integrated digital technologies, data-driven decision-making, and digital processes into their operations to enhance care delivery, improve efficiency, and meet the evolving needs of individuals they serve.

There are three research goals in this paper:

1. To introduce and evaluate the model of digital maturity of people care sector organisations,
2. To define and evaluate the levels of digital maturity inside social organisations in the people care sector,
3. To present the digital tools that are used in people care sector.

This paper is organised as follows: Section 2 presents a literature review on digitalisation in the people care sector. Section 3 presents digitalisation in people care sector in three countries. Section 4 presents the methodology of the research. Section 4 presents the proposed model to assess digitalisation in the social economy. Section 5 presents the results. Finally, Section 6 concludes the paper and gives recommendations to foster the process of digitalisation.

2 Assessing digitalisation

Despite the recognised importance of digital transformation within economic organisations, there is a noticeable dearth of studies proposing models to evaluate digital maturity in these entities. This gap is even more pronounced when focusing on organisations within the social economy, where existing literature is limited and often originates from professional rather than academic circles. The available studies tend to adopt a qualitative approach, rather than a quantitative one (Roman and Rusu, 2022), and are sometimes managed by social organisations interested in the development of the social economy within their contexts. (Taula del Tercer Sector, 2024, Scottish Council for Voluntary Organisations, 2024) A segment of academic research links the digitalisation process directly to digital innovation initiatives (Khin and Ho, 2019).

To advance the process of digital innovation, organisations must commit to embracing emerging technologies (Khin and Ho, 2019). It is beneficial for firms to initially outline potential applications of digital technologies and the value they can add (Sjodin et al., 2018). However, digital components are sometimes integrated without a thorough understanding of customer needs or a unique value proposition Gebauer et al., 2005). The digitalisation process within an organisation should be considered holistically, encompassing both back and front office operations, and taking into account both internal resources and external factors, including political variables. Collaboration both within and outside the organisation is crucial. The digitalisation process generates value through collaboration among various stakeholders within the ecosystem. As highlighted by Parida (Parida et al., 2019), often, the realisation of digital value creation extends beyond a firm's boundaries, involving collaborative efforts across networks (Ehret and Wirtz, 2016; Hakanen and Rajala, 2018; Loebbecke and Picot, 2015). To analyse the level of digitalisation in a one sector, therefore, it is vital to consider the influence of external and internal factors that encourage digitalisation.

Regarding external factors, it is necessary to consider two variables:

- the support from public administrations and the importance that organisations that operate inside the sector give to the process. The support from governments to promote digitalisation within organisations can lead to a favourable environment for implementing digital processes and reinforce positive attitudes towards digitalisation; effectively, as a consequence of COVID 19, a lot of European countries promoted policies to foster the process of digitalisation inside organisations. One example of these programmes was the programme: Next Generation EU.
- the second external factor that must be taken is the support of the sector where organisation develop their activity. Inspired by the Theory of Planned Behaviour (Ajzen, 1991) we consider that Relational Support plays an important role in the process of digitalisation. This concept refers to

support for introducing and developing the process of digitalisation from the whole of organisations that belong to a sector, which can exert pressure to adopt such practices.

The definition of internal factors that influence the development of the process of digitalisation is more complex. Reviewing the academic literature, we conclude that concepts such as digital orientation and digital capability are important. A strong technological orientation is essential. Gatignon et al. (1997) defined technological orientation as a firm's dedication to applying new technology and being responsive to technological changes. According to Khin and Ho (2019), alongside the importance of a digital technology orientation, a firm must also possess the capability to manage and optimise the use of digital technology efficiently.

Thus, enhancing an organisation's digitalisation process requires it to have the digital capability to utilise digital technology effectively. Grant (1996) and Ritter and Pedersen (2019) suggest that an organisation is capable if it can perform a productive task repeatedly, directly or indirectly contributing to its capacity for creating value by transforming inputs into outputs.

When applied to digital capacity, this concept emphasises focusing on digital skills. These skills range from basic competencies essential for functioning in the current economic environment to more advanced abilities for working with data generated through digitalisation. This includes the capability to use communication tools, business process support tools, and platforms for website creation. Moreover, employees must possess skills for storing and analysing data, as well as the permissions for its use. Data management involves not only its generation, storage, and analysis but also sharing it with other organisations within the sector. Therefore, digitalisation, value creation, and collaboration should be considered beyond the boundaries of individual firms, especially in SMEs. Knowledge creation, development, transfer, and exchange inside the ecosystem are operated among economic agents and non-economic parties, such as technology, institutions, sociological interactions, and culture (Liu et al, 2022).

Additionally, under the project DIMCARE, an instrument to measure the digital level of social economy organisations was developed by Kadoić et al (2024). This instrument is mainly focused on internal factors and is used at the level of individual organisations in the people care sector while the external factors are only implicitly included.

Since the development of the digitalisation of the organisation in the people care sector, a certain geographical area is not a matter of the individual organisation and is not only influenced by internal factors; there is a need to develop a more comprehensive model that will include external factors and be used at the level of all organisations that are active in different geographical locations.

3 The digitalisation in people care sector in Prato, Mataro and Varaždin

The theoretical model proposed in this paper and the questionnaire used to validate it were the result of a qualitative field analysis of digitalisation in the people sector in three countries: Croatia, Italy and Spain. Here we describe the state in three cities and countries regarding the digitalisation in people care sector.

1. Croatia: Even, currently the term of social entities (associated with people care sector) in Croatia is still a blur, Croatia embedded the notion of solidarity a long time ago. Its strong legacy has nourished the development of new forms of economic cooperation promoting solidarity and mutual self-help. The Croatian legal systems, although not conceived for social enterprises, the existing legislation allows for a certain degree of flexibility to identify the following legal forms adopted by social enterprises in Croatia: Associations, Social Co-operatives, Private Limited Liability Companies. The legislation related to social entities is still a big challenge in Croatia. There is no systematic tracking of the application of the digital transformation paradigm in any sector as a whole, nor tracking of digital transformation in social entities (the term “social entities” by its definition is still unclear). However, there is some partial research in this manner, related to specific types of institutions. The first example is related to some research that focuses on the digital maturity of higher education institutions. The second example is related to the Croatian company Apsolon, which created the HDI (cro. *Hrvatski digitalni indeks*, eng. Croatian Digital Index). Both of those studies resulted in the instrument for measuring digital maturity level, but for different purposes: the first for higher education institutions, and the second for business organisations.
2. Italy: Italy has a long-lasting tradition in social economy (which mathes the term people care sector). The first organisations date back to the Middle Ages, namely the catholic brotherhoods, the “Misericordie” in Florence in 1244, and the “Monti di Pietà” set up by Franciscan friars in 1462 in Turin, which provided credit to people in need. The social economy sector in Italy, is vibrant and diverse, encompassing a range of organisations, such as cooperatives, mutuals, and non-profit organisations. These organisations provide essential services and support to the local community, including employment opportunities, training, and access to affordable goods and services. In Italy, according to the report “Digital transformation: Shaping the future of European Healthcare” (Deloitte Center for Helth Solutions, 2020), 6% of respondents at the Italian level

declared that their organisation has increased the use of digital technologies to support the work of health professionals following the COVID-19 emergency.

3. Spain: Social Economy is considered in Spain an economic sector of great relevance since it represents 10% of the Spanish GDP and it represents 12,5% of total Spanish employment. 64% of the social economies that develop their activity in the people care sector have a level of income lower than 2 million euros. 22,8% have a level of income bigger than 50 million euros. A study carried out by the Mataró Council (2022) about the level of digitalisation in the companies that develop their activity in the territory concluded that 75% of the companies know about "Digital Transformation", but only 55% of companies affirm that they have a Digital strategy. 75% of those companies consider that the digital strategy implies profit for the companies, the other 25% have not obtained results yet. The Board of Entities of the Third Social Sector carried out a study on digitalisation in social organisations (Taula del Tercer Sector, 2024). This study analyses the level of digitalisation in the third social sector in Catalonia. The conclusions of this study are that there are a lot of things to do if we are speaking about digitalisation and the people care sector. In Catalonia 62% of the entities have an intermediate level of digital maturity, which means that they are entities that are beginning to formalise a digital plan in some specific area. Most of them act only when it is necessary. They have a reactive strategy. It is very difficult for entities to make decisions according to the data obtained thanks to the process of digitalisation.

The state in the area of digitalisation of the people care sector in three countries shows big differences in the development of the people care sector, but also in the digitalisation of the people care sector. This variability enabled a comprehensive view of the overall state of the people care sector in the EU.

4 Methodology

The methodology developed for assessing and fostering digitalisation in the social economy's people care sector within the DIMCARE project follows a structured approach, incorporating insights from the categorisation of digital maturity levels developed by the Board of Entities of the Social Third Sector of Catalonia (2024) and the literature revision This methodological framework guides the selection of organisations and the design of the questionnaire to comprehensively evaluate digitalisation efforts within the sector.

This research is an example of the mixed research type since it combines qualitative and quantitative aspects. The research has been implemented on different organisations in the people care sector in Prato, Mataro and Varaždin. The methodology developed is multifaceted, incorporating both qualitative and quantitative approaches to capture a comprehensive understanding of the digital maturity levels of the organisations participating in the study. The mixed research design is used. The selection criteria for participating organisations were informed by the guidelines provided by Creswell and Creswell (2017) for qualitative and quantitative research design.

Quantitative research: conducting a survey related to digitalisation in the people care sector using a questionnaire. The quantitative aspect of the research utilised an online questionnaire distributed to a selected group of social organisations dedicated to the people care sector across the three territories. Local and regional authorities identified relevant and representative social organisations within their respective territories that aligned with the project's scope. The total number of organisations that have received the questionnaire was 83. The questionnaire design was informed by established practices in survey research (Regmi et al, 2016). The survey questions covered various aspects related to digitalisation in the people care sector, including the organisation's starting point on digitalisation, the attitude of staff towards digitalisation, sector-wide perspectives on digitalisation, and barriers and challenges faced. The questionnaire, disseminated via Google Forms, consists of 31 questions. The questionnaire was completed by a total of 49 social organisations (response rate is 59%), with 26 from Mataró, 12 from Varaždin, and 11 from Prato (Fig. 1). The research sample to which the survey was a full sample – the questionnaire was sent to all the organisations in the people care sector in those three geographical locations. Based on the conceptual framework, we provided a model to evaluate digital maturity. The results of the questionnaire were used in (1) providing and validating the model to assess and develop digitalisation in the social economy's people care sector and (2) defining the levels of digital maturity inside social organisations in the people care sector.

Qualitative research: study visits to the selected organisations in the area of people care sector in Prato, Mataro and Varaždin. A total of 3 social organisations were visited in Mataró, 2 in Varaždin, and 4 in Prato. The organisations were selected for visits based on importance and influence in their local environment, their responses on the quantitative survey and their availability to visit them at the particular moment, considering the project schedule. A semi-structured interview format was employed to allow for flexibility and depth in exploring the digitalisation efforts, challenges, and attitudes prevalent within the social organisations in each territory (Knott et al, 2022). The results of the qualitative research were used to polish the results of the quantitative results and to finalise the model to access and develop digital maturity in the

organisations in the people sector. Furthermore, study visits served as a tool to evaluate and update the descriptions of organisations in the people care sector that apply different levels of digital maturity considering different criteria.

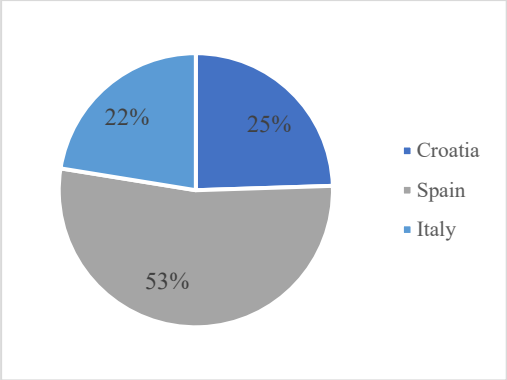


Figure 1. The number of survey participants (N=49)

The qualitative aspect of the research involved site visits and in-depth semi-structured interviews conducted in the three territories: Mataró, Varaždin, and Prato. This approach allowed for an initial exploration of the importance of digital transformation within social organisations operating in these regions.

5 The results

5.1 The model of digital maturity of people care sector organisations

After the questionnaire and study visits were implemented, it was possible to establish a model for the digital maturity of organisations. One of the tasks during the methodology development was to elaborate a model that considers both internal and external factors. To propose a model, we analysed the obtained data, formed an initial model, and further, through focus groups and deep analysis, updated the initial model. This model is depicted in Fig. 2.

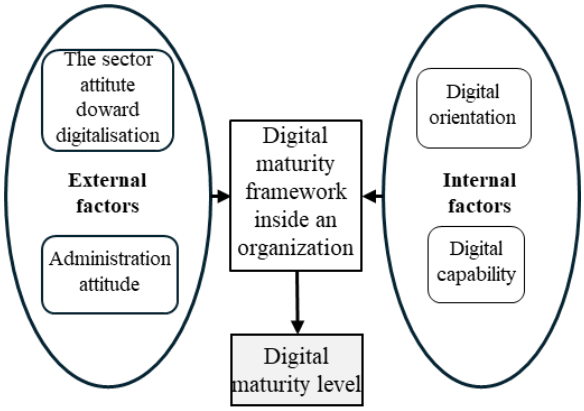


Figure 2. The model used in the DIMCARE project.

The model proposed has four dimensions: two internal, Digital Orientation and Digital Capability, and two external, sector attitudes toward digitalisation and Administration Attitudes. For evaluating the four dimensions we have identified the variables:

1. Implementation of the process of digitalisation in the front office and in the back office
2. The level of budget focused on digitalisation. Organisations with a strong digital orientation will devote more resources to digital transformation than those that do not.
3. Areas of organisation where the digitalisation process is implemented. Organisations with a high degree of Digital Orientation will implement digital transformation in more areas.
4. There is a department or a specific person in charge. An organisation with a person or department in charge of digitalisation demonstrates a greater digital orientation than organisations without it.
5. Digital capability is related to digital skills. Employees with positive attitudes toward digitalisation are more likely to have digital skills or to be interested in training them.
6. The sector attitudes. It is important the organisations that develop their activity inside the sector consider the process of digitalisation as a priority. In this way, the digitalisation process will be more efficient since organisations can share infrastructure, solutions or even share data that is produced as a consequence of the activity carried out. Administration attitudes toward digitalisation. If public administrations are aware of the importance of the process of digitalisation, they will promote different programmes with the aim of increasing the level of digitalisation inside organisations. All of those dimensions are influenced by the specific characteristics of entities that belong to the sector, like size defined by the number of employees and the amount of income, the average age of employees and the legal structure, among others.

The variables used to measure internal and external factors have been established according to the existing literature, to which we have referred in Section 2.

According to the existing literature, the support of public administrations is essential in promoting the digitalisation process in organisations; the Next Generation EU programme is proof of this. The attitude of the sector is another of the external factors that most influence the digitalisation process. Social organisations that develop their activity in the people care sector are characterised by being small in size and having few resources. For this reason, the digitalisation process in an organisation will hardly occur if there are no synergies/collaboration between the different organisations within the sector. This is why the attitude of the sector has been used as one of the variables to measure external factors.

The variables that define the internal factors have been defined according to the contributions made by Gatignon et al (1997) when they define the concept of technological orientation (identified by us as technological orientation) and the contributions made by Khin and Ho (2019), among others, when they state that organisations, beyond technological orientation, must have the capacity to manage digital technology. The variables that define digital capability have been established based on the definitions provided by Ritter et al. (2020), among other authors.

5.2 Levels of digital maturity inside social organisations in the people care sector

Different authors describe different scales to define the level of digital maturity in a sector. According to Remane et al. (2017), the predominant literature is argumentative. The authors refer, among others, to the work carried out by Lichtblau et al. (2015), which defined three archetypes: newbies, beginners, and pioneers, and the report issued by PWC(2024), which defined four archetypes: digital novice, vertical integrator, horizontal integration, and digital champion. Pursuing the main objective of the present study, a methodological approach has been followed, based on a categorisation developed by the Board of Entities of the Social Third Sector of Catalonia (2024), that defines four possible levels of digital maturity for the social organisations:

1. Beginners: the social organisations are not really interested in the process of digitalisation and the digital transformation. They have no digital skills and they are not interested in obtaining them. They do not trust in the data available.
2. Intermediate: the social organisations have started some process of digital transformation in some areas or departments. They are aware of the importance of digital transformation but they present a reactive attitude. They develop some kind of transformation when necessary.
3. Advanced: the social organisations are implementing or have implemented digital planning in the organisation. They are aware of the importance of digitalisation and the importance of having ratios that measure the level of digital transformation. They have a team that is already engaged in digital transformation.
4. Senior: the social organisations innovate continuously. Technology and digitalisation are a key element in the development of their activity.

Consequently, when dimensions of the model are combined with the levels of digital maturity, it is possible to describe the characteristics of typical organisations, considering dimensions and levels of digital maturity. The result is visible in Table 1.

Table 1. Levels of digital maturity inside social organisations in the people care sector (defined by DIMCARE)

Dimension	Variables	LEVEL OF DIGITAL MATURITY			
		Beginners	**Intermediate**	**Advanced**	**Senior**
Digital Orientation	Implementation of the process of digitalisation in the front office and in the back office	None	Only one	It is entirely developed in one of them and has been started in the other one	Both front and back office
	The level of budget focused in the process of digitalisation	0%	0-10%	10%-25%	More than 25%
	Areas of organisation where the process of digitalisation is implemented	0	1-2	3-4	+4
	Digital solutions used in the process of digitalisation	No	Simple tools	Machine learning	Machine learning and artificial intelligence.
Digital Capability	There is a department or a specific person in charge	No	Periodically the management has to report about it	There is a person in charge	There is a department
	Attitude of employees versus the process of digitalisation	None	Reactive	Active	Active with knowledge
	Level of digital competences of employees that belong to the organisation	None	Intermediate	Advanced	Specialist
Sector Attitudes	Data (analysis and sharing)	No	There is an analysis of data only when there is a problem	There is an analysis of data periodically	There is an analysis of the data periodically.The organisations share the data
	Importance in the sector	None	Moderate	High	Very high
Administration attitudes	Programmes carried out to foster digitalisation	None	Proposed programmes without carried out them	Proposed programmes with a moderate level of development	Proposed programmes with a high level of development

Also, according to the questionnaire responses, a correlation model was established to provide insights into how different aspects of digitalisation relate to the overall digital maturity of social organisations in the people care sector across the three territories of Mataró, Varaždin, and Prato. This model aimed to identify relationships between the level of digitalisation and various factors, including Digitalization Process Metrics: presence of digitalisation initiatives, departmental involvement, employee attitudes, extent of digitalisation across different areas, and resource allocation; Importance of Digitalization Metrics: sector perception of digitalisation importance, types of digital tools utilised, and data utilisation practices.

From our point of view, organisations are currently more reactive than active, although there are some exceptions. In those exceptions, collaboration between organisations and local administrations plays an important role, and it is very important that the employees have digital skills, especially at the management level.

The digitalisation enablers:

1. The entity's size is determined by the number of employees and the level of turnover. Larger entities tend to have a higher level of digitalisation.
2. While the presence does not always imply digitalisation, having advanced technological infrastructure facilitates the implementation of digitalisation in the long term.
3. The extent of technical expertise within the management team directly correlates with the organisation's level of digitalisation.
4. A higher proportion of young employees in the organisation generally makes digitalisation easier, although it doesn't necessarily guarantee a higher predisposition.
5. If the sector prioritises digitalisation and public administrations actively promote it, organisations feel compelled to pursue it.

5.3 Digital tools in the social economy

In the DIMCARE project, tools used in the context of digitalisation of social enterprises and other organisations in the field of social economy occupy a central place. Based on a review of the available literature and research in the field, we have divided the digital tools used in this area into three areas. These areas are:

1. Digital tools and devices used for the needs of the main business processes in the area,
2. Digital tools as support for business processes,
3. Communication tools (which are actually a subgroup of the second group, but due to their greatest use they are singled out separately).

Below are descriptions of those tools. Additionally, we gave examples of concrete software for each group to describe the group better.

5.3.1 Communication tools

Communication tools include various computer and mobile applications that are used for everyday communication in all sectors and on an individual level. Considering that we are used to different applications for sending and receiving instant messages and have different profiles on social networks, it is precisely these tools that are easiest to keep as the ones through which we communicate in the field of social economy. These include:

1. E-mail addresses - most often those with which we are registered on various other services; these are most often Google mail and Apple mail, which we connect to Android and Apple mobile devices and can also be used via a computer. In other cases, these are all other e-mail addresses from different services, or official web addresses that users use at work.
2. Applications for sending and receiving instant messages and audio and video calling, such as Viber, WhatsApp, Telegram, etc.
3. Different social networks that allow both sending and receiving instant messages and audio and video calling, such as Facebook and its Messenger, Instagram or Tik Tok.
4. More recently, tools for holding real-time video meetings and conferences, including: Zoom, Webex, Jitsi, BigBlueButton, MS Teams and others.
5. For older users, we are talking about standard SMS messages.

These communication tools in the field of social economy are used by employees and volunteers for mutual coordination, but also for communicating with their users, clients or their representatives.

5.3.2 Digital tools as support for business processes

This group of tools includes the previously mentioned communication tools that we singled out because of their specificity, but also all other tools that serve to administer business processes, documenting, dissemination, and the like. These include:

1. Platforms for creating websites (Wordpress, Joomla, Wix, ...)
2. Tools from the Office package (Word, Excel, PowerPoint, Access...) or another similar package (Libre Office or Google documents) that are used to create documents, slides, budgets, analyses, databases, records and the like.
3. Tools for organising documents within the organisation (Google Drive, Dropbox, OneDrive, Alfresco Share, etc.)
4. Tools for electronic learning (Moodle)
5. Tools for member registration, online payment, event management and other association needs (MemberClicks Oasis and MemberClicks Atlas; Wild Apricot, Perosinify360, Mailchimp)

6. Crowdfunding tools (Indiegogo, Kickstarter, GoFundMe, Kiva, Zopa)
7. Tools for sale (webshop) (Shopify, Wix, BigCommerce, WooCommerce, PrestaShop, Weebly, ...)
8. At the end of this analysis, it is worth mentioning the large language models that can support the business processes.

Here, we include all the tools accepted at the social organisation level in the context of digital transformation and serve various support functions in the organisation.

5.3.3 Digital tools and devices used to meet the needs of the main business processes in the field of social economy

When we use a tool or device for the main activity of a social enterprise, we classify it in this group and not in the previous one. These are very specialised tools and devices that are targeted for a specific activity. We include:

1. The equipment and programs used in the sensory room used in the care of specific types of patients,
2. Equipment and programs for communicating with people who have specific disabilities (e.g. equipment and software that the famous Steven Hawking used to communicate with other people),
3. Various tools that help visually impaired and blind people monitor content from computers and mobile phones.

6 Conclusions

This paper was developed within the framework of the DIMCARE project. The primary objective of this paper is to consolidate the results produced during the 18-month project into a concise document that highlights the key scientific contribution: a model for assessing the digital maturity of organizations in the people care sector. Digital maturity in this sector is a multifaceted concept influenced by both internal and external factors. In this paper, we present and apply the model to organizations within the people care sector from the countries involved in the project. This model can serve as a valuable tool for future researchers and practitioners in the digitalization of the people care sector.

Based on the findings, we formulated four main recommendations for advancing digitalization in the people care sector. These recommendations are as follows:

1. Partnerships and Collaboration between Organizations: Collaboration is vital and can take two forms: (a) Collaboration within the same sector operating in the same geographical area (e.g., sharing websites, maintaining services, and ensuring cybersecurity) and (b) Collaboration between organizations, local governments, and universities/research centers, with universities and research centers playing a key role in training, designing, implementing, and promoting digitalization. Local governments should maintain a registry of social organizations and incentivize digital transformation by offering subsidies and training programs within communities.
2. Adopt a Proactive Approach: Organizations should be proactive rather than reactive in anticipating future needs. Collaborating with universities and research centers is crucial in this process. Additionally, establishing a dedicated department responsible for digital transformation is important. This department should possess broad expertise in areas such as healthcare, engineering, education, economics, and more. .
3. Digitalization of Both Back and Front Office Operations: Ensuring the continuity and efficiency of organizational processes requires the digitalization of both back-office and front-office functions.
4. Investing in Cybersecurity: Given the sensitive nature of data handled in sectors like healthcare and finance, investing in cybersecurity is essential.

The main limitation of this research lies in the limited number of organizations included in the model's evaluation and the reliance on self-reported data. Future research will involve further evaluation and refinement of the model, with adjustments made based on technological advancements.

Acknowledgments

The DIMCARE project fully supports this paper (https://sbflorence.org/dimcare/, accessed on October 28, 2024). The project included the following organisations: The Pompeu Fabra University (coordinator), Social Business City Barcelona (co-coordinator), the Universitat Politècnica de Catalunya, Polo Universitario Città di Prato (PIN, University of Florence), University of Zagreb Faculty of Organization and Informatics, Municipality of Prato, City of Mataro, and Varaždin County.

We thank participants of the project DIMCARE:

- Aina Tella, Universitat Politècnica de Catalunya BarcelonaTech (UPC), Spain,
- Alejandra Lara, The municipality of Mataro, Spain
- Aleksandar Bedeniković, Varaždin County, Croatia,
- Andrea Balletbo, Tandem Social SCCL, Spain
- Barbara Palla, University Center City of Prato, Italy,
- Berta Sellarès Gómez, Pompeu Fabra University, Spain,

- Didac Ferrer, Universitat Politècnica de Catalunya BarcelonaTech (UPC), Spain,
- Francesca Silipo, City Council of Prato, Italia,
- Giada Selvi, City Council of Prato, Italia,
- Josep Niubó, Pompeu Fabra University, Spain,
- Larisa Križan Breljak, Varaždin County, Croatia,
- Laura Strupar, University of Zagreb Faculty of organisation and informatics, Croatia
- Luca Mattesini, University Center City of Prato, Italy,
- Marta Tobias, Varaždin County, Croatia,
- Miguel García, Tandem Social SCCL – during the project, Spain
- Noemi Imprescia, City Council of Prato, Italy,
- Paolo Fusi, University Center City of Prato, Italy,
- Sara Monti, City Council of Prato, Italy,
- Sergio Morales, The municipality of Mataro, Spain,
- Silvia Lizana, The municipality of Mataro, Spain,
- Sonia Garcia Esteban, Universitat Politècnica de Catalunya BarcelonaTech (UPC) – during the project, Spain.

References

Schwab, K. The Fourth Industrial Revolution; Portfolio: 2017.

Roman, A.; Rusu, V.D. Digital technologies and the performance of small and medium enterprises. Studies in Business and. Economics. 2022, 17(3), 190–203. https://doi.org/10.2478/sbe-2022-0055

Taula del Tercer Sector (TTS). Estudi sobre l'estat de la digitalització al tercer sector. Accessed April 18, 2024. Available online: https://www.tercersector.cat/sites/default/files/2022-02/t3s_2022-02-03_informe_estat_digitalitzacio_tercer_sector.pdf

Scottish Council for Voluntary Organisations (SCVO). Taking a place-based approach to digital inclusion: The story so far. Accessed April 18, 2024. Available online: https://files.scvo.scot/2022/03/SCVO-Digital-Taking-a-place-based-approach-to-digital-inclusion-the-story-so-far.pdf

Khin, S.; Ho, T.C. Digital technology, digital capability and organisational performance. International. Journal of Innovation Science 2019, 11(2), 177–195. https://doi.org/10.1108/ijis-08-2018-0083

Sjödin, D.; Parida, V.; Leksell, M.; Petrović, A. Smart factory implementation and process innovation. Research-Technology Management 2018, 61(5), 22–31. https://doi.org/10.1080/08956308.2018.1471277

Gebauer, H.; Fleisch, E.; Friedli, T. Overcoming the service paradox in manufacturing companies. European Management Journal 2005, 23(1), 14–26. https://doi.org/10.1016/j.emj.2004.12.006

Parida, V.; Sjödin, D.; Reim, W. Reviewing literature on digitalisation, business model innovation, and sustainable industry: Past achievements and future promises. Sustainability 2019, 11(2), 391. https://doi.org/10.3390/su11020391

Ehret, M., & Wirtz, J. Unlocking value from machines: business models and the industrial internet of things.Journal of Marketing Management 2016, 33(1–2), 111–130. https://doi.org/10.1080/0267257X.2016.1248041

Hakanen, E.; Rajala, R. Material intelligence as a driver for value creation in IoT-enabled business ecosystems. Journal of Business & Industrial Marketing 2018, 33(6), 857–867. https://doi.org/10.1108/jbim-11-2015-0217

Loebbecke, C.; Picot, A. Reflections on societal and business model transformation arising from digitisation and big data analytics: A research agenda. The Journal of Strategic Information Systems 2015, 24(3), 149–157. https://doi.org/10.1016/j.jsis.2015.08.002

Ajzen, I. The theory of planned behavior. Organisational Behavior and Human Decision Processes 1991 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Gatignon, H., & Xuereb, J.M. Strategic orientation of the firm and new product performance. Journal of Marketing Research, 1997, 34(1), 77–90.

Grant, R.M. Prospering in dynamically-competitive environments: Organisational capability as knowledge integration. Organisational Science 1996, 7(4), 375–387.

Ritter, T.; Pedersen, C.L. Digitization capability and the digitalisation of business models in business-to-business firms: Past, present, and future. Industrial Marketing Management 2020, 86, 180–190. https://doi.org/10.1016/j.indmarman.2019.11.019

Liu, Z.; Shi, Y.; Yang, B. Open innovation in times of crisis: An overview of the healthcare sector in response to the COVID-19 pandemic. Journal of Open Innovation 2022, 8(1), 21. https://doi.org/10.3390/joitmc8010021

Kadoić, N.; Monllau Jaques, T.; Lacueva Rueda, J. Measuring the digital maturity of organisations in the people care sector. Sustainabi lity 2024, 16(20), 9053. https://doi.org/10.3390/su16209053

Deloitte Center for Health Solutions. Digital Transformation: Shaping the Future of European Healthcare; September 2020.

Creswell, J.W.;Creswell, J.D.; Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (4th ed.). Sage, Newbury Park, 2017

Regmi, P.R.; Waithaka, E.; Paudyal, A.; Simkhada, P.; van Teijlingen, E. Guide to the design and application of online questionnaire surveys. Nepal Journal of Epidemiology 2016, 6(4), 640–644. https://doi.org/10.3126/nje.v6i4.17258

Knott, E.; Rao, A.H.; Summers, K.; Teeger, C. Interviews in the social sciences. Nat. Rev. Methods Primers 2022, 2, 73. https://doi.org/10.1038/s43586-022-00150-6

Remané, G.; Hanelt, A.; Wiesboeck, F.; Kolbe, L.M. Digital maturity in traditional industries – An exploratory analysis. Conference Paper, 2017. Available online: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1009&context=ecis2017_rp

Lichtblau, K.; Stich, V.; Bertenrath, R.; Blum, M.; Bleider, M.; Millack, A.; Schmitt, K.; Schmitz, E.; Schröter, M. Impuls-Industrie 4.0-Readiness. Impuls-Stiftung des VDMA: Aachen-Köln, 2015.

PWC. Industry 4.0: Building the digital enterprise. Accessed April 25, 2024. Available online: https://www.pwc.com/gx/en/industries/industries-4.0/landing-page/industry-4.0-building-your-digital-enterprise-april-2016.pdf

Caliński, T.;Harabasz, J.; A dendrite method for cluster analysis. Communications in Statistics 1974, 3(1), 1–27. https://doi.org/10.1080/03610927408827101

Gender equality in the field of healthcare

Nikica Daraboš
Permanent Mission Republic Croatia at UN and WHO
Geneva 1202, Switzerland
nikica.darabos@mvep.hr

Violeta Vidaček Hainš
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
vvidacek@foi.hr

Nevena Ladić
Univesity North
Jurja Križanića 31 b, 42 000 Varaždin, Croatia
nbiskup@unin.hr

Abstract. *Gender equality refers to equal rights pertaining to men and women in their private and professional lives. The objective of this paper is to analyze policies focused on equal representation of women and men in the fields of ICT in healthcare at the global and national levels. The analyzed sample was collected from publicly available data provided by the UN, WHO, OECD, Development, Eurostat, and the Croatian Bureau of Statistics, mostly covering the period between 2015 and 2023, depending on the publication, with the most recent statistics from 2023. Quantitative research methods were employed, with the most commonly used statistical techniques being descriptive statistics, time series analysis, and comparative statistics based on secondary data sources. Results showed a slight rise in women's equality in healthcare, depending on the development level of their respective areas. While our paper presents data and analysis on gender equality among medical doctors at global, European, and Croatian levels, the rapid growth of ICT in healthcare, combined with insights from other professions, highlights significant potential for advancing gender equality in this area within Croatia's healthcare system. More ICT efforts are needed to utilize these data for training and awareness.*

Keywords. Gender Equality, Healthcare, ICT, Information and Communication Technology, Croatia

1 Introduction

According to the European Institute for Gender Equality (2022) gender equality is the cornerstone of European policies and is of utmost importance for further development of any society. It reflects itself in equal opportunities for men and women in all spheres of life and includes equal pay for the same level of education, elimination of gender stereotypes, representation in decision-making, and prevention of gender-based violence (European Commission, 2021).

Women have increased their education and representation in professions so far dominated by men in the areas of Science, Technology, Engineering and Mathematics – STEM (UNESCO, 2017).

Information and communication technology in healthcare includes digital solutions for secure data processing, analysis of health information, and more efficient management of the healthcare system that improve the quality and accessibility of healthcare services. Despite the growing demand for professionals in the fields of information and communication technology in healthcare, women remain significantly underrepresented (OECD, 2018)

The gender gap is defined as the measurable disparity between women and men in the fields of information and communication technology in healthcare, and often stems from systemic discrimination, stereotypes, and unequal access to rights and resources (Gender Equality Strategy 2020-2025, 2020).

Gender equality in healthcare is one of the most important factors influencing health and high-quality healthcare. Despite slight progress made towards greater gender equality at work in healthcare, women are still encountering obstacles preventing them from equal promotion in their careers (World Health Organization, 2021). Hence, the principles of gender equality in the fields of information and communication technology within healthcare should be aligned with those upheld in STEM disciplines. (UNESCO, 2017). Gender bias in healthcare is opposite and describes the discrimination according to gender in the healthcare services with the goal to prevent, diagnose, and treat illness and injury in order to maintain or improve health (Bargeri et al, 2025).

The identified research question will be answered by analysing the theoretical background in healthcare at the global, European, and Croatian levels, with the

goal of identifying disparities and proposing actionable measures for the improvement of the status of gender equality in healthcare in the Republic of Croatia. The management of information in health care organizations supported by information and communication technology needs to be an important part of the health care organization's strategy (Fierz, 2004).

1.1 Gender Equality

Gender equality has been achieved when women and men have equal rights, responsibilities, and opportunities in all parts of society and when different interests, needs, and priorities of men and women are valued equally (Official European Union websites).

The European Union has developed different varieties of gender regimes, reflecting diverse institutional approaches among member states, which significantly influence the effectiveness of gender equality policies (Walby, 2004).

Despite growing global awareness, gender equality in science, medicine, and global health remains far from achieved, with systemic biases and structural barriers continuing to limit women's representation and advancement in leadership roles (Shannon et al., 2019).

Although there are certain positive signs of progress when it comes to representation of both genders in managerial roles within organizations, the term 'Glass Ceiling' used when referring to preventing women from being promoted to higher hierarchical levels in workplaces still exists (Nedović et al., 2015). Research has shown that women are as successful in performing their jobs as men; in spite of that, they still occupy a very small percentage of managerial positions. Legal standards on this issue have not been implemented, and actors of this breach have not been subject to any sanctions (Burušić-Barčan & Burušić., 2021).

Major obstacles faced by women in their careers can be classified into three groups: social, organizational and personal. (Pološki Vokić et al., 2017):

Considering the leadership style, the difference between women and men can vary and depends on numerous factors such as personal characteristics, social expectations, etc. Research results demonstrate that the style of female leaders is significantly different from the style of male leaders. (Nidogon Višnjić et al., 2018).

1.1.1 Gender Equality Policies

Gender equality policies refer to contemporary political measures whose goal is to promote the protection of women's rights as universal human rights (European Institute for Gender Equality, 2008).

Initiated by international human rights movements and organizations like the United Nations, gender equality policies emerged as an important global management aspect (United Nations, Gender Equality Policy). Namely, the UN policy on gender equality includes the Convention on the Elimination of All Forms of Discrimination against Women and mandates that states enact all adequate measures to abolish discrimination. The Sustainable Development Goals, SDG Agenda 2030, and their Goal 5, pursue achieving gender equality and empowering all women and girls (Gender Equality Policy Hub, 2020; Welch et al., 2022).

The UN Gender Equality Strategy (2022-2025) has set an ambitious plan to identify new solutions in achieving gender equality.

In its programmes and activities, the Strategy has included gender-responsive policies that address impediments based on gender, observe gender differences, and develop systems sensitive to gender inequalities (INEE Guidance Note on Gender, 2019).
Different gender equality policies have become an integral part of different UN organizations, states, and associations' strategies:

- The Gender-responsive Policy (INEE Guidance Note on Gender, 2019)
- The Gender Transformative (European Institute for Gender Equality, 2016)
- The Gender Diversity Policy (The Fundamental Principles of the International Red Cross and Red Crescent Movement, 2020)
- The Gender-Neutral Policy (European Institute for Gender Equality, 2017)
- The Gender Inclusive Policy (Thailand Policy Lab, 2023)

Some of the different gender equality polices are listed above.

1.1.2 Promoting Gender Equality in Healthcare

Gender equality policies impact equal access to medical services, healthcare quality, and medical outcomes for men, women, and gender-diverse persons. Promoting gender equality in the healthcare not only ensures equity but also increases the overall efficiency of healthcare system and improves medical outcomes for everyone.

Addressing gender equality issues at the global, EU, and national (Croatia) levels, while focusing primarily on gender equality, some topics also cover broader workforce, digital transformation, and policy issues that affect gender equality outcomes. This distinction highlights the complexity of advancing gender equality in healthcare and ICT (Information and Communication Technology) across multiple levels.

1.1.2.1 Global State of Affairs

World Health Organization (WHO) has adjusted its activities to the Sustainable Development Goals, in particular SDG 3, which focuses on ensuring healthy lives and promoting well-being for all at all ages, and SDG 5, aiming at achieving gender equality and empowering all women and girls. (Welch et al., 2022). Within the Commission on the Status of Women (CSW), the WHO, Global Healthcare Professional Network (GHP Network), and Women in Global Health WGH (2023) presented a report describing social and economic factors (projected stereotyping) due to which there are so few women occupying managerial positions in global healthcare. Global research results indicate that women take up 67% of the global workforce in healthcare and social care, but only 25% of them hold high managerial positions. Over 80% of nurses and more than 90% of midwives perform most of their unpaid care work in homes, families, and communities, and make the majority of decisions regarding purchasing and using healthcare services. The financial value equivalent of women contributing to healthcare systems is estimated at over 3 trillion USD every year.

On the global level, healthcare is mostly led by men; the heads of 69% of healthcare organizations are men, who, at the same time, in 80% of cases act as Chairmen of these organizations' Board of Directors. Such stereotyping significantly contributes to salary differences among genders, with the gender gap amounting to 25%, which is higher than in other sectors. These facts are even more evident considering the estimate that over 40 million new jobs will have become a necessity in healthcare and social security sectors by 2030, and that currently there is already an 18-million-shortage of healthcare workers. Having regard to all of the above, unless target measures are put in place to address the gap, the estimated period for achieving workplace gender equality is 202 years (Women in Global Health, 2023)

WHO plays the key role in the development of standards and guidelines on gender-sensitive healthcare services and supports national initiatives to strengthen the response of the healthcare sector to promote gender equality in the development of the healthcare workforce (World Health Organization, 2020).

1.1.2.2 EU State of Affairs

According to the European Institute for Gender Equality (2022) the EU healthcare sector workforce is dominated by women, and 78% of the EU-28 workers are women. When positions of women and men in healthcare are compared, vertical and horizontal professional segregation can be observed, and women are insufficiently represented in managerial positions.

The European Union aims to ensure equal opportunities for all, regardless of gender, to participate in the digital development closely linked to the quality of healthcare. Although women make up 51% of the EU population, only one in five specialists in the field of information and communication technology is a woman.

According to the Declaration on the Commitment to the Issue of Women in the Digital World (2019), women make up 52% of the European population, yet only 15% hold jobs related to information and communication technology.

In its core, the healthcare personnel crisis is a gender crisis, highlighted Roopa Dhatt, the co-founder of the Women in Global Health organization. The COVID-19 pandemic has additionally emphasized the existing inequalities: women, being 70% of healthcare staff, were more exposed to the virus, had worse access to the PPE, and were faced with a permanent gender wage gap (20% less compared to men). Therefore, gender equality in leadership when it comes to the representation of women would reflect their majority in the healthcare system (European Public Health Alliance, 2023).

Despite the current situation, the EU has achieved significant progress in gender equality in the last few decades through legislation, i.e., gender mainstreaming (International Union against Tuberculosis and Lung Disease, 2020) and special women empowering measures.

Furthermore, the (Gender Equality Strategy 2020-2025, 2020) aims at further development, and has combined gender mainstreaming basic guidelines with targeted actions, emphasized intersectionality, and complies with the EU's gender equality external policies (Gender Equality Strategy, 2020-2025, 2020).

1.1.2.3 Croatian State of Affairs

The results of the last research conducted on gender equality and discrimination in Croatia reflect the respondents' belief that progress has been achieved when it comes to improving gender equality in Croatia.

A strong example of commitment to promoting equal opportunities for all, regardless of gender, was provided by the Faculty of Organization and Informatics, University of Zagreb, in its (Gender Equality Plan for the period 2023–2030, 2023). In line with national and European strategies and the principles of the European Commission, the plan focuses on building a gender-sensitive and inclusive academic environment, increasing the representation of women in academic and teaching positions and in leadership roles, aiming to ensure balanced decision-making and leadership.

In Croatia, on 14th July 2003 Croatian Parliament adopted the Act on Gender Equality (Rodin & Vasiljevic, 2023).

Furthermore, in 2017, the Croatian Government's Office for Gender Equality issued a Manual on Gender Mainstreaming and Promotion of Gender Equality, a manual intended for coordinators in state administrative bodies, dealing with gender mainstreaming and reasons for including the gender perspective as a responsibility of public bodies. (Handbook on Gender Mainstreaming and Promotion gender equality, 2017).

Based on bibliography data and available statistics, the following goals have been identified for this paper.

The global goal is to present gender equality policies at different levels. Specific goals are to present information related to:

- At the global level, in accordance with publicly available information;
- At the European level using the example of the European Union;
- At the national level, in the Republic of Croatia.

The analysis indicators used details from the healthcare sector as an example.

2 Materials and Methods

In order to achieve the goals of this paper, which refer to the collection of secondary statistics on gender equality in healthcare, the used sample included global and national organizations performing healthcare activities.

The analysis included publicly available global data collected by the United Nations (UN), World Health Organization (WHO), European Commission (EC), and the Organization for Economic Cooperation and Development (OECD), publicly available European data from Eurostat, and those available nationally from the Croatian Bureau of Statistics. This sample was selected because it presents data on gender equality in healthcare (WHO), includes economic indicators (OECD), and enables comparisons at the national level (Croatian Bureau of Statistics).

The data, predominantly collected between 2015 and 2023, were analyzed using descriptive statistics, time series analysis, and comparative methods based on secondary sources.

3 Results and Discussion

Statistical gender equality indicators are presented through secondary statistical metrics for organizations operating within the global and European contexts.

3.1 Indicators of Global Gender Equality: UN, WHO, and OECD Data

WHO data (2019) show the percentage of women working in the healthcare sector being 70.3%, which is higher than the overall percentage of women working in all other business activities, i.e., 39.5% (International Labour Organization Geneva ILO Geneva, 2017).

In the OECD member states (OECD, Gender wage gap, GWG, 2023), the percentage of female doctors is 46%.

This is particularly emphasized in the fields of information and communication technology within the healthcare sector. Therefore, projections by the EU have presented that increasing the share of women in this sector to around 45% by 2027 could significantly boost Europe's GDP by as much as 260 to 600 billion euros, making Europe more competitive and prosperous (Women in Digital, European Commission, 2025).

Furthermore, the GWG in various business activities is different and shows higher income for men, defined as the difference between the median earnings of men and women relative to the median earnings of men (as per OECD definition).

In the European Union GWG is 10.8, and, in the OECD, member states the index is 11.4. This index has been slowly increasing in the OECD member states and the European Union. In 2005, the GWG in the EU Member States amounted to 14.1, and in 2022 to 10.8. In the OECD member states, the 2005 GWG was 16.0, and in 2022. 11.3. Data for Croatia are only partially available, and in 2010, the GWG was 3.8, whilst in 2022 it was 3.2 (OECD, Gender wage gap).

According to the World Economic Forum (2016) there is a difference between the number of completed informal and unpaid work hours of actively employed healthcare professionals. An average work day for a man is 7 hours and 47 minutes, and 1 hour and 30 minutes of unpaid informal time. For women, the completed paid hours amount to 8 hours and 39 minutes, and the unpaid informal time is 4 hours and 47 minutes.

3.2 Women and Men in Healthcare Systems of the European Union and Croatia

The example used in the analysis was taken from details available on the professionals actively employed in the healthcare system.

As defined by the Eurostat (2020) the EU healthcare sector employees cover 7% of the overall EU workforce, amounting to 14.3 million. Of the overall healthcare workforce, 78% are women over 22% men. The percentage of women employed in healthcare is slightly lower than the EU average, i.e., Croatia is ranked 11th on the list of lower percentages of women employed in healthcare. These details refer to all healthcare sector employees: doctors, caretakers, nurses, and other healthcare professionals.

According to the Gender Equality Act (2008), gender equality means that women and men are equally

present in all areas of public and private lives, have equal status and equal opportunities to enjoy their rights and benefit from their results.

The analysis of statistical indicators conducted on employees of the healthcare sector in Croatia has shown the following men-women ratio. Different sources explored different periods for results in Fig. 1 (1960-2022) and Fig. 2 (2022).

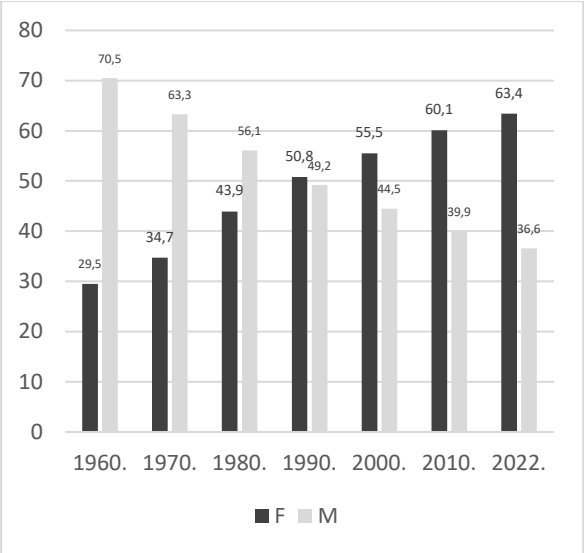


Figure 1. Doctors of medicine in Croatia from 1960 to 2022, the values shown are percentages (Croatian Bureau of Statistics, 2024)

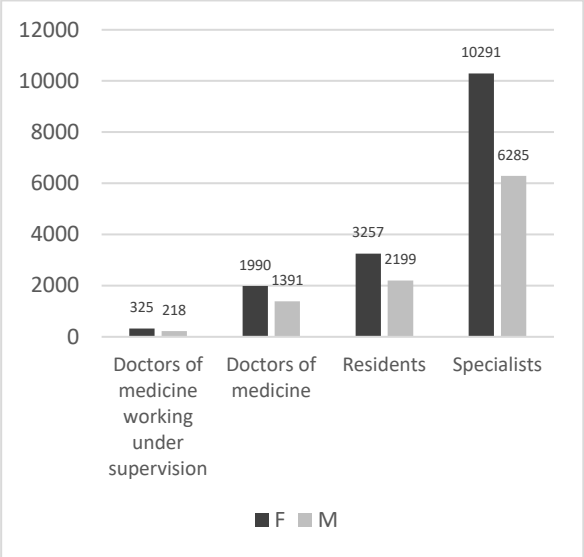


Figure 2. Doctors of medicine employed in the Croatian healthcare system in 2022 (Croatian Bureau of Statistics, 2022)

As shown in Fig. 1, Croatian Bureau of Statistics, 2024, the number of female doctors of medicine has gradually increased throughout the period from 1960 to 2022. In 1960 the ratio of female and male doctors of medicine was 29.5% to 70.5% respectively. In 1990, the percentage of female and male doctors of medicine was almost the same, 50.8% female and 49.2% male. Today, the percentage of female doctors is higher (63.4%) compared to 36.6% of male doctors. This data shows an increasing trend of women graduating from medical schools.

In 2022, the healthcare system employed 63.4% women compared to 36.6% men. There are 69.9% female Doctors of Medicine compared to 30.1% male Doctors of Medicine, 67.5% female Residents over 32.5% male residents. As shown in Fig. 2 Croatian Bureau of Statistics, 2022, the lowest value refers to female Specialists, 61.1%, compared to their male peers, 38.9%, although the area of Specialists also shows a dominant presence of women.

4 Conclusion

The key takeaway of this study is that despite progress, women in the healthcare remain underrepresented in leadership and ICT roles, facing wage gaps, stereotypes, and limited career advancement.

Gender inequality is still a challenge amongst healthcare workforce, at global, EU, and Croatian levels. Women in healthcare commonly occupy lower-ranked, less-paid, and often unpaid positions, without the possibility to make key decisions in this sector.

As stated in the Gender Equality Plan for the period 2023–2030, the Faculty of Organization and Informatics at the University of Zagreb FOI is developing a work culture of equal opportunities for professional development by integrating a gender perspective into educational content through training, workshops, and consultations, promoting the use of inclusive language, and continuously monitoring and evaluating the implementation of planned activities.

Organizations in the healthcare sector can, through additional efforts in the same direction, actively work on removing barriers that may affect the academic and professional advancement of women in the fields of information and communication technology in healthcare.

The focus of research conducted on the global healthcare workforce should be changed. It should prioritize countries with low and medium income, apply a gender and intersectional perspective, and include data classified by gender and gender roles.

Addressing gender inequalities in the healthcare workforce may considerably contribute to achieving Sustainable Development Goals. Letting the development of gender equality take its natural course is not an option. Countries should adopt policies that address fundamental causes of gender inequalities.

Gender equality in the fields of information and communication technology in healthcare is still a very poorly researched topic. Analyses and graphs on the current state of gender equality in healthcare at the global, European, and Croatian level, stated in our paper, refer exclusively to relationships, structure, and number in the profession of doctor of medicine. Concerning the accelerated development and greater

use of information and communication technology in modern healthcare, in correlation with the analysis of data from the literature and information on gender equality in other and sectors of healthcare, it also indicates that in Croatia there is great potential for improving equality in this segment of the healthcare system. The healthcare sector and ICT sector, as an important part of the healthcare system, represent the problem of gender bias and inequality between women and men in healthcare.

The findings of this research underscore the importance for Croatian healthcare professionals and policymakers to adopt gender-responsive and transformative measures that support equal pay, increase women's participation in leadership roles, and strengthen their presence in health-related ICT fields.

Methodology limitations of this paper refer to the fact that the analysis of statistical indicators was focused on organizations that were part of the eligible sample, and that the method of analysis was used only on that eligible sample. Furthermore, different organizations included in the analysis use different methodologies, and data should be interpreted taking into consideration all these facts.

Finally, the analysis of global and national secondary statistical data gives us an insight into the current state of affairs in healthcare. The presented data can inform targeted policy reforms, organizational strategies, and educational initiatives aimed at closing gender gaps, improving workplace equality, and ensuring that the healthcare system in Croatia fully benefits from the potential of its diverse workforce.

The paper provides a valuable descriptive overview by presenting and compiling important data, highlighting opportunities for further development through enhanced ICT efforts aimed at utilizing this information for training and awareness.

The research focusing on gender equality indicators may be continued in other business activities, such as education, social care, and the economy. However, available analyses indicate the need for additional efforts, especially in information and communication technology, to encourage the inclusion of women and better utilize their potential in the healthcare sector.

References

Bargeri, S., Schaap, L.A., Castellini, G., Gianola, S., Innocenti, T., Ostelo, R., Tomaiuolo, R., Vidal-Itriago, A., Rubinstein, S. (2025). Exploring the definitions of gender bias in healthcare literature: a scoping review protocol, MethodsX,15, Retrieved from https://www.sciencedirect.com/science/article/pii/S2215016125003899 Accessed March 15, 2025.

Burušić-Barčan, I., i Burušić, J. (2021). 'Stakleni strop – zastarjeli pojam ili realnost?', ET²eR – ekonomija, turizam, telekomunikacije i računarstvo, 3(3), pp 1-7. Retrieved from: https://hrcak.srce.hr/277681 Accessed May 5, 2025.

Croatian Bureau of Statistics, Žene i Muškarci u Hrvatskoj (2024). Retrieved from: https://ravnopravnost.gov.hr/istaknute-teme-12/zene-i-muskarci-statistika/3453 Accessed May 3, 2025.

Declaration on Commitment to Women in Digital, EU Member States (2019). Retrieved from https://mpudt.gov.hr/UserDocsImages/RDD/SDURDDdokumenti/Deklaracija%20o%20posvecenosti%20pitanju%20zena%20u%20digitalnom%20svijetu_prijevod.pdf Accessed May 5, 2025.

European Commission (2021). She Figures 2021, Gender in Research and Innovation. Luxembourg: Publications Office of the European Union

European Institute for Gender Equality (2022). Gender Equality Index 2022: The COVID-19 pandemic and care. Luxembourg: Publications Office of the European Union.

European Institute for Gender Equality. (2008) Retrieved from: https://eige.europa.eu/publicationsresources/thesaurus/terms/1071?language_content_entity=en Accessed March 15, 2025.

European Institute for Gender Equality. (2016) Retrieved from https://eige.europa.eu/publicationsresources/thesaurus/terms/1090?language_content_entity=hr Accessed March 15, 2025.

European Institute for Gender Equality. (2017) Retrieved from: https://eige.europa.eu/gendermainstreaming/policyareas/health?language_content_entity=en Accessed March 15, 2025.

European Public Health Alliance. (2023) Retrieved from: https://epha.org/the-health-workforce-crisis-is-a-gender-crisis/ Accessed March 20, 2025.

Eurostat (2020). Retrieved from: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20210308-1 Accessed March 25, 2025.

Fierz, W. (2004). Cultivating complex ICT in health care: the gardeners' perspective. Retrieved from https://www.researchgate.net/publication/307545325_Cultivating_information_infrastructure_in_health_care_neither_purely_top-down_nor_purely_bottom-up_but_locally_on_all_scales Accessed March 30, 2025.

Gender Equality Act (Croatian Official Gazette (OG) 82/08, OG 69/17). Retrieved from: https://narodnenovine.nn.hr/search.aspx?upit=Zak

on+o+ravnopravnosti+spolova&naslovi=da&sortiraj=1&kategorija=1&rpp=10&qtype=3&pretraga=da Accessed March 20, 2025.

Gender Equality Plan for the period 2023–2030, Faculty of Organization and Informatics, University of Zagreb FOI (2023). Retrieved from: https://www.foi.unizg.hr/sites/default/files/plan_rodne_ravnopravnosti_foi_2023_bez_priloga.pdf Accessed March 22, 2025.

Gender Equality Policy Hub (2020). Retrieved from: https://www.parliament.uk/mps-lords-and-offices/offices/commons/scrutinyunit/gender-equality-policy-map/ Accessed March 30, 2025.

Gender Equality Strategy 2020-2025. (2020) Retrieved from: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-equality-strategy_en Accessed March 28, 2025.

Gender Equality Strategy UN 2022-2025. (2020) Retrieved from: https://genderequalitystrategy.undp.org/ Accessed April 15, 2025.

Handbook on Gender Mainstreaming and Promotion gender equality (2017). Zagreb

INEE. (2019). INEE Guidance Note on Gender. Retrieved from: https://inee.org/eie-glossary/gender-responsive Accessed June 15, 2025.

International Labour Organization, Geneva (2017). Improving employment and working conditions in health services; In World Health Organization (2019). Delivered by women, led by men: A gender and equity analysis of the global health and social workforce. (Human Resources for Health Observer Series No. 24). Retrieved from: https://cdn.who.int/media/docs/default-source/health-workforce/delivered-by-women-led-by-men.pdf?sfvrsn=94be9959_2 Accessed February 15, 2025.

International Union against Tuberculosis and Lung Disease. (2020) Retrieved from: http://theunion.org/sites/default/files/2020-09/The-Union_Gender-Policy.pdf Accessed February 15, 2025.

Nedović, M.; Ivanković, D.; Mišćević, D.: Stakleni strop - položaj žena u sustavu znanosti, Obrazovanje za poduzetništvo, 5 (1), 2015, p. 91.

Nidogon Višnjić, S., Begičević Ređep, N., i Vidaček-Hainš, V. (2018). Stavovi i percepcije žena na rukovodećim pozicijama o njihovom položaju na radnom mjestu, Ekonomska misao i praksa, 27(1), pp 287-313. Retrieved from: https://hrcak.srce.hr/202285 Accessed March 15, 2025.

OECD (2018) Bridging the Digital Gender Divide: Include, Upskill, Innovate. Paris: OECD Publishing. Retrieved from https://www.researchgate.net/publication/329144075_BRIDGING_THE_DIGITAL_GENDER_DIVIDE_INCLUDE_UPSKILL_INNOVATE_-_Key_findings Accessed March 20, 2025.

OECD (2023) Gender Wage Gap – definition. Retrieved from: https://www.oecd.org/en/data/indicators/gender-wage-gap.html Accessed March 11, 2025.

OECD (2023) Gender wage gap; Measure: Combined unit of measure: Percentage of wages of men in the same decile. Retrieved from: https://dataexplorer.oecd.org/vis?tm=wage&pg=0&snb=10&vw=tb&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_EARNINGS%40GENDER_WAGE_GAP&df[ag]=OECD.ELS.SAE&df[vs]=1.0&pd=2005%2C&dq=MEDIAN.T&to[TIME_PERIOD]=false Accessed March 25, 2025.

OECD (2023) Gender wage gap. Retrieved from: https://www.oecd.org/en/topics/policy-issues/gender-equality.html?oecdcontrol-90d4bb70b2-var2=DNK%7CTUR Accessed March 5, 2025.

Official European Union websites. Retrieved from: https://www.eurofound.europa.eu/hr/topic/rodna-ravnopravnost Accessed May 15, 2025.

Pološki Vokić, N., Sinčić Ćorić, D., i Obadić, A. (2017). 'Biti ili ne biti žena? Percepcija visokoobrazovanih žena o rodnoj jednakosti na radnom mjestu', Revija za socijalnu politiku, 24(3), pp 276-276. Retrieved from: https://doi.org/10.3935/rsp.v24i3.1432 Accessed April 15, 2025.

Rodin, Siniša; Vasiljević, Snježana Zakon o ravnopravnosti spolova, Revija za socijalnu politiku, 10 (2003), 3; 397-402. Retrieved from: doi: 10.3935/rsp. v10 I 3.118 Accessed March 1, 2025.

Shannon, G., Jansen, M., Williams, K., Cáceres, C., Motta, A., Odhiambo, A., Mannell, J. (2019). Gender equality in science, medicine, and global health: Where are we at and why does it matter? The Lancet, 393(10171), pp. 560–569.

Thailand Policy Lab. (2023). Retrieved from: https://thailandpolicylab.com/en/checklist-gender-inclusive-workplace/ Accessed February 15, 2025.

The Fundamental Principles of the International Red Cross and Red Crescent Movement. (2020) Retrieved from: https://www.ifrc.org/sites/default/files/GD-Policy-v4.pdf Accessed March 10, 2025.

UNESCO (2017) Cracking the code: Girls' and women's education in STEM. Paris: UNESCO. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000262730 Accessed June 15, 2025.

United Nations, Gender Equality Policy. Retrieved from https://www.sciencedirect.com/topics/social-sciences/gender-policy Accessed March 25, 2025.

Welch V, Dewidar O, Tanjong Ghogomu E., Abdisalam S., Al Ameer A, Barbeau V.I., Brand K., Kebedom K., Benkhalti M., Kristjansson E., Madani M.T., Antequera Martín A.M., Mathew, C.M, McGowan J., McLeod W., Park, H.A., Petkovic, J., Riddle A., Tugwell, P., Petticrew, M., Trawin, J., Wells, G.A. (2022). How effects on health equity are assessed in systematic reviews of interventions., The Cochrane Database of Systematic Reviews, 18; 1(1). Retrieved from: https://pubmed.ncbi.nlm.nih.gov/35040487/#full-view-affiliation-8 Accessed May 25, 2025.

Walby, S. (2004). The Cost of Domestic Violence. London: Women and Equality Unit (DTI). Retrieved from https://openaccess.city.ac.uk/id/eprint/21681/Accessed May 1, 2025.

Women in Digital, European Commission (2025). Retrieved from: https://digitalstrategy.ec.europa.eu/en/policies/women-digital. Accessed May 10, 2025.

Women in Global Health. Policy Brief: The State of Women and Leadership in Global Health: The XX Paradox. Women in Global Health; 2023. Retrieved from https://womeningh.org/sheshapes/ Accessed April 21, 2025.

World Economic Forum (2016) p 16. According to the World Health Organization (2019). Delivered by women, led by men: A gender and equity analysis of the global health and social workforce. (Human Resources for Health Observer Series No. 24) Retrieved from https://www.weforum.org/meetings/world-economic-forum-annual-meeting-2016/ Accessed March 15, 2025.

World Health Organization (2025). Retrieved from: https://www.who.int/healthtopics/gender#tab=tab_2 Accessed May 11, 2025.

Developing an LLM-specific Coding Framework for Prediction and Feedback Modeling of Science Teachers' Instructional Expertise

Jin Eun Yoo
Korea National University of Education
Department of Education
250 Taeseongtabyeon-ro, Cheongju, Republic of Korea
jeyoo@knue.ac.kr

Nam-Hwa Kang
Korea National University of Education
Department of Physics Education
250 Taeseongtabyeon-ro, Cheongju, Republic of Korea
nama.kang@knue.ac.kr

Suna Ryu
Korea National University of Education
Department of Chemistry Education
250 Taeseongtabyeon-ro, Cheongju, Republic of Korea
sunaryu@knue.ac.kr

Jun-ki Lee
Jeonbuk National University
Institute of Science Education
567 Baekje-daero, Jeonju, Republic of Korea.
junki@jbnu.ac.kr

Youngsun Kwak
Korea National University of Education
Department of Earth Science Education
250 Taeseongtabyeon-ro, Cheongju, Republic of Korea
kwak@knue.ac.kr

Hyeong Gwan Kim
Korea National University of Education
Department of Education
250 Taeseongtabyeon-ro, Cheongju, Republic of Korea
milphy@nate.com

Abstract. *This study introduces a novel coding framework for classroom discourse tailored to large language models (LLMs). To address limitations of prior studies relying on global-level observation tools, this framework adopts utterance-level analysis and fine-grained coding. Predicate-level chunking, coding, and rating were collaboratively conducted using a dual-expertise approach, involving teams of science education faculty and experienced teachers, across 125 science sessions from 27 teachers. Over the course of the process, a glossary of terms was also developed via a primarily bottom-up approach. Quadratic Weighted Kappa results demonstrated reasonable rater consistency, supporting the validity of the framework.*

Keywords. Large Language Models (LLMs), classroom discourse, science education, instructional analysis, utterance-level coding, coding framework, rater consistency, teacher expertise, science teacher

1 Introduction

Teachers play a pivotal role in education. It is widely acknowledged that 'the quality of education cannot exceed the quality of its teachers,' underscoring how teachers' instructional and assessment practices influence student achievement. In particular, teacher language has long functioned as a primary tool for instructional delivery in classroom discourse. For example, the specific lexical choices, the timing of verbal interventions, and the form of feedback provided can have a decisive impact on students' learning. Classroom discourse studies are grounded in this perspective. Experts in traditional studies typically evaluate one or more complete lessons—either through direct observation or via recorded sessions—using established observation frameworks. While such approaches allow for a holistic evaluation of instruction, they heavily rely on subjective judgment, lack scalability, and tend to concentrate on exemplary teachers—thus limiting their applicability to the

practical needs of teachers with diverse backgrounds and varying levels of experience.

Recent advances in large language models (LLMs) have opened new possibilities for analyzing and modeling classroom discourse in ways that were not feasible with traditional observation-based approaches. While LLM-based studies are increasingly applied to classroom discourse analysis, they continue to rely on structured coding schemes such as CLASS (Classroom Assessment Scoring System) or MQI (Mathematical Quality of Instruction)—widely used in traditional observation research. These frameworks were designed for global-level lesson analysis, not for the local, utterance-level applications. This misalignment presents a critical challenge to the proper granularity required for effective LLM implementation.

To address this issue, the present study aims to develop a coding scheme tailored to LLM analytical units in modeling teacher expertise prediction and feedback. Specifically, this study integrates authentic classroom discourse from in-service secondary science teachers with state-of-the-art deep learning techniques and instructional analysis methods to construct both theoretical and empirical models for predicting teacher expertise and generating formative feedback. This manuscript reports year-one findings of a three-year project, focusing on the development of the coding scheme for LLM modeling.

Conventionally, raters are responsible for assigning scores or ratings—such as on a 1-to-5 scale—whereas coders categorize or classify data according to predefined themes or categories. In the present study, the act of chunking teacher utterances based on semantic coherence is defined as coding, while assigning a numerical score to each chunked segment is regarded as rating. As the same individuals performed both tasks in sequence, the terms ‘rater’ and ‘coder’ are used interchangeably throughout this study.

2 Literature Review

A number of LLM-based studies on classroom discourse have utilized transcript data provided by the National Center for Teacher Effectiveness (NCTE). The NCTE dataset consists of classroom observation records collected and annotated between 2010 and 2013 from four school districts in the New England region. It captures mathematics instruction delivered by 317 elementary teachers to 4th and 5th grade students, most of whom came from low-income and historically marginalized backgrounds (Kane et al., 2015).

A distinguishing feature of the dataset is its inclusion of 1,660 annotated transcripts—each corresponding to a 45- to 60-minute instructional session—evaluated by expert raters using two validated classroom observation protocols: the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) and the Mathematical Quality of Instruction (MQI; Hill et al., 2008). Several prior studies, including Alic et al. (2022), Wang and Demszky (2023), Xu et al. (2024), and Hardy (2025), have employed the built-in annotated data from NCTE. Accordingly, a brief overview of the two rating instruments—CLASS and MQI—is presented below.

2.1 CLASS and MQI

The CLASS is an observational tool designed for evaluating instructional quality in K–12 classrooms. It involves segmenting a lesson into observation cycles of 15 to 20 minutes, typically conducted over two to six cycles per session. Each cycle is scored on a 7-point scale (ranging from 1 to 7) across 10 to 11 items, organized under three core domains: emotional support, classroom organization, and instructional support. CLASS emphasizes the quality of teacher–student interactions and intentionally excludes curriculum-specific content and the physical classroom environment from its evaluation criteria.

Similarly, the MQI is a structured classroom observation protocol specifically designed to assess the mathematical rigor and interactive quality of instruction in K–12 mathematics settings. A complete mathematics lesson is divided into segments of approximately 5 to 7 minutes, which are individually rated and then aggregated to yield a composite session score. The MQI rubric comprises five dimensions: richness of the mathematics, errors and imprecision, working with students and mathematics, student participation in meaning-making and reasoning, and connections between classroom work and broader mathematical ideas. Each dimension is rated using a three-level ordinal scale (low, mid, high).

2.2 Studies Using Built-in NCTE Ratings without Additional Human Raters

Among previous studies, Xu et al. (2024) and Hardy (2025) leveraged the built-in annotations provided by the NCTE dataset, without recruiting additional human raters. Xu et al. (2024) applied Transformer-based models—including BERT, DistilBERT, XL-Net, RoBERTa, and Llama2-7B with QLoRA—to 9,886 segments of MQI transcripts, each approximately 7.5 minutes in length. Based on performance metrics such as F1 score, majority F1, and Spearman’s rho, the results indicated that LLM-based models performed comparably to human raters on tasks requiring lower levels of pedagogical reasoning. However, their performance declined significantly in evaluating more complex teaching practices. Notably, the study also found that using only teacher utterances as input yielded results similar to those obtained when both teacher and student utterances were included.

Hardy (2025), on the other hand, conducted a secondary analysis focusing on four of the thirteen MQI items. The study compared the rating performance of GPT-based models and encoder-based models using evaluation criteria such as concordance,

confidence, validity, bias, fairness, and helpfulness. Analytical methods included assessments of rater consistency, generalizability and decision studies, and hierarchical rater modeling. Overall, encoder-based models outperformed GPT-based models and achieved performance levels comparable to those of human raters.

2.3 Studies Involving Additional Human Raters and Calibration Procedures

In some studies utilizing the NCTE dataset, researchers additionally recruited human raters and implemented rater training procedures, depending on the research objectives. For instance, Alic et al. (2022) focused on identifying teachers' use of *focusing* and *funneling* questions and demonstrated that a supervised RoBERTa model achieved a strong linear correlation of 0.76 with expert-coded labels. To construct the dataset, they applied three sampling criteria—mathematical relevance, follow-up on a prior student utterance, and inclusion of a question—to select 2,348 examples of teacher–student exchanges. These were then annotated using three MQI items (student-provided explanations; overall student participation and meaning-making and reasoning; and mathematical quality of instruction), with a three-category coding scheme (0 = not meeting criteria; 1 = funneling; 2 = focusing). Thirteen raters underwent structured rater training and calibration sessions, resulting in a Fleiss' Kappa of 0.415 on the randomly assigned utterance pairs.

In comparison, Wang and Demszky (2023) used the built-in annotations provided in the NCTE dataset but recruited two additional human raters to compare against ChatGPT-generated responses. Specifically, they randomly selected ten transcripts each for the CLASS and MQI instruments from the total of 1,660 transcripts. The utterances within each segment were grouped into equal-sized bins. Two mathematics teachers were then tasked with evaluating ChatGPT's zero-shot performance on a range of teacher coding tasks. Depending on the task, three or four evaluation criteria were used on a 3-point scale (yes, somewhat, no). However, the study did not provide detailed information regarding the training of these raters or how discrepancies in the actual ratings were addressed. According to Spearman's rho analyses, ChatGPT performed comparably to human raters on relatively simple instructional tasks but failed to capture more complex teaching practices effectively.

2.4 Rationale of the Study

Traditional classroom analysis studies have typically involved the collection of instructional sessions conducted by confident teachers in controlled, evaluative settings. These studies generally rely on structured observation frameworks and domain experts who analyze one or more sessions using extended time intervals—such as full-class sessions or 15-minute blocks—as the primary unit of analysis. In contrast, LLM-based studies, including the present one, analyze classroom discourse in the form of text data, often focusing on teacher utterances.

Despite their strengths in processing large-scale language data, prior research indicates that LLMs perform well on relatively simple, low-inference tasks but fall short of human-level performance on more cognitively demanding or pedagogically nuanced tasks (Wang & Demszky, 2023; Xu et al., 2024). These limitations may stem from two critical challenges: the absence of an LLM-specific evaluation framework and the misalignment between traditional units of analysis and those suitable for LLM-based approaches.

To enhance the effectiveness of LLMs in classroom discourse analysis, it is therefore necessary to shift the unit of analysis toward smaller, utterance-level segments and to develop evaluation frameworks tailored to the characteristics of LLMs. However, to date, no study has proposed a coding framework explicitly designed for LLM-based approaches. Most existing studies remain secondary analyses of data originally collected under traditional observation paradigms, typically applying global-level frameworks such as CLASS and MQI. Even in cases involving newly collected data, no novel coding schemes have been introduced that align with the analytical granularity required for effective LLM implementation.

Due to the fundamental difference in analytical approach, it was not feasible to directly apply existing science education observation frameworks. Therefore, it became essential to develop and implement a coding scheme tailored to LLM-based modeling. To conclude, a coding framework for full-class instructional sessions should be developed, incorporating coding schemes for smaller, semantically meaningful segments to enable more effective implementation of LLM-based analysis. Table 1 presents the comparison between the present study and traditional classroom discourse research.

Table 1. Comparison between the Present Study and Traditional Research

	Present Study	Traditional Research
Data	classroom transcripts composed of teacher utterances (text data)	audio-visual data, typically from direct observation
Data collection	authentic teaching practices	often collected from controlled settings
Unit of analysis	utterance-level chunked segments	full-class session or minute blocks; research question-specific

		chunking (e.g., teacher-student utterance pairs)
Coding framework	no clearly established or widely agreed-upon coding framework	structured frameworks (e.g., CLASS, MQI)
Purpose	prediction and feedback modeling of teachers' instructional expertise	evaluation and feedback on the given lesson

3 Methods

3.1 Materials

A total of 125 classroom sessions were collected from 27 middle school science teachers during the second semester of 2024 and the first semester of 2025. In South Korea, middle school science teachers typically major in one of the four subjects—physics, chemistry, life science, or earth science—but are required to teach all areas. Each participating teacher recorded at least three instructional sessions, each approximately 45 minutes long, in both their major and non-major subject areas.

Table 2 presents the distribution of instructional sessions by teachers' majors and the subjects they taught. The table details the number of participating teachers for each major and breaks down the total number of sessions they conducted across four subject areas: Physics, Chemistry, Life Science, and Earth Science. Percentages indicate the proportion of sessions within each major that were devoted to each subject. Notably, teachers frequently taught subjects outside their major fields, reflecting the interdisciplinary nature of middle school science instruction in the given context.

Table 2. Number of Teachers and Instructional Sessions by Academic Major and Subject Area

Major	# of Teachers (%)	Subjects Taught	Sessions (%)
Physics	10 (37.01%)	Physics	12 (29.27)
		Chemistry	9 (21.95)
		Life Science	14 (34.15)
		Earth Science	6 (14.63)
		Subtotal	41(100%)
Chemistry	4 (14.81%)	Physics	4 (20.00)
		Chemistry	8 (40.00)
		Life Science	6 (30.00)
		Earth Science	2 (10.00)
		Subtotal	20(100%)
Life Science	8 (29.63%)	Physics	6 (16.22)
		Chemistry	14 (37.84)
		Life Science	14 (37.84)
		Earth Science	3 (8.11)
		Subtotal	37(100%)
Earth Science	5 (18.52%)	Physics	9 (33.33)
		Chemistry	3 (11.11)
		Life Science	3 (11.11)
		Earth Science	12 (44.44)
		Subtotal	27(100%)
Total	27 (100%)		125

The collected audio data were transcribed into text data using a speech-to-text (STT) software, and each participating teacher manually corrected their transcript. For each subject area, experienced teachers with acknowledged instructional expertise were nominated to serve as coders. Eight teacher-raters participated in the study, with two each holding undergraduate degrees in physics, chemistry, life science, and earth science education. Their teaching experience ranged from over 6 to 22 years, and all were either enrolled in or had completed a master's or doctoral degree in science education. Several had additional experience in national assessment development, teacher training, or instructional material review, demonstrating both subject-matter expertise and familiarity with evaluation practices. Of note, the teacher-coders contributed solely to the coding process and were not among those who provided instructional recordings.

Based on the transcribed data, paired teams of university faculty and coders—each representing the four subjects of science—collaboratively constructed instructional data. The instructional dataset consisted of four components: *input*, *theme*, *output*, and *rating*. For each meaningful teacher utterance selected from the corrected transcript, the utterance was entered into the *input* column, with its corresponding thematic category and qualitative evaluation recorded in the *theme* and *output* columns, respectively. A numerical rating on a scale of 1 to 5 was also assigned in the *rating* column (1: educationally inappropriate utterance; 2: utterance requiring improvement; 3: average-level utterance; 4: educationally effective utterance; 5: educationally exemplary utterance).

This study retained only teacher utterances in the transcripts, excluding all student speech. The inclusion of student utterances in classroom recordings entails logistical complexities, such as the deployment of multiple microphones, and introduces additional ethical considerations, including more stringent IRB procedures due to the involvement of minors. Notably,

Xu et al. (2024) demonstrated that, even for student-related variables, using teacher utterances alone as input for large language models produced results that were comparable to—or for some MQI variables exceeded—those obtained when student speech was also included.

3.2 Establishing a Robust Coding Framework

3.2.1 Chunking Units

In studies involving relatively short texts—such as essay-type prompt-response pairs or teacher–student question–answer exchanges—the chunking unit is generally well-defined. However, when analyzing teacher discourse over the course of a 45-minute instructional session, determining appropriate analytical units for LLM-based analysis presents a more nuanced and challenging task.

Considering data-processing efficiency and computational cost depending on chunking units in LLM-based modeling, two distinct chunking approaches were initially applied: small-unit chunking, which involved detailed coding at the sentence level, and large-unit chunking, which involved chunking at the paragraph level. Comparison of the two chunking methods was also one of the research questions. However, coders reported substantial difficulty in consistently distinguishing between the two chunking units during the actual coding process. Consequently, considerable discrepancies were observed across subject areas, prompting the revision of the initial coding scheme.

A review of the recorded instructional sessions further revealed notable variation in teachers' speech patterns. While some teachers spoke in long, complex utterances, others used brief sentences, phrases, or even single words. The STT software segmented the audio based on pause detection, which led to substantial variability across sessions and speakers—some segments appeared as full paragraphs, whereas others consisted of only one or two words.

To ensure consistency and support fine-grained feedback generation appropriate for LLM-based modeling, chunking was conducted at the predicate level wherever possible. In instances where teachers produced extended utterances without natural pauses, manual chunking was applied based on semantic coherence. Moreover, utterances that had already been segmented into separate lines by the software were not merged during the coding process.

3.2.2 The Need for Systematic Coding and Rating

A commonly noted limitation of LLM-based models is their tendency to generate overly generic feedback—functionally akin to returning the median in statistical terms. Previous studies have reported that response generated by these models are often characterized as 'not novel or insightful' (e.g., Wang & Demszky, 2023). In recognition of this limitation, the present study allowed coders the flexibility to develop and assign themes during the coding process. Nonetheless, pronounced differences emerged in the descriptive statistics of the number of themes and the distribution of ratings across subject areas.

The discrepancies observed in chunking units and theme development during the initial coding process highlighted the necessity of systematic coding and rating and the development of a standardized glossary of terms. To ensure the validity of subsequent analyses, two raters were assigned per subject area, and a targeted, systematic coding and rating approach was implemented. The specific procedures are detailed in the following section.

3.3 Coding and Rating Procedures

3.3.1 Glossary of Terms Development: A Primarily Bottom-Up Approach

The development of the glossary followed a primarily bottom-up approach, albeit subtly informed by theoretical background and prior disciplinary training. The initial coding yielded 172 terms in physics, 71 in chemistry, 63 in life science, and 373 in earth science. When organized in alphabetical order, the total number of terms reached 678. However, many redundancies were identified due to variations in particle usage (e.g., checking students' level of understanding vs. checking student understanding) as well as semantically identical (e.g., self-directed learning vs. autonomous learning) or closely related expressions (e.g., task presentation vs. task guidance). Through collaborative and iterative refinement rounds involving four science education professors—each representing one of the four subject areas—and one professor specializing in educational evaluation, the initial glossary was consolidated into 223 terms. Each science education professor then developed exemplary annotation responses based on the initial glossary, which were used in multiple rounds of coding and rating. Feedback from these sessions led to further refinements of both the glossary and the coding scheme.

Finally, the revised glossary comprises 144 finalized themes, each accompanied by a simple definition, a list of thesaurus terms, and a higher-order category. These categories specify the instructional phase and functional role associated with each theme and provide corresponding rating guidelines. For example, themes under the category 'Basic Instructional Management' (e.g., lesson introduction, closure, review of prior learning) are capped at a maximum rating of 3 points, whereas more pedagogically important themes such as 'inducing cognitive conflict for conceptual change' may warrant ratings of 4 or 5 points.

3.3.2 Coding and Rating: A Dual-Expertise Approach

The academic expertise of science education faculty is a sine qua non in the evaluation of science teachers' instructional and assessment practices. This need is particularly salient in the present study, which developed and implemented a novel *LLM-specific instructional expertise coding framework*. Equally critical, however, is the experiential knowledge of experienced science teachers who continuously engage in, and reflect on, instructional and assessment practices within authentic classroom settings. The study thus recognizes the complementary and essential roles of both theoretical expertise in science education and the practical expertise of in-service science teachers.

To operationalize this dual-expertise approach, coding and rating were conducted by a team composed of one science education faculty member and two experienced science teachers per subject, with each member's disciplinary and practical knowledge fully respected and leveraged. The feedback obtained throughout this process informed the refinement of the thematic framework and coding manual, including the glossary of terms, thereby enhancing the consistency of the ratings and contributing the validity of the approach.

The process proceeded in several iterative stages. In Step 1, a faculty member with expertise in educational evaluation provided a comprehensive set of guidelines outlining the procedures for coding and rating, including protocols for constructing the instructional data set.

In Step 2, each of the four science education faculty members independently selected one instructional session within their respective major subject areas (45 minutes in length) and produced subject-specific annotations. The two teacher-coders then independently coded the same session. Their results were iteratively reviewed through collaborative discussion to reach consensus within the team. Discrepancies were of particular concern, and any related misconceptions or inappropriate coding decisions were carefully examined.

Step 3 involved full-faculty meetings. Issues arising during the Step 2—including challenges, cautions, and suggestions—were addressed, and annotations on common factors (e.g., Basic Instructional Management) were collaboratively generated to ensure consistency across subjects.

In Step 4, the common-factor annotations derived in Step 3 were fed back to the respective teams. Incorporating the feedback, each team generated subject-specific annotations within their domain, producing a finalized annotation set for one full instructional session, known as the consensus version. The procedures and efforts used to generate this finalized set were subsequently applied to the annotation of the remaining sessions. Fig. 1 illustrates the procedures.

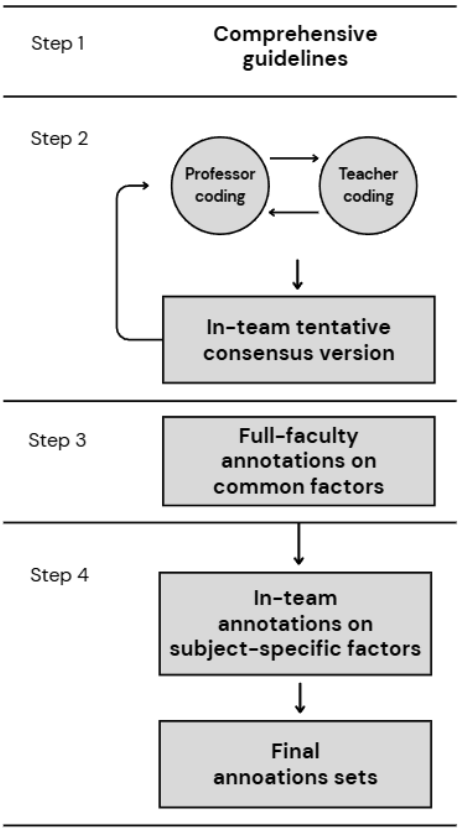


Figure 1. Coding and Rating Procedures

3.4 Analysis Methods

This study employed Quadratic Weighted Kappa (QWK) to assess rater consistency. QWK is an extension of Cohen's Kappa that adjusts for agreement occurring by chance. Unlike Cohen's Kappa, which assigns equal weight to all disagreements, QWK applies quadratic weights to discrepancies, such that larger differences between ratings incur greater penalties.

Equation 1 expresses the unweighted form of Cohen's Kappa, which is calculated as one minus the ratio between q_o and q_c (Cohen, 1968). q_o is the proportion of observed disagreement ($p_o = 1 - q_o$), and q_c is the proportion of disagreement by chance ($p_c = 1 - q_c$).

$$\kappa = 1 - \frac{q_o}{q_c}. \quad (1)$$

Suppose item i is rated by rater 1(j) and rater 2(k) on a K-ordinal scale. The weighted Kappa incorporates weights to the unweighted Kappa as in equation 2 (Vanbelle, 2016). The sole distinction between the equations is whether the disagreement rates are weighted. Equation 2 provides a general formulation of the weighted Kappa.

$$\kappa^w = 1 - \frac{q_o^w}{q_c^w}, \quad (2)$$

where $q_o^w = \sum_{j=1}^{K}\sum_{k=1}^{K} v_{jk} \cdot p_{jk}$ and $q_c^w = \sum_{j=1}^{K}\sum_{k=1}^{K} v_{jk} \cdot p_{j\cdot} \cdot p_{\cdot k}$.

When the quadratic weights defined in equation 3 are applied to equation 2, the resulting coefficient is referred to as the Quadratic Weighted Kappa (QWK).

$$v_{jk} = \left(\frac{j-k}{K-1}\right)^2, \quad (3)$$

Where $0 \le v_{jk} \le 1,\ v_{jj} = v_{kk} = 0\ (j,k = 1,\dots,K)$.

Although Spearman's rank-order correlation (rho) is also frequently used to examine rater consistency in LLM research, it measures the monotonic relationship between raters by converting ratings into ranks. In contrast, QWK directly evaluates the degree of actual agreement by preserving the original rating values. Therefore, in contexts such as the present study—where tied scores are prevalent—Spearman's rho is not among the most suitable statistics, as it neither adjusts for chance agreement nor performs robustly in the presence of numerous ties.

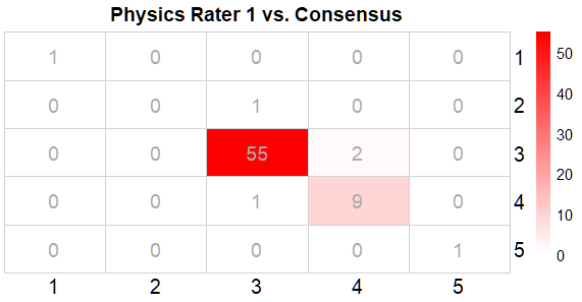
QWK ranges from –1 to 1, with values closer to 1 indicating stronger agreement between raters, and a value of 0 suggesting agreement no better than chance. The QWK results were calculated for every iteration, and shared with the coding team, which contributed to improving rater consistency. Specifically, the consensus rating from the final round was regarded as the ground truth. Accordingly, the QWK between the consensus and each individual rater in the final round was calculated and presented, accompanied by heatmaps visualizing the alignment between each rater and the consensus. The R packages *Metrics* and *pheatmap* were used to compute QWK and to generate heatmaps, respectively.

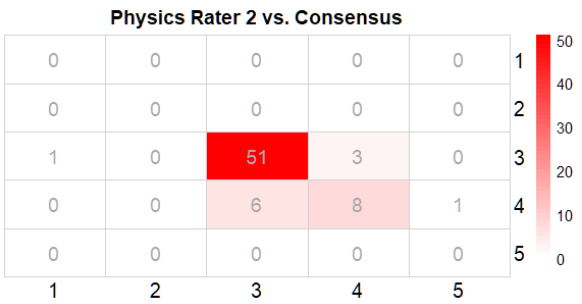
4 Results

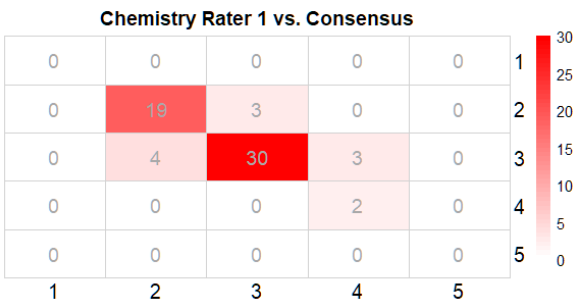
After a full iteration of Steps 1 through 4 in Fig. 1, the within-subject rater consistency results were summarized in Table 3. Specifically, the QWK between each rater and the consensus rating was calculated. Based on the interpretation criteria proposed by Landis and Koch (1977), all four subjects exhibited at least moderate agreement, with some reaching levels categorized as substantial or even almost perfect. These outcomes are further illustrated through heatmaps (Fig. 2).

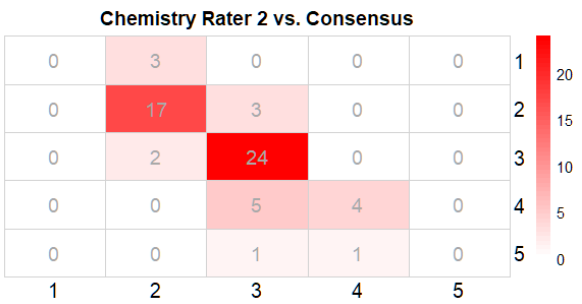
Table 3. QWK between Raters and Consensus by Subject

Subject	Rater 1 vs. Consensus	Rater 2 vs. Consensus
Physics	0.89	0.52
Chemistry	0.75	0.74
Life Science	0.41	0.69
Earth Science	0.96	0.98









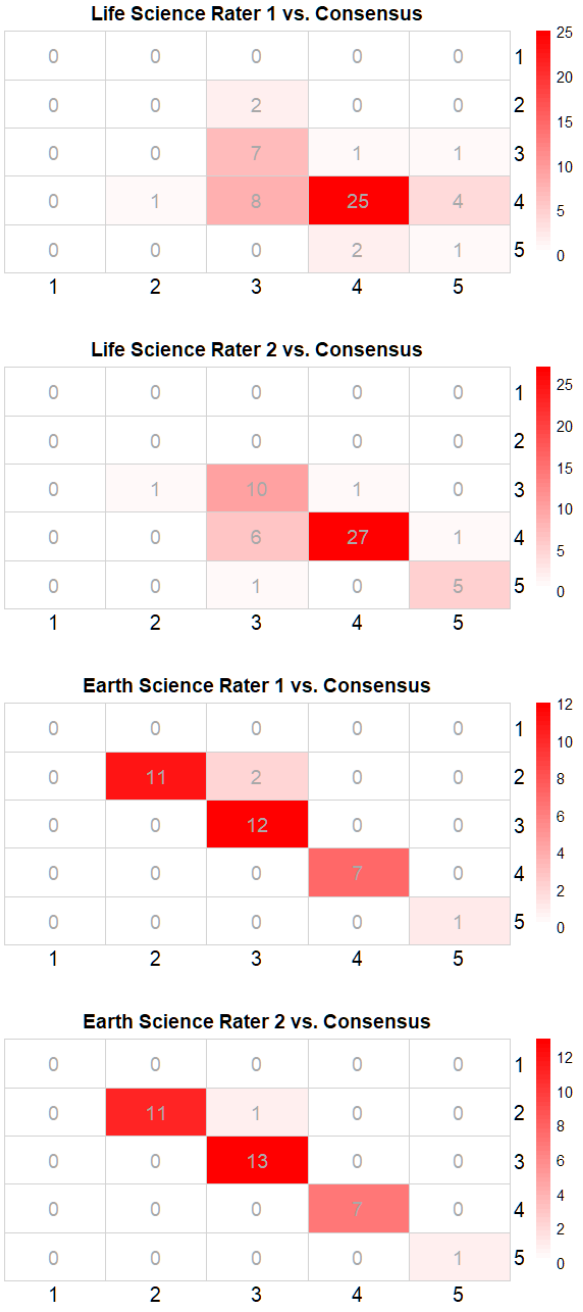


Figure 2. Rater–Consensus Agreement Heatmaps by Subject

5 Discussion

In response to the limitations identified in previous LLM research, this study designed and implemented predicate-level chunking, coding, and rating wherever possible. However, this fine-grained approach introduced a critical challenge: such segmenting could at times obscure the overarching instructional intent. For instance, when explanatory discourse is structured through rhetorical questioning followed by self-answering, the isolated coding of teacher utterances—coupled with the absence of student responses—can lead to mislabeling. To address this issue, coders engaged in repeated, holistic reviews of entire lessons to contextually infer the teacher's pedagogical intent prior to conducting coding and rating. Continuous refinement of the coding scheme and iterative updates to the coding guidelines were integral to the methodological rigor of this study.

Moreover, to elevate moderate levels of rater consistency observed in some subject areas to substantial or near-perfect agreement, additional coder training appears warranted. In this study, inter-rater consistency was calculated at the session level using rater pairs within each subject. Future research will involve re-running the cross-subject calibration process for shared coding factors (corresponding to Step 3 in Fig. 1) and subsequently evaluating consistency across all eight raters. This will support a more nuanced and systematic refinement of the coding and rating protocol.

To date, LLM research on modeling teacher instructional expertise has predominantly relied on traditional, global evaluation frameworks such as CLASS and MQI, resulting in alignment limitations for fine-grained analysis. To address this gap, we introduce a predicate-level, utterance-centric coding framework developed through a dual-expertise approach that combines science education faculty with experienced teachers. Specifically, teacher utterances are chunked into smaller units of analysis, and the chunked segments are coded and rated according to instructional expertise glossary, which were developed as a part of the process via a primarily bottom-up approach. The Quadratic Weighted Kappa results demonstrate reasonable rater consistency.

This study lays the groundwork for domain-adaptive LLM research. Future work will integrate the coding outputs into LLM pipelines for quality feedback modelling, leveraging and benchmarking latest techniques such as data augmentation and prompt engineering to identify optimal strategies for teacher expertise modeling. While the present framework was developed in the context of science education, exploring its adaptation to other subjects (e.g., social sciences, language, and mathematics courses) would be highly worthwhile. Depending on the specific subject domain, some areas may lend themselves more readily to LLM modeling, whereas others may require the integration of additional factors to improve prediction. Extending the framework to such areas could provide critical insights into its flexibility and broader applicability. Ultimately, this line of research aspires to advance both the methodological rigor and practical impact of LLM-based educational analytics, fostering more effective, data-driven support for teaching and learning.

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2024S1A5C3A0104264212).

References

Alic, S., Demszky, D., Mancenido, Z., Liu, J., Hill, H., & Jurafsky, D. (2022). Computationally identifying funneling and focusing questions in classroom discourse. *arXiv.* https://doi.org/10.48550/arXiv.2208.04715

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213–220.

Hardy, M. (2025). "All that glitters": Techniques for evaluations with unreliable model and human annotations. *Findings of the Association for Computational Linguistics: NAACL 2025*, 2250–2278.

Hill, H., Blunk, M., Charalambous, C., Lewis, J., Phelps, G., Sleep, L., & Ball, D. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430-511.

Kane, T., Hill, H., & Staiger, D. (2015). National center for teacher effectiveness main study. *Inter-university Consortium for Political and Social Research*. icpsr36095-v2.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Vanbelle S. (2016). A new interpretation of the weighted kappa coefficients. *Psychometrika, 81*(2), 399-410.

Wang, R., & Demszky, D. (2023). Is ChatGPT a good teacher coach? Measuring zero-shot performance for scoring and providing actionable insights on classroom instruction. *arXiv.* https://doi.org/10.48550/arXiv.2306.03090

Xu, P., Liu, J., Jones, N., Cohen, J., & Ai, W. (2024). The promises and pitfalls of using language models to measure instruction quality in education. *arXiv.* https://doi.org/10.48550/arXiv.2404.02444

Systems for Smart Mobility

Dunja Srpak, Josip Srpak, Miljenko Mustapić
Application of Internet of Thins in urban mobility

Sirje Virkus, Marianne Paimre
Telepresence Robots as Emerging Technologies in Elderly Care and Wellbeing: The Case of Estonia

Mihael Radovan
Dynamic Lane Reversal Review

Application of Internet of Things in urban mobility

Dunja Srpak, Josip Srpak
University North
Department of Electrical Engineering
104. brigade 3, Varaždin, Croatia
dsrpak@unin.hr; jsrpak@unin.hr

Miljenko Mustapić
University North
Department of Logistics and Sustainable Mobility
Trg dr. Žarka Dolinara 1, Koprivnica, Croatia
mmustapic@unin.hr

Abstract*. This paper presents the results of research on the integration of Internet of Things (IoT) technologies within smart city initiatives, focusing on mobility solutions in Croatia and selected European urban areas. Through comparative analysis, the study investigates how IoT-based innovations enhance service availability, streamline urban movement, and support the development of sustainable city environments. The findings underscore the value of real-time data, accessible digital platforms, and citizen engagement in developing smarter, more resilient communities.*

Keywords. Internet of Things (IoT), public transport innovation, smart cities, urban mobility

1 Introduction

Smart management of urban entities, popularly called "smart cities," is gaining importance with the accelerating trend of urbanization. At the centre of this transformation is often the Internet of Things (IoT) technology, which enables the connection of different urban systems through a network of devices and sensors. Through the implementation of IoT in sectors such as transport, energy, security, and communal services, energy consumption, the use of public resources, and compliance with Environmental, Social, and Governance (ESG) standards are optimized. This contributes to a more sustainable, efficient, and dynamic urban future, moving towards smarter, greener cities.

The Internet of Things (IoT) is a network of interconnected devices that communicate and exchange data in real time (Srpak et al., 2024). In the context of smart cities, IoT is a key set of technologies that enable the connection and collection of data from a network of sensors and data processing, for intelligent decision-making and efficient city management. In addition, the comfort of living in an urban environment, both for local residents and tourists/visitors, depends on the accessibility of city services. Public transport services come first, followed by information on the availability of sports and cultural content at the current location, in real time.

The application of IoT in urban mobility and traffic management has been widely examined, with particular attention to its benefits, challenges, environmental impact, and improvements in user experience (Vaidian et al., 2019). However, a gap persists in models that would lead to an increase in the use of public transport within urban areas. Therefore, the focus of this paper is on the impact of technological innovation as a tool for sustainable urban mobility and reducing dependence on private cars, through the following research questions:

RQ1: What are the main barriers to implementing IoT-based models for increasing public transport use in urban areas?

RQ2: How can the implementation of IoT technologies affect the motivation of residents to use public transport instead of private cars in urban areas?

The article is organized as follows: after the introduction, the background section provides an overview of the relevant literature, followed by a description of the research methodology. The following section presents a comparative analysis of the application of modern technologies in public transport in several cities in Croatia and Europe. After a discussion of the findings, the corresponding conclusions are drawn at the end.

2 Background

In the future, the majority of the world's population (85% by 2050) is expected to live in cities. This demographic shift will require urban areas to enhance their services and resource management. The rapid advancement of IoT technologies is driving innovation in smart cities, improving system efficiency and quality of life, while also raising important questions about sustainability and environmental impact. Research by Nižetić et al. (2020) highlights both the opportunities and challenges of integrating IoT into urban environments, emphasizing the need for careful assessment to maximize benefits and minimize potential risks.

A study by Rajapaksha et al. (2017) proposed improving urban mobility through the introduction of an integrated iTrip (intelligent travel) framework,

which addresses urban sustainability issues by utilizing data generated in cities - such as traffic flow, people mobility, goods distribution, and environmental data collected from various sensors. Smart cities employ monitoring and communication technologies to optimize operations and increase sustainability, with sensors having an essential role. Given the diverse applications of these sensors, a review by Moreno et al. (2021) categorizes their use into six main groups: energy, health, mobility, security, water, and waste management, with a particular emphasis on efforts to create smart cities. A review study by Allwinkle and Cruickshank (2011) first explores what it means for cities to be "smart", and then offers an alternative path to smarter cities based on a critical and more realistic assessment. Additionally, the book by Coletta et al. (2019) draws on a series of international case studies and proposes social, policy, and practical interventions that recognize the benefits of smart city initiatives while also minimizing some of their potential drawbacks.

The paper by Bellini et al. (2022) provides an overview of the latest literature on IoT-based smart cities, highlighting the main trends and challenges in applying these technologies to develop sustainable and efficient cities. It also analyses the implementation of innovative services and advanced applications for various city stakeholders. In the paper Hurem and Kadrić (2023), present an overview of research on the application of IoT in urban areas, focusing on smart traffic systems, energy management, waste management, air quality monitoring, citizen safety, and related fields. Mobile applications that citizens can use to access information about city services, especially in problematic situations (natural disasters, traffic accidents, public protests...) were also analysed in particular. The study by Apata et al. (2024) highlights how IoT solutions in smart transportation infrastructure and fleet management enable more efficient, safer, and more sustainable transport systems through real-time monitoring and data-driven management. The authors emphasize the importance of integrating IoT into existing transportation systems, addressing data security and privacy issues, and fostering innovation, while recommending that future research analyse the long-term effects of IoT on preventive maintenance and real-time data processing.

Several articles, such as Hancke et al. (2013), address the challenges of increasing traffic congestion, pollution, and traffic accidents that accompany the rising number of vehicles in urban areas, and propose intelligent solutions for mobility and traffic control. Other studies focus on efforts to enhance pedestrian safety as a crucial aspect of improving urban mobility (Hussein et al., 2016). The IoT significantly improves transportation by automating processes and increasing both safety and efficiency; however, it is essential to address security and integration challenges with existing infrastructure. Application examples such as smart traffic lights, autonomous vehicles, and cargo monitoring systems demonstrate the considerable potential of this technology, as highlighted by Ziuziun and Parasiuk (2024).

From the other side, research by Belli et al. (2020) points out that discussions about smart cities often concentrate too heavily on technological and technical aspects, while the needs of citizens are sometimes overlooked. The IoT paradigm includes multiple scenarios and contexts, through a series of connected devices that collaborate to adapt the physical world to a changing environment.

Scientific research regarding the concept of smart cities, i.e. various technological and social innovations to improve urban life, predominantly focuses on urban mobility as a key subset, particularly regarding traffic and movement within cities. In these discussions, public transportation consistently emerges as a primary topic. Public transport users in all cities report varying levels of satisfaction with factors such as station accessibility, service frequency, and transfer options. Additionally, the availability of digital services for obtaining information or purchasing tickets has become an important factor influencing the overall user experience in public transport systems.

The primary motivation for this research was the identified lack of an effective model to encourage greater use of public transport among citizens. Given the interdisciplinary nature of the topic, the research team included experts in electrical engineering as well as logistics and transport sciences. This approach is essential for a comprehensive analysis of the challenges and potential solutions related to public transport usage. The paper presents initial comparisons of personal experiences within smaller groups, using examples of good practice from several EU cities. These findings can serve as a valuable foundation for future research on a larger scale, such as more systematic studies involving detailed surveys in multiple cities, carried out in partnership with city authorities and public transport providers.

3 Methodology

Relevant insight into the research area, (specifically, the connection between modern technologies and smart cities, particularly in the context of urban mobility) was obtained by searching the Web of Science platform and the Scopus database. At the same time, gaps were identified in the field of research, especially regarding the application of IoT to user experiences in public city transportation, as outlined in the introduction.

To encourage the wider implementation of IoT technology in the future, it is necessary to analyze the current situation and practices in different Croatian cities (both smaller and larger ones) and compare them with advanced systems in more developed European countries. Due to the impact of disruptions (such as traffic accidents, weather conditions, protest marches,

the passage of extraordinary cargo, VIP movements, etc.), changes to timetables should be updated in real time and communicated to users.

Through a case study conducted in several Croatian and European cities, different methods of organizing public transport systems were compared, with an emphasis on possible ways for users to access up-to-date information. Finally, the summary identifies the factors that influence the dynamics of implementing modern technologies and the actions worth taking to encourage the use of public transport over private cars in city centres.

4 Application of modern technologies in public transport

This section presents a comparative analysis of the application of modern technologies in public transport in Croatia and Europe. Among Croatian cities, the public transport systems in the author's hometowns (Varaždin and Koprivnica) and Zagreb, the largest city and capital of Croatia, are compared in more detail. The analysis of the public transport situation in several European cities is also summarized based on the experience of the authors and the people close to them (rather than on a public survey).

4.1 Public transport in Varaždin

Varaždin is connected to other places and cities in Croatia and Europe by trains or buses, although passengers use buses more often than trains, mainly due to better timetables and shorter travel times. However, attempts to organize a functional public bus transport within the city of Varaždin have been failing for a number of years, especially in an economically sustainable way. Without majority public funding of municipal passenger transport, it is not possible to maintain lines on which buses run almost empty. Varaždin currently has five bus lines that complete a circuit through part of the city and suburban. They meet only in the city centre, and some lines touch or intersect at some other stations, as shown in Fig. 1 (Gradski-buseki, n.d.-a).

Although the bus routes are arranged to circulate through several streets, some residents of Varaždin still have a long way to get to the nearest station. On the other hand, this makes some parts of the route faster to cover on foot (by taking a shortcut) than by bus, which circulates in this way. In addition, the lines are not frequent enough to be suitable for going to school or work, so pensioners who have enough time and free tickets mainly use them. For example, (Fig. 2) to get to work in the center of Varaždin at 7:30 am, from suburbs like Zbelava or Kučan Marof, you need to leave a little after 6 am (6:08 - 6:15), line 5 arrives in the center at 6:32 where transfers to other lines are possible (towards the University, hospital, high schools, bus or train station, etc.). The next departure from Kučan Marof is at 7:18 and arrives in the centre at 7:38, which is too late. The distance is about 4 km, which can be reached by car (even during the morning rush hour) in about ten minutes, and on foot in less than an hour. The same situation exists on the other side of the city, line 1 (Hrašćica). Therefore, private cars are still the most common means of transportation in the city, possibly bicycles or scooters, when the weather is nice or on weekends (during free time).

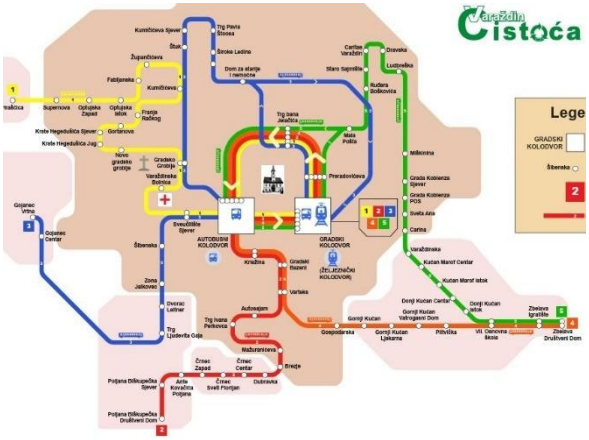


Figure 1. Public transport network in Varaždin

LINE 5 — Zbelava - Donji Kućan - Kućan Marof - Banfica - Bus Station - City Station

No.	St Name / Address of Stop	DEPARTURE TIMES							Possibility of transfer to the line				
		M-F	M-S	M-S	M-S	M-S	M-S	M-S	M-F	M-F	M-F	M-F	M-F
1	Zbelava * / Varaždinska ulica 119	5:03	6:08	7:13	8:23	10:28	11:43	12:58	14:43	15:48	16:58	19:48	20:58	
2	Zbelava ** / Varaždinska ulica 65	5:04	6:09	7:14	8:24	10:29	11:44	12:59	14:44	15:49	16:59	19:49	20:59	
3	VII. Osnovna škola / Varaždinska ulica 131	5:05	6:10	7:15	8:25	10:30	11:45	13:00	14:45	15:50	17:00	19:50	21:00	
4	Donji Kućan Istok / Varaždinska ulica 111	5:06	6:11	7:16	8:26	10:31	11:46	13:01	14:46	15:51	17:01	19:51	21:01	
5	Donji Kućan Centar / Varaždinska ulica 73	5:07	6:12	7:17	8:27	10:32	11:47	13:02	14:47	15:52	17:02	19:52	21:02	
6	Kućan Marof Istok / Varaždinska ulica 161	5:08	6:13	7:18	8:28	10:33	11:48	13:03	14:48	15:53	17:03	19:53	21:03	
7	Kućan Marof Centar / Varaždinska ulica 93	5:09	6:14	7:19	8:29	10:34	11:49	13:04	14:49	15:54	17:04	19:54	21:04	
8	Varaždinska / Varaždinska ulica 27	5:10	6:15	7:20	8:30	10:35	11:50	13:05	14:50	15:55	17:05	19:55	21:05	
9	Customs / Ulica Braće Krajanac 18	5:11	6:16	7:21	8:31	10:36	11:51	13:06	14:51	15:56	17:06	19:56	21:06	
10	Sveta Ana *** / Ulica Vilka Novaka 30/A	5:13	6:18	7:23	8:33	10:38	11:53	13:08	14:53	15:58	17:08	19:58	21:08	
11	Grada Kobenza POS / Ul. Grada Kopenza 4	5:14	6:19	7:24	8:34	10:39	11:54	13:09	14:54	15:59	17:09	19:59	21:09	
12	Grada Kobenza **** / Ul. Grada Kopenza 22	5:15	6:20	7:25	8:35	10:40	11:55	13:10	14:55	16:00	17:10	20:00	21:10	
13	Miškinina / Ulica M.P. Miškine 64	5:16	6:21	7:26	8:36	10:41	11:56	13:11	14:56	16:01	17:11	20:01	21:11	
14	Ludbreška / Sored (Vukovarska 2A)	5:20	6:25	7:30	8:40	10:45	12:00	13:15	15:00	16:05	17:15	20:05	21:15	
15	Dravska / Dravska ulica 30	5:21	6:26	7:31	8:41	10:46	12:01	13:16	15:01	16:06	17:16	20:06	21:16	
16	Caritas Varaždin / Bombel esova cesta 15	5:22	6:27	7:32	8:42	10:47	12:02	13:17	15:02	16:07	17:17	20:07	21:17	
17	Staro Sajmište / Trg Matije Gupca 54	5:23	6:28	7:33	8:43	10:48	12:03	13:18	15:03	16:08	17:18	20:08	21:18	
18	Rudera Boškovića / U. Rudera Boškovića 21	5:25	6:30	7:35	8:45	10:50	12:05	13:20	15:05	16:10	17:20	20:10	21:20	
19	***** Ulica Ivana Kukuljevića Sakcinskog 35	5:26	6:31	7:36	8:46	10:51	12:06	13:21	15:06	16:11	17:21	20:11	21:21	
20	Trg Bana Jelačića / Trg bana Josipa Jelačića 9	5:27	6:32	7:37	8:47	10:52	12:07	13:22	15:07	16:12	17:22	20:12	21:22	1,2,3,4
21	Bus Station / Kapucinski trg 6	5:31	6:36	7:41	8:51	10:56	12:11	13:26	15:11	16:16	17:26	20:16	21:26	1,2,3,4
22	City Station / Trg kralja Petra Svačića 3	5:35	6:40	7:45	8:55	11:00	12:15	13:30	15:15	16:20	17:30	20:20	21:30	1,2,3,4
1	City Station / Trg kralja Petra Svačića 3	5:40	6:45	7:50	10:00	11:05	12:20	14:15	15:20	16:25	19:00	20:25	22:15	1,2,3,4
2	Preradovićeva / Ulica Petra Preradovića 15	5:43	6:48	7:53	10:03	11:08	12:23	14:18	15:23	16:28	19:03	20:28	22:18	1,2,3,4
3	Mala Pošta / Međimurska ulica 2	5:44	6:49	7:54	10:04	11:09	12:24	14:19	15:24	16:29	19:04	20:29	22:19	
4	Rudera Boškovića / U. Rudera Boškovića 14a	5:45	6:50	7:55	10:05	11:10	12:25	14:20	15:25	16:30	19:05	20:30	22:20	
5	Staro sajmište / Trg Matije Gupca 54	5:47	6:52	7:57	10:07	11:12	12:27	14:22	15:27	16:32	19:07	20:32	22:22	
6	Caritas Varaždin / Bombel esova cesta 15	5:48	6:53	7:58	10:08	11:13	12:28	14:23	15:28	16:33	19:08	20:33	22:23	
7	Dravska / Dravska ulica 15	5:49	6:54	7:59	10:09	11:14	12:29	14:24	15:29	16:34	19:09	20:34	22:24	
8	Ludbreška / Ludbreška ulica 70	5:50	6:55	8:00	10:10	11:15	12:30	14:25	15:30	16:35	19:10	20:35	22:25	
9	Miškinina / Ulica M.P. Miškine 47	5:54	6:59	8:04	10:14	11:19	12:34	14:29	15:34	16:39	19:14	20:39	22:29	
10	Grada Kobenza **** / Ul. Grada Kopenza 37	5:55	7:00	8:05	10:15	11:20	12:35	14:30	15:35	16:40	19:15	20:40	22:30	
11	Grada Kobenza POS / Ul. Grada Kopenza 1	5:56	7:01	8:06	10:16	11:21	12:36	14:31	15:36	16:41	19:16	20:41	22:31	
12	Sveta Ana *** / Ulica Vilka Novaka 32	5:57	7:02	8:07	10:17	11:22	12:37	14:32	15:37	16:42	19:17	20:42	22:32	
13	Customs / Ulica Braće Krajanac 18	5:59	7:04	8:09	10:19	11:24	12:39	14:34	15:39	16:44	19:19	20:44	22:34	
14	Varaždinska / Varaždinska ulica 30	6:00	7:05	8:10	10:20	11:25	12:40	14:35	15:40	16:45	19:20	20:45	22:35	
15	Kućan Marof Centar / Marofska ulica 1	6:01	7:06	8:11	10:21	11:26	12:41	14:36	15:41	16:46	19:21	20:46	22:36	
16	Kućan Marof Istok / Varaždinska ulica 148	6:02	7:07	8:12	10:22	11:27	12:42	14:37	15:42	16:47	19:22	20:47	22:37	
17	Donji Kućan Centar / Varaždinska ulica 66	6:03	7:08	8:13	10:23	11:28	12:43	14:38	15:43	16:48	19:23	20:48	22:38	
18	Donji Kućan Istok / Varaždinska ulica 106	6:04	7:09	8:14	10:24	11:29	12:44	14:39	15:44	16:49	19:24	20:49	22:39	
19	VII. Osnovna škola / Varaždinska ulica 132	6:05	7:10	8:15	10:25	11:30	12:45	14:40	15:45	16:50	19:25	20:50	22:40	
20	Zbelava ** / Varaždinska ulica 67	6:06	7:11	8:16	10:26	11:31	12:46	14:41	15:46	16:51	19:26	20:51	22:41	
21	Zbelava * / Varaždinska ulica 119	6:07	7:12	8:17	10:27	11:32	12:47	14:42	15:47	16:52	19:27	20:52	22:42	

* Community Center / ** Playground / *** Konzum / **** North / ***** Little Tos: Office - Bakery Žito / M-F (Monday-Friday) / M-S (Monday-Saturday)

Figure 2. Timetable of public transport line 5 in Varaždin

Regarding the application of modern technologies, the timetable is available on the web and also delivered by mail with the garbage collection bill (since the same

company provides both household waste collection and public city transport services). In timetables the arrival time and the addresses of the stops for each line (e.g. at house number 27 or 161) are indicated, while most bus stops have no signs at the location, not even a metal plate with a bus sign, let alone a screen with the arrival time of the next line(s). The timetables for Line 5 is shown as an example in Fig. 2 (Gradski-buseki, n.d.-b). The last column shows at which stops it is possible to transfer to other lines. In addition, the official website of the public transport provider contains a table with a list of stops, geolocation and a link to the location of the stop in Google Maps.

4.2 Public transport in Koprivnica

Koprivnica's connectivity with surrounding towns is similar to that of Varaždin (Hrvatske ceste, 2022.; Feletar, 2016), and lines within the city itself face the same problems. Free public transport by electric buses (BusKo, Fig. 3) was introduced as part of the CIVITAS DYN@MO project (2012-2015), which is largely financed by the European Union's programme for scientific research and development in the context of competitiveness. It aimed to develop the most effective transport sustainability measures based on electromobility and sustainable urban transport planning, and was implemented in four cities: Aachen, Palma de Mallorca, Gdynia and Koprivnica. Line 1 has five stops and the interval between two arrivals (the frequency of arrivals) is more than an hour. as in Varaždin, while line 2 has more stops, but runs only twice a day, from Monday to Saturday.



Figure 3. Public transport lines in Koprivnica

These buses can carry 12 passengers, their maximum speed is limited to 90 kilometers per hour, and they have been in operation since 2018. The city of Koprivnica did not have an established public transport system before that, but it is arguable whether the number of daily passengers justifies the predicted vision of this project.

In addition to electric buses, this project also introduced a public bicycle transport system - BicKo, which can be used seven days a week from 6:00 to 22:00 during the season (April - October). Information about available bicycles and stands is available to users at each pylon and on the system's website.

The BicKo system has 70 bicycles (60 regular and 10 electric) available at eight stations throughout the city. Data shows it is a successful project and that the most popular pick up and drop off stations are at the train station and the city market, which means both locals and visitors, have embraced this modern way of getting around. Usage varies throughout the year: in 2022, there were 2,950 rentals, while in 2021 there were 4,380, though that year saw a drop due to construction work at Zrinski Square, which made one terminal harder to reach. Since the system started, there have been over 89,000 rentals, and in 2022, there were 112 active users. Bearing in mind there is a significant increase in new users of 10% each year, this project will continue its success. Currently, more than 1,400 people are registered, and the system is increasingly used by tourists. Figure 4 shows one of the BicKo system stations in Koprivnica, (Podravina, n.d.).



Figure 4. BicKo system station in Koprivnica

When the system was introduced, to start using the bikes, it was first necessary to sign a rental agreement with the BicKo system operator. The user was then issued an Radio Frequency Identification (RFID) card or a one-time bike rental code, which allowed them to unlock an available bike at any BicKo terminal. After a maximum of 120 minutes, the bike had to be returned to the next available terminal. Today, the bike stations are clearly marked and located at key spots in the city, such as the train station, city market, and university campus. At each station, users can check bike availability and station status via the Go2bike website and mobile app, which provide real-time updates on available bikes and free docking spots. This makes it easier to plan trips and ensures a smooth rental experience. The app lets users register, scan a QR code on the bike, and start riding - no paperwork needed.

4.3 Public transport in Zagreb

The company Zagreb Electric Tram (ZET - Zagrebački električni tramvaj) manages public transport in Zagreb. The tram network covers public transport in the city centre of Zagreb, while buses run to the outskirts of the city. An overview of the tram network and stations in the inner city is shown in Fig. 5, (ZET n.d.). Zagreb is a much larger city and significantly more people use public transport than in small towns, such as Varaždin and Koprivnica.

Two mobile applications are available to passengers: "ZET info" and "moj ZET" (my ZET). ZET info contains information on the current status, changes and warnings about the movement of trams and buses, and in the event of works or events (races, protests, accidents, track works). The website also provides a fixed timetable, which is often not relevant due to congestion or road conditions. The ZET info application provides the current status, i.e. the location of all vehicles, detected by their built-in GPRS, with an overview by lines or stations. Also, a fixed timetable is posted at each station, but it is often outdated or worn/faded due to the passage of time and non-renewal. In addition, at certain stations (about 50%, mainly at larger stations) there is a screen with a live display of upcoming trams, which is very reliable, like the ZET info application.

The my ZET application is newer (May 2023), and can be used to purchase tickets for all city zones, including the funicular. Funds need to be transferred into the application in advance, which are deducted from the account when purchasing a ticket (prepaid). When boarding a tram or bus, it is necessary to scan the QR code or Near Field Communication (NFC) tag built into each door using the application (with a mobile phone), which activates the ticket.

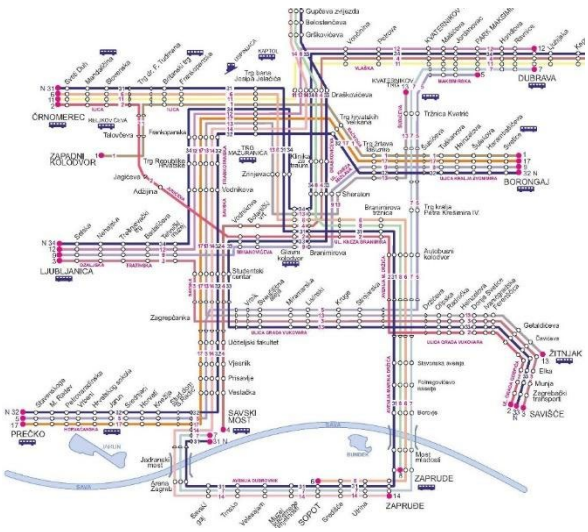


Figure 5. Tramlines overview in Zagreb

Waiting times for trams during the day are usually 10-15 minutes, and for buses are up to 20-30 minutes. For everyday use (going to school or work), it is convenient to use public transport within the city centre of Zagreb. However, for longer distances, if the route has multiple transfers, the journey can be quite long, especially with delays on some lines.

4.4 Experiences with public transport systems from Europe

Many European cities have been using public transport for a long time and have developed many different variants of it, but what they all have in common is the constant improvements trough the integration of new technologies.

In Stuttgart, one of Germany's biggest cities, there is an intricate network of buses, street trams (U-Bahn) and light rail (S-Bahn) spanning tens of kilometres in every direction. Most stops are equipped with screens indicating live updates of arrival times, in addition to the printed schedule. Multiple apps (VVS, DB...) also provide live updates and travel schedule suggestions, which have certain minor limitations and assumptions that can affect the suggestions. For example, when switching between a bus and a light rail, the program assumes that a person takes 5 minutes to travel the distance between those stops. Therefore, a connection within 5 minutes of the bus arrival will not be shown, even though someone running may arrive in 2 minutes. Nevertheless, the apps are absolutely necessary to navigate the public transport system due to frequent construction works and cancellations. Despite occasional issues, the German public transport system is one of the most affordable in the world, thanks to the Deutschland Ticket initiative. That is a monthly ticket that enables unlimited travel on local and regional trains, buses, trams, and the underground across Germany for a set fee.



Figure 6. Screen at the bus stop in Stuttgart with arrival times updated live, recorded on May 26, 2025.

In Copenhagen, the official Danish public transport app DSB offers several options for purchasing a ticket.

It is possible to buy a ticket with a limited duration for a certain number of zones and not think about it any further. However, one can save money by enabling GPS, i.e. allowing access to personal data stored in other apps (such as health-related apps). Furthermore, a person needs to continuously map their journey in the app, marking the change of transport mode (bus, metro, light rail). Exchanging personal data for a discount may seem like a strange choice, but the data collected in this way are used to improve the service in the future. Therefore, the quality and punctuality of the public transport system there is unquestionable.

In Zürich, one can always rely on high punctuality and very detailed updates on any possible issues. Multiple apps (ZVV, SBB) allow passengers to navigate the transport smoothly. Zürich has no metro, and light rail mainly connects suburbs further away, but the buses and trams come frequently and cover the whole city. SBB also has a tracking method by which the app can track the location of the passenger and automatically charge accordingly.

In Stockholm, public transport is organized by bus, tram, and boat. The mobile app allows easy navigation through the different transport options, with a built-in map of the lines included. Using the same app, it is possible to buy a ticket (single, daily, monthly, etc.) to access any form of transport. The waiting time at the station until the next line arrives is comparable to that Zagreb, i.e., the gap between two arrivals of the same line is about ten minutes, during the day.

Cyprus, as an island country, is in the EU, but it is still a bit isolated. It has city buses and intercity buses, and a couple of apps that show a fixed timetable, but they can be confusing to navigate. The waiting time for the next bus is 30-60 minutes, depending on the line. Cyprus is relatively small, so a 30-minute wait for a bus that runs for 1 hour to the other end of the island is acceptable. Apart from these apps, there is no digitalization, and tickets are purchased with cash on the bus.

5 Discussion

By comparing examples of good practice for an efficient and accessible public transport system in some European cities, ideas applicable in Croatia can be obtained. An affordable monthly ticket for a larger area and different modes of transport significantly encourages greater use of public transport and facilitates the payment system, although it is not completely free. The idea is similar to the vignette payment system for motorways, unlimited travel for a fixed monthly amount as in many European countries (Slovenia, Hungary, Slovakia, Czech Republic, Austria), as applied to the Deutschland Ticket.

A prepaid system like the one in Zagreb via the "my ZET" application, if the app is activated when entering a bus or tram, would correspond to the current motorway payment system in Croatia via the Electronic Toll Collection (ENC) system. It is planned to replace the ENC payment system so that vehicles are tracked and charged for exactly the number of kilometers of motorway used in the near future. This requires the installation of stations with cameras that automatically read vehicle license plates at multiple locations, so the existing toll stations on motorways can be removed. Such a system can be applied to track public transport users (via mobile phone applications) and charge accordingly, as in Zurich. This is where the application of new technologies such as IoT comes to the fore.

A good example is also the BicKo system in Koprivnica, created as a free, simple, and accessible way to get around the city, aiming to reduce CO2 emissions and promote sustainable mobility. While it is popular and increasingly used, especially by tourists, its usage still varies by season and there is room for further growth. Key improvements should include expanding the network to nearby municipalities, adding more e-bikes, and better integrating the system with other modes of public transport.

In the process of transitioning towards a smart city, it is essential to consider not only the opportunities and challenges related to smart urban infrastructure, but also how these new developments and the advancement of digitalization can enhance the overall quality of life for citizens. Smarter communities are characterized by active community engagement in adopting new technologies, which are shaped by local and regional social needs and values. In addition, the integration of artificial intelligence (AI) into traffic and resource management is becoming increasingly prevalent. As highlighted by Gheorghe and Soica (2025), AI is most commonly used for advanced data analysis and the prediction of traffic patterns.

To enable effective cooperation between cities, data openness and adherence to common standards are particularly important. However, the widespread adoption of digital technologies such as IoT in everyday life, especially when it involves broader sharing of personal data, introduces significant vulnerabilities, particularly in the areas of cybersecurity and privacy. The risk of cyberattacks and data breaches remains a major concern, as sensitive personal and institutional information may be exposed. It is therefore crucial to implement robust security protocols, raise public awareness, and ensure compliance with data protection regulations to safeguard both individuals and the community as a whole.

6 Conclusion

Based on the analysis of conditions in selected cities in Croatia and Europe, a review of the available literature, and the discussion presented, answers to the research questions have emerged.

RQ1: The main obstacles to implementing IoT-based models in urban public transport are the lack of digital infrastructure, especially in smaller cities, and the fragmentation between different transport providers, which hinders system integration. Additionally, concerns about data privacy and cybersecurity, as well as the need for data standardization, present significant challenges. Limited public awareness and the financial demands of building and maintaining IoT systems further complicate widespread adoption.

RQ2: The implementation of IoT technologies can motivate residents to use public transport instead of private cars by providing real-time information on arrivals, delays, and route changes, thereby increasing convenience and reliability. Integrated digital platforms and mobile apps also make ticketing and multimodal travel simpler and more accessible, enhancing the overall user experience. As public transport becomes more predictable, efficient, and user-friendly through IoT solutions, it becomes a more attractive alternative to private vehicles, helping to reduce congestion and emissions in urban areas.

To maximize the benefits of smart city initiatives, several improvements can be made to urban mobility and digital services:

- Integrated Transport Recommendations:

Link notifications from different transport modes (such as trains, trams, buses, ferries, catamarans, and funiculars), even if they are operated by different providers or fall under different national or local jurisdictions. This would enable seamless, multimodal travel planning and real-time updates for all users.

- Visitor Information Points:

Establish digital information points for visitors who are unfamiliar with the city. These points should provide suggestions for destinations that are accessible on foot or by available means of transport, including tourist attractions, cultural and sports facilities, universities, hospitals, police stations, and other key locations.

- Incentivizing Public Transport Use:

In smaller cities such as Varaždin, Koprivnica, and even Zagreb, motivating the population to use public transport more frequently can be achieved by making it free. This approach has been shown to increase ridership and reduce reliance on private vehicles, contributing to reduced traffic congestion and lower emissions.

- Expanding IoT Applications:

In addition to public transport, urban mobility should be supported by the broader application of the Internet of Things (IoT) across strategic and operational levels. This includes the deployment of intelligent waste containers, smart parking systems, and dynamic travel recommendation systems. Furthermore, cross-city applications targeting tourists and other visitors can be implemented to provide relevant tourist information and facilitate seamless navigation throughout the urban environment.

By addressing these opportunities and challenges, cities can foster smarter, more resilient communities that leverage digital innovation to improve the quality of life for all residents and visitors.

In future research, the goal is to collect, in a systematic manner, additional and varied data from a wider selection of cities, using surveys along with other relevant sources. By expanding the research team to include experts from the social sciences, as well as electrical and traffic engineering, the study will be better positioned to provide deeper analysis of emerging trends, explore innovative solutions, and develop original methodological approaches for improving urban mobility.

References

Allwinkle, S., & Cruickshank, P. (2011). Creating Smart-er Cities: An Overview. Journal of Urban Technology, 18(2), 1–16. doi.org/10.1080/10630732.2011.601103

Apata SB, Oyenuga M, Adeoba M, Michael U, Kenneth M, Adeyinka O. (2024). Internet of Things (IoT) solutions for smart transportation infrastructure and fleet management. Tuijin Jishu/Journal of Propulsion Technology. 45(4), 1492–509.

Belli, L., Cilfone, A., Davoli, L., Ferrari, G., Adorni, P., Di Nocera, F., Dall'Olio, A., Pellegrini, C., Mordacci, M., & Bertolotti, E. (2020). IoT-Enabled Smart Sustainable Cities: Challenges and Approaches. Smart Cities, 3(3), 1039-1071. https://doi.org/10.3390/smartcities3030052

Bellini, P., Nesi, P., & Pantaleo, G. (2022). IoT-Enabled Smart Cities: A Review of Concepts, Frameworks and Key Technologies. Applied Sciences, 12(3), 1607. doi.org/10.3390/app12031607

Coletta, C., Evans, L., Heaphy, L. & Kitchin, R. (2019). Creating smart cities. Abingdon, Oxon: Routledge.

Feletar, Petar, (2016). Geografski položaj i prometna povezanost jezera 'Šoderica'. Journal od multidisciplinary research 'Podravina', 29(12-23)

Gheorghe, C., & Soica, A. (2025). Revolutionizing Urban Mobility: A Systematic Review of AI, IoT, and Predictive Analytics in Adaptive Traffic Control Systems for Road Networks. Electronics, 14(4), 719. https://doi.org/10.3390/electronics14040719

Gradski-buseki (n.d.-a), Public transport network in Varaždin, Varaždin.hr/gradski-buseki/JP mreža (2025). Retrieved May 29, 2025, from https://varazdin.hr/upload/2025/03/varazdin_jp-mreza_03032025_67d81ea9af79d.pdf

Gradski-buseki (n.d.-b), Timetable of public transport line 5 in Varaždin Varaždin.hr/gradski-buseki /Linija 5 (2025). Retrieved May 29, 2025, from https://varazdin.hr/upload/2025/03/linija_5_67d81e9c2e152.pdf

Hancke, G. P., Silva, B. D. C. e., & Hancke, Jr., G. P. (2013). The Role of Advanced Sensing in Smart Cities. Sensors, 13(1), 393-425. doi.org/10.3390/s130100393

Hrvatske ceste d.o.o. (2022). Environmental impact study. Environmental impact study for the intervention: Expressway DC10 Dubrava interchange (DC26)- Križevci-Koprivnica-Gola GP border crossing.

Hurem, A., Kadrić. M. (2023). Impact of the internet of things on the development of smart cities: transformation of urban infrastructure towards a sustainable world, Proceedings of the 27. International conference Economic, legal and media transformation through the green economy of the western Balkan countries with special reference to Bosnia and Herzegovina, pp. 322-332, Travnik: Internacionalni univerzitet Travnik

Hussein, A., García, F., Armingol, J.M., Olaverri-Monreal, C. (2016). P2V and V2P communication for Pedestrian warning on the basis of Autonomous Vehicles," 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC 2016,), Rio de Janeiro, Brazil, pp. 2034-2039, doi: 10.1109/ITSC.2016.7795885

Nižetić, S., Šolić, P., Gonzalez-De, D. L. D. I., & Patrono, L. (2020). Internet of things (IoT): Opportunities,issues, and challenges toward a smart and sustainable future. Journal of Cleaner Production, 274,122877. 1-32. doi: 10.1016/j.jclepro.2020.122877

Patačko, Bruno. Master thesis (2022). The impact of transport infrastructure on the development of Koprivnica-Križevci County. Faculty of Economics and Business, Zagreb, https://repozitorij.efzg.unizg.hr/en/islandora/object/efzg%3A9059/datastream/PDF/view

Podravina (n.d.) BicKo stations, ePODRAVINA.HR, News portal (2023). Retrieved May 29, 2025, from https://epodravina.hr/bicko-se-vec-devet-godina-vozika-gradom-plan-je-siriti-stanice-na-okolne-opcine-a-posudba-je-sad-jos-jednostavnija-medurecan-uvjet-je-kartica-ali-sve-je-i-dalje-besplatno/

Rajapaksha, P., Farahbakhsh, R., Nathanail, E., Crespi, N. (2017). iTrip, a framework to enhance urban mobility by leveraging various data sources, Transportation Research Procedia, Volume 24, Pages 113-122,

Ramírez-Moreno, M. A., Keshtkar, S., Padilla-Reyes, D. A., Ramos-López, E., García-Martínez, M., Hernández-Luna, M. C., Mogro, A. E., Mahlknecht, J., Huertas, J. I., Peimbert-García, R. E., Ramírez-Mendoza, R. A., Mangini, A. M., Roccotelli, M., Pérez-Henríquez, B. L., Mukhopadhyay, S. C., & Lozoya-Santos, J. d. J. (2021). Sensors for Sustainable Smart Cities: A Review. Applied Sciences, 11(17), 8198. doi.org/10.3390/app11178198

Srpak, D., Šac, V., Tomičić, E. (2024). Introducing the students to the world of IoT. Proceedings of 16th International Conference on Education and New Learning Technologies, EDULEARN24, Palma, Spain, pp. 7104-7111. doi: 10.21125/edulearn.2024.1682

Vaidian, I., Azmat, M., Kummer, S. (2019). Impact of Internet of Things on Urban Mobility. In Innovation Arabia 12 Proceedings: Health and Environment Conference (Vol. 12, pp. 4-17). https://innovationarabia.ae/wp-content/uploads/2020/10/IA-12-Proceedings-Health-and-Environment.pdf

ZET (n.d.), Tram lines in Zagreb, Zagrebački električni tramvaj d.o.o. Mreža tramvajskih linija (2025). Retrieved May 29, 2025, from https://www.zet.hr/UserDocsImages/Prilozi/pdf/Mre%C5%BEe%20tramvajskih%20linija%202018-dnevne%20i%20no%C4%87ne/tram%20mre%C5%BEa%20va%C5%BEe%C4%87a%20A3_4.4.25..pdf?vel=1523880

Ziuziun, V., Parasiuk, L. (2024). IoT in the transportation sector: opportunities and challenge. XLIII Internationalscientific and practical conference “Modern Challenges and schievements of the scientific community of the 21st century”, P. 65–70

Telepresence Robots as Emerging Technologies in Elderly Care and Wellbeing: The Case of Estonia

Sirje Virkus, Marianne Paimre
School of Digital Technology
Tallinn University
Narva mnt 25, 10120 Tallinn, Estonia
sirje.virkus@tlu.ee, marianne.paimre@tlu.ee

Abstract. *This study investigates the role of telepresence robots (TPRs) in supporting older adults through a mixed-methods approach combining a systematic literature review and 54 semi-structured interviews with caregivers, medical professionals, and older adults in Estonia. Findings highlight TPRs' potential to reduce loneliness, improve access to care, and enhance autonomy. A persona-driven analysis reveals diverse needs and attitudes among caregivers, medical professionals, and elderly individuals. Key challenges include ethical concerns, technological unfamiliarity, and infrastructural constraints. The study offers actionable design and policy recommendations to support user-centered TPR integration in eldercare, emphasizing inclusivity, trust-building, and contextual adaptability to ensure successful deployment.*

Keywords. telepresence robots, elderly care, caregivers, medical professionals, older adults, literature review, qualitative research, personas, Estonia

1 Introduction

Telepresence robots (TPRs) are remotely operated mobile devices equipped with video, audio, and navigational capabilities. Typically integrating cameras, microphones, speakers, and display screens, these robots enable users to experience a sense of physical presence in remote environments. As a form of technological mediation, TPRs facilitate real-time interaction across geographical distances, offering new possibilities for communication and engagement (Virkus, 2025).

The pioneering work by Paulos and Canny (1998) marked the first systematic exploration of mobile robotic telepresence. Their research was primarily motivated by a desire to understand the social and psychological dimensions of sustained human-to-human interaction through technological mediation. Since then, TPRs have attracted interests across various sectors, including healthcare, education, industry, museums, and corporate environments (Virkus et al., 2023).

One particularly promising domain for TPRs is elder care. The aging of populations presents substantial challenges to healthcare and welfare systems globally (EC, 2021; WHO, 2015). At the same time, the expansion of digitalization has created opportunities for promoting independent and healthy aging (Nimrod, 2020). In many high-income countries, digital health applications now empower older adults to monitor their well-being and manage minor medical issues autonomously (Arthanat, 2021; Chu et al., 2021). However, despite the growing availability of smart technologies, a considerable portion of the elderly population remains either reluctant or unable to use them effectively (Paimre et al., 2023).

Estonia, a recognized leader in digital transformation, exemplifies the potential and limitations of this shift. It has achieved notable progress in e-government and e-health, consistently ranking among the top EU countries in digital public services (European Commission, 2022; Ojaperv & Virkus, 2023). However, Estonia's social welfare provision, especially for vulnerable groups like older adults, still trails behind that of more developed Western nations (Leppiman et al., 2021; Paimre et al., 2023, 2024).

This paper investigates the role of TPRs in providing support and companionship for the elderly, with a specific focus on the Estonian context. Combining a systematic literature review with qualitative interviews of caregivers, medical professionals, and elderly individuals, the study identifies key research gaps and stakeholder perspectives. Thematic analysis of the interviews informed the creation of user personas, offering nuanced insights into needs, preferences, and barriers. Together, these methods provide a foundation for adapting and implementing TPRs to enhance the quality of life and care for Estonia's aging population.

The structure of this paper is as follows: Section 2 outlines the research methodology. Section 3 presents the findings of the literature review. Section 4 provides an account of the personas derived from interviews with caregivers, medical professionals, and elderly

individuals. Section 5 offers a discussion and concluding remarks.

2 Methodology

This study employed a mixed-methods approach that combined a systematic literature review with a persona-driven approach grounded in qualitative interviews. The research design was structured in two sequential phases to ensure both breadth of understanding from existing literature and depth of insight from real-world experiences. Phase one focused on synthesizing current academic knowledge, while phase two aimed to explore the perspectives of key stakeholders through semi-structured interviews and develop user personas grounded in the data.

The literature review was conducted in February 2024 following the PRISMA *(Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) framework (Moher et al., 2009), a widely recognized standard for enhancing transparency, completeness, and replicability in systematic reviews.

Searches were performed in the Web of Science and Scopus databases using the terms "telepresence robot*" AND ("elderly" OR "older adults"). In Web of Science, queries were conducted by topic, while in Scopus, they covered the title, abstract, and keywords fields. The search yielded 25 results from Web of Science and 114 from Scopus. After deduplication and screening for relevance, 81 articles were selected for in-depth analysis. These studies were examined using thematic analysis, enabling the identification of recurring themes, knowledge gaps, and divergent findings within the literature. This phase provided an overview of current research and informed the design of the interview protocol for the second phase.

To explore user experiences and expectations regarding TPRs, a qualitative research design using semi-structured interviews was adopted. This approach allowed for flexibility in probing participant perspectives while maintaining consistency across interviews. The interviews sought to capture rich, first-person insights from key stakeholders involved in elderly care.

A total of 54 participants were interviewed, drawn from three target groups: caregivers (n=15), medical professionals (n=15), and elderly individuals (n=24). A convenience sampling method was employed to recruit caregivers and medical professionals. Elderly individuals were selected based on inclusion criteria that required the individuals to be aged 65 or older, with varying degrees of technological familiarity, to ensure a broad range of viewpoints. The caregivers ranged in age from 40 to 70 years, the medical professionals from 25 to 82 years, and the elderly participants from 65 to 86 years. To enhance understanding among participants unfamiliar with robotics, a brief video demonstration of a TPR was shown prior to the interviews. Each interview was conducted face-to-face, lasted approximately 60 minutes, and adhered strictly to ethical protocols, including informed consent, confidentiality, and participants autonomy.

The interview protocol was structured around ten thematic areas:

1. General understanding and prior experience.
2. Needs and challenges in maintaining social connections.
3. Desired features and functionalities.
4. Comfort and trust in using the technology.
5. Social and emotional impact.
6. Levels of support and autonomy.
7. Preferences for appearance and customization.
8. Safety and privacy concerns.
9. Technical support and training needs.
10. Future use and adoption potential.

Interview data were transcribed verbatim and analyzed through inductive thematic analysis. Coding occurred in multiple iterative phases, enabling the identification of patterns and the emergence of overarching themes. This thematic structure formed the analytical basis for interpreting the interviews in light of the study's research objectives.

Drawing on the insights from the interviews, a set of user personas was systematically developed to represent archetypal stakeholders in the TPR ecosystem. Personas were developed through a systematic analysis of interview data using thematic coding and clustering techniques. These personas encapsulate the diverse motivations, behaviors, needs, and barriers associated with TPR adoption and usage. Each persona reflects a unique cluster of user characteristics and provides a conceptual model to: 1) Illustrate variation within and across stakeholder groups. 2) Inform the user-centered design of TPRs. 3) Guide implementation strategies and policy development in elderly care contexts.

Grounding the design of telepresence technologies in the experiences of end-users, these personas serve as a practical framework for creating inclusive, responsive, and adaptable robotic solutions that align with the real-world needs of older adults and their care networks.

3 Literature review

The systematic literature review was guided by the five research questions (RQs), which structured the data collection, analysis, and interpretation processes:

RQ1: What are the main topics addressed in publications concerning TPRs in elderly care?

The literature encompasses a broad range of themes related to the development, design, and implementation of TPRs, with particular emphasis on their potential to enhance social connectedness, well-being, and independent living among older adults (Moyle et al., 2014). A significant body of work investigates the perspectives of key stakeholders, including individuals living with dementia, caregivers,

and healthcare professionals, to better understand the needs and concerns surrounding TPR adoption in care contexts (Shin et al., 2022). In addition, researchers examine the technological attributes of TPRs, such as mobility, interface usability, and communication capabilities, and how these influence user experience and acceptance (Ragno et al., 2023). Evaluative studies of specific TPR models highlight both the practical applications and limitations of existing technologies in real-world settings (Casiddu et al., 2015). Furthermore, investigations into psychophysiological responses provide insights into how older adults emotionally and cognitively engage with TPRs during interaction (Tiberio et al., 2012), contributing to a deeper understanding of human-robot dynamics in aging populations.

Research during the COVID-19 pandemic highlights the utility of TPRs in supporting remote exercise and care (Addas, 2023), particularly in long-term care settings (Cesta et al., 2016), emphasizing the importance of staff training, family connections, and implementation strategies (Moyle et al., 2020). Other studies explore the educational potential of TPRs for older adults (Hiyama et al., 2017).

Attitudes toward TPRs are generally positive across user groups, including older adults, caregivers, and healthcare providers (Cesta et al., 2016; Shin et al., 2022). Perceived benefits include improved communication, emotional well-being, and independent living (Koceski & Koceska, 2016; Cesta et al., 2012; Reis et al., 2018), although concerns remain about their limitations in providing physical and emotional care (Cesta et al., 2013).

Challenges identified include privacy, security, ethical concerns, technical complexity, and resistance to change (Cesta et al., 2012; Fischedick et al., 2023; Niemelä et al., 2021). Cultural differences also influence acceptance, underlining the need for broader, cross-cultural studies (Gonzalez-Jimenez et al., 2013; Zhang & Hansen, 2022). Usability, trustworthiness, and reliability are critical factors influencing adoption (Aaltonen et al., 2017; Smith et al., 2021), alongside regulatory, licensing, and liability issues (Fiorini et al., 2022).

Positive user experiences are closely linked to trust, ease of use, and prior exposure to the technology (Michaud et al., 2007). In addition to facilitating social interaction, TPRs may serve in roles such as remote health monitoring, telemedicine, cleaning, and disinfection, potentially enhancing care quality and infection control (Fiorini et al., 2022). However, successful integration requires addressing technical, ethical, and user-centered challenges (Niemelä et al., 2021).

RQ2: What are the primary applications of TPRs in elderly care, and how effective are these applications in enhancing older adults' overall well-being?

Research consistently underscores the potential of TPRs to enhance older adults' well-being by fostering social connectedness, enabling remote care, and promoting autonomy. While outcomes vary due to contextual and methodological differences, several key applications and benefits emerge:

1) *Remote monitoring and assistance:* TPRs allow caregivers and healthcare professionals to monitor older adults remotely, provide timely interventions, and enhance safety - facilitating early detection of health issues and reducing caregiver burden (Beraldo et al., 2018; Fiorini et al., 2022; Zhang et al., 2018).
2) *Social connectedness*: TPRs support virtual interactions with family, caregivers, and healthcare providers, helping to alleviate loneliness and enhance social engagement (Addas, 2023; Moyle et al., 2019; Ragno et al., 2023).
3) *Quality of life*: By supporting independence, emotional well-being, and continuous communication, TPRs improve quality of life (Beraldo et al., 2021; Koceski & Koceska, 2016).
4) *Cognitive support:* Some evidence suggests cognitive benefits, particularly for individuals with dementia, through increased mental stimulation and engagement (Michaud et al., 2007; Moyle et al., 2019).
5) *Health outcomes:* TPRs support virtual consultations, medication adherence, and mental health interventions, contributing to better health outcomes (Cortellessa et al., 2017; Sorrentino et al., 2020).
6) *Physical activity:* TPRs can guide older adults in exercise and rehabilitation routines, promoting mobility, strength, and physical wellness (Almeida et al., 2022; De Benedictis et al., 2022).
7) *Cost-effectiveness:* By reducing reliance on continuous in-person care, TPRs may help lower healthcare costs and improve resource efficiency (Ragno et al., 2023).
8) *Privacy and security:* The use of TPRs raises significant privacy and security issues, necessitating the implementation of strong technical safeguards (Hung et al., 2022).

In sum, TPRs offer strong potential to enhance older adults' well-being, though broader adoption depends on usability, ethical safeguards, and long-term evaluation.

RQ3: What are the key technological features of TPRs used in elderly care, and how do these features influence user acceptance and engagement?

Telepresence robotics is rapidly advancing, driven by developments in AI, machine learning, and hardware design. In elderly care, effective TPRs incorporate features such as audiovisual communication, intuitive user interfaces, remote control, obstacle detection, autonomous docking, adjustable height, and customizable functionality to accommodate diverse user needs (Ballesteros et al., 2019; Casiddu et al., 2015; Loza-Matovelle et al., 2019; Sorrentino et al., 2020).

Recent innovations focus on improving accuracy, usability, reliability, and adaptability to domestic

environments. Enhanced obstacle detection, floor pattern recognition, and improved maneuverability are central to optimizing performance. Addressing sensor errors and physical limitations remains a key challenge (Casiddu et al., 2015; González-Jiménez et al., 2012; Sorrentino et al., 2020).

User acceptance is closely linked to the perceived ease of use and reliability. Complex or unintuitive interfaces, high latency, and poor connectivity contribute to user frustration and decreased engagement (Loza-Matovelle et al., 2019). Hybrid control models have been introduced, integrating caregiver or software support while preserving user autonomy (Koceska et al., 2019). Features such as simplified interfaces, large buttons, and voice commands enhance usability for users with sensory, cognitive, or mobility impairments (Casiddu et al., 2015; Ballesteros et al., 2019).

Training and support are crucial for adoption, especially among less tech-savvy older adults; simple interfaces boost satisfaction and engagement (Kiselev et al., 2015).

Semi-autonomous interaction and robotic-mediated communication represent promising approaches, particularly in clinical and long-term care settings. When thoughtfully designed, TPRs can effectively enhance quality of life by supporting communication, independence, and emotional well-being.

RQ4: What empirical evidence exists regarding the impact of TPRs interventions on cognitive functioning, emotional well-being, and quality of life among older adults?

Evidence on the impact of TPRs on cognitive functioning, emotional wellbeing, and quality of life in older adults remains limited and methodologically inconsistent. Nonetheless, existing studies report generally positive outcomes (Cesta et al., 2016; Koceski & Koceska, 2016; Michaud et al., 2007).

TPR interventions have been associated with improvements in attention, memory, problem-solving, and executive function through cognitively stimulating activities, social engagement, mental exercises, and access to educational resources (Cesta, 2013; Beraldo et al., 2020; Hiyama et al., 2017; Chen et al., 2013; Koceski & Koceska, 2016).

Studies suggest TPRs can enhance mood, reduce anxiety, and alleviate loneliness by facilitating virtual social interaction, maintaining social ties, and providing companionship (Cesta et al., 2016; Cortellessa, 2017; Reis, 2018; Zhang, 2018).

TPRs support daily activities, education, communication, and inclusion, contributing to improved quality of life and high user satisfaction (Hiyama et al., 2017; Narasimha, 2017). Positive effects have also been observed in individuals with mild cognitive impairment or dementia (Tiberio et al., 2013; Boman & Bartfai, 2015).

Emerging approaches, such as shared-control systems, further aim to enhance the effectiveness of TPRs in supporting well-being among older adults with cognitive challenges (Beraldo, 2021).

RQ5: What critical research gaps remain in the study of TPRs in elderly care, and what opportunities do these gaps present for future investigation and innovation?

Current research on TPRs in elderly care reveals key gaps and future directions:

1. Rigorous studies are needed on TPRs' impact on cognition, emotional well-being, and quality of life (Addas, 2023; Smith et al., 2021).
2. Financial viability, scalability, and long-term sustainability require further evaluation (Boissy et al., 2007; Moyle et al., 2020).
3. Cultural, ethical, and privacy factors affecting acceptance remain underexplored (Hung et al., 2022; Mascret & Temprado, 2023).
4. Research should focus on personalization for specific conditions and disabilities (De Benedictis et al., 2022).
5. User-centered design and interface features need further investigation (Fiorini et al., 2020).
6. Enhancing social presence and emotional responsiveness is essential for user trust (Fraune et al., 2022).
7. Technical improvements in navigation, durability, and adaptability are needed (Mascret & Temprado, 2023).
8. Longitudinal studies should assess long-term effects on well-being (Winterstein et al., 2021).
9. Integration into existing care systems requires addressing operational and regulatory barriers (Hiyama et al., 2017).

Addressing these gaps will inform theory, guide implementation, and support evidence-based innovation in TPR-enabled eldercare.

4 Persona-based findings

The following subsections provide an account of the personas derived from interviews with caregivers (Section 4.1), medical professionals (Section 4.2), and elderly individuals (Section 4.3).

4.1 Caregivers

Based on semi-structured interviews conducted with 15 caregivers, six distinct personas were identified to reflect the diverse attitudes, experiences, and expectations regarding the integration of TPRs in elderly care settings. These personas vary across care environments, levels of technological familiarity, and motivational drivers, offering a view of the socio-technical dynamics that influence TPR adoption.

Each persona captures a representative caregiving profile and serves to highlight differing priorities and barriers associated with robotic integration.

Persona 1: Socially-oriented female caregiver open to TPRs
- *Setting:* Care home
- *Focus:* Enhancing residents' well-being and social interaction through TPRs.
- *Attitude:* Cautiously optimistic, open to training.
- *Challenges:* Limited technological exposure, residents' cognitive decline.
- *Goals:* Reduce social isolation, improve family communication.

Persona 2: Task-oriented female caregiver
- *Setting:* Care home
- Focus: Reducing workload and improving emergency responsiveness through technology.
- *Attitude:* Functional view of TPRs; secondary interest in social use.
- *Challenges:* Time constraints, resistance to technology.
- *Goals:* Streamline monitoring and caregiving efficiency.

Persona 3: Skeptical and tradition-oriented female
- *Setting:* Care home
- *Focus:* Prioritizes personal care and traditional methods.
- *Attitude:* Highly skeptical of TPRs' usefulness.
- *Challenges*: Lack of clarity about robot benefits, concerns about resident safety and ethical issues.
- *Goals:* Maintain dignity and human-centric care.

Persona 4: Human-centered, resistant to change female
- *Setting:* Small care facility
- *Focus:* Preserving direct, in-person interaction.
- *Attitude:* Disinclined to integrate robots; distrust in technology.
- *Challenges:* Space limitations, ethical/privacy concerns, perceived lack of added value.
- *Goals:* Ensure comfort and emotional support via personal interaction.

Persona 5: Tech-positive and experienced male
- *Setting:* Care home
- *Focus:* Enhancing both social and practical caregiving tasks with TPRs.
- *Attitude:* Experienced with and receptive to TPRs.
- *Challenges:* Limited robot functionality, resistance from residents.
- *Goals:* Improve communication, reduce isolation, support staff workload.

Persona 6: Female home caregiver with conditional optimism
- *Setting:* Home care
- *Focus:* Using TPRs to assist in care and reduce elderly isolation.
- *Attitude:* Cautious but sees potential if implemented appropriately.
- *Challenges:* Elderly resistance, infrastructure limitations, privacy concerns.
- *Goals:* Foster emotional bonds, increase safety and independence via remote monitoring.

The persona analysis revealed several common themes that transcend individual caregiving profiles:
- *Motivational drivers.* A shared emphasis on improving resident well-being, maintaining dignity, enhancing caregiving efficiency, and strengthening communication with family members.
- *Key challenges.* Limited technological experience among caregivers, infrastructure constraints, especially in home settings, ethical concerns regarding privacy and autonomy, cognitive impairments among older adults impeding effective TPR use.
- *Desired robot features.* Simplicity, safety, user-friendly interfaces, emotional engagement, and support for both social and practical tasks.
- *Training needs.* Practical, hands-on guidance for both caregivers and residents; ongoing support.

This persona framework offers an understanding of the complex landscape of caregiver attitudes toward TPRs. It serves as a foundational tool for guiding user-centered design, deployment strategies, and training programs tailored to the real-world needs and concerns of caregivers. Recognizing and accommodating these diverse perspectives is essential for ensuring successful and ethical integration of TPRs in elderly care environments.

4.2. Medical professionals

Interviews with 15 medical professionals, including nurses, physiotherapists, internal medicine doctors, and educators, revealed a spectrum of perspectives on the use of TPRs in elderly care. Through a persona-driven analysis, three representative personas were constructed to capture key differences in attitudes, technological familiarity, and readiness for adoption. While many participants shared overlapping motivations and concerns, generational and experiential differences significantly influenced their views on the potential of TPRs in healthcare contexts.

Persona 1: Tech-comfortable female nurse and educator
- *Profile:* Female nurse and educator, 35 years old, with 10+ years in elderly care.
- *Environment:* Hospital, home care, and educational settings.
- *Technological outlook:* Comfortable with technology; limited experience with TPRs.
- *Goals:* Promote emotional and social well-being for elderly patients. Use TPRs to increase access

to care while preserving human connection. Encourage co-development of robots with elderly users for usability and trust.
- *Challenges:* Cognitive limitations in elderly patients hinder remote interaction. Financial and training barriers to adoption.
- *Concerns:* Ethical and privacy risks, especially in dementia care. Navigation in cluttered homes, trust in automation.
- *Desired features:* User-friendly interface, high-quality video/audio, health monitoring tools. AI-driven companionship, customizable robot design and voice. Integration with medical records and emergency support functions.
- *Training needs:* Interactive tutorials for caregivers and patients. Technical support, privacy training, and slow, supported introduction for elderly.
- *Success indicators:* Increased elderly engagement, improved care outcomes, family satisfaction, and national-level trust programs.

Persona 2: Low-tech female physiotherapist and nurse
- *Profile:* Female physiotherapist, aged 56, with 20+ years in rehabilitation and nursing.
- *Environment:* Nursing homes, in-home care, rehab centers.
- *Technological outlook:* Limited technological comfort and no TPR experience.
- *Goals:* Facilitate communication and motivation for elderly clients. Prevent mental isolation through meaningful interactions.
- *Challenges:* Poor digital literacy among elderly patients. Safety risks and cost barriers in TPR integration.
- *Concerns:* Suitability for dementia patients. Risk of injury or distress from unfamiliar or intimidating robots. Data privacy and lack of emergency responsiveness.
- *Desired features:* Compact, intuitive design; safe for small and cluttered spaces. Real-time communication; monitoring of vitals like temperature, blood pressure.
- *Training needs:* Practical workshops for staff. Continuous support and privacy training.
- *Success indicators:* Improved social engagement and reduced loneliness. Positive usability feedback from both elderly and staff.

Persona 3: Senior female physician skeptical of TPRs
- *Profile:* Female doctor, aged 82, with ~50 years in internal and palliative care.
- *Environment:* Nursing and care clinic.
- *Technological outlook:* Very limited; only basic computer and smartphone use. No hands-on experience with TPRs; limited exposure through AI testing in ward.
- *Challenges:* Strongly skeptical of TPR feasibility in elderly care and believes elderly cannot realistically adapt to robots due to cognitive decline.
- *Concerns:* Criticizes impersonality and high cost of TPRs.
- *Desired features:* Functional assistance (e.g., carrying food, delivering packages). Support communication when human staff are unavailable.
- *Training and success metrics:* Unable to specify due to lack of knowledge or experience.

The persona analysis highlighted a number of common motivations, barriers, and design requirements across professional roles and settings:
- *Motivational drivers.* Maintain emotional well-being of elderly patients, increase access to care without compromising human touch, use TPRs as supplementary tools rather than replacements for personal care.
- *Key challenges.* Technological unfamiliarity among elderly users, infrastructure and financial constraints, persistent ethical and privacy concerns.
- *Desired robot features.* Simplicity and ease of use, safe operation in constrained environments, adaptability for cognitively or physically impaired individuals.
- *Implementation needs.* Gradual, staged integration of TPRs into care workflows, hands-on training for both caregivers and care recipients, participatory, user-centered design and continuous technical support.

This analysis underscores the heterogeneous landscape of professional attitudes toward TPRs in elder care. While some healthcare professionals see strong potential for enhancing access and well-being, others, particularly older or more traditionally trained staff, express skepticism rooted in practical, ethical, and generational concerns. These findings highlight the need for tailored strategies that address both readiness and resistance within professional communities, ensuring that TPR implementation is inclusive, ethical, and grounded in real-world care practices.

4.3. Elderly individuals

To explore the diversity in needs, capabilities, and expectations among the elderly population, five personas were developed using thematic analysis of interview data. These personas represent a spectrum of health conditions, social settings, and technological familiarity, offering insight into how TPRs might enhance their well-being and autonomy.

Persona 1: Technologically confident female multigenerational resident
- *Age:* 72

- *Living:* With daughter and grandchildren abroad.
- *Health:* Fair, suffers from high blood pressure; easily fatigued.
- *Tech comfort:* High; daily smartphone and social media user.
- *Goals:* Save time and energy for herself, overcome language barriers.
- *Challenges:* Fatigue from daily chores, unfamiliar cultural context.
- *Concerns:* Privacy, robot safety in household spaces, uncanny appearance.
- *Desired features:* Multilingual support, household assistance, translation.
- *Training:* In-person demonstration preferred, expects regular updates.

Persona 2: Independent but socially isolated female

- *Age:* 74 |
- *Living:* Alone; widowed.
- *Health:* Good mobility, hearing and vision impairments. uses aids.
- *Tech comfort:* Moderate, learned from grandchildren.
- *Goals:* Reduce loneliness, feel closer to family.
- *Challenges:* Distrust of online privacy and data security.
- *Desired features:* Simple interface, reminders, video calling.
- *Training:* Hands-on with family present, 24/7 phone support preferred.

Persona 3: Autonomous but technologically hesitant male

- *Age:* 65
- *Living:* Alone in suburbia, limited support network.
- *Health:* Mild hypertension and back pain.
- *Tech comfort:* Basic, uses smartphone, unfamiliar with assistive tech.
- *Goals:* Maintain autonomy, improve family contact, manage health tasks.
- *Challenges:* Tech complexity, privacy concerns, dependency fears.
- *Desired features:* Voice control, health monitoring, entertainment.
- *Training:* Short tutorials with caregivers, automated guidance prompts.

Persona 4: Enthusiastic but physically dependent female

- *Age:* 75
- *Living:* On a farm with extended family.
- *Health:* Wheelchair user due to osteoarthritis.
- *Tech comfort:* High, very open to technology.
- *Goals:* Lighten the family's caregiving burden, increase self-reliance.
- *Challenges:* Severe mobility limitations.
- *Desired features:* Voice functions, vital sign monitoring, entertainment.
- *Training:* In-home, family-supported setup and training.

Persona 5: Independent female rural dweller with entrepreneurial spirit

- *Age:* 75
- *Living:* In the countryside with spouse.
- *Health:* Arthritis; walks with cane.
- *Tech comfort:* Very open, uses smartphone for voice/video calls.
- *Goals:* Reduce rural isolation, support daily tasks without reliance.
- *Challenges:* Loneliness from family migration, security concerns.
- *Desired features:* Gas detectors, audio reminders, video call screen.
- *Training:* In-person guidance, simple video tutorials, continuous support.

Across all personas, several shared themes emerged:

- *Motivational drivers.* Reduce loneliness and increase family interaction, maintain independence while managing health, gain support without replacing human relationships.
- *Key challenges.* Data privacy and cybersecurity, complexity of technology use, reliability and safety of robots in physical environments.
- *Desired robot features.* User-friendly interfaces with voice control, video calling, entertainment, and health monitoring, emergency alert systems and task reminders.
- *Training needs.* In-person, hands-on instruction with family or caregivers, ongoing support via telephone or regular visits, easy-to-follow user guides and troubleshooting aids.
- *Indicators of Success.* Improved communication and emotional connection, better health self-management, increased comfort and confidence using the robot.

These five personas reflect a multifaceted elderly population characterized by differences in digital literacy, health status, physical mobility, and social engagement. The findings underscore the need for adaptive, inclusive, and empathetic design principles in the development of TPRs for eldercare. The persona framework serves as a practical foundation for guiding user-centered innovation, deployment strategies, and policymaking that respects the diversity of older adult experiences and expectations.

5 Discussion

This study explored the potential of TPRs to support elderly individuals in Estonia by combining a systematic literature review with qualitative interviews analysed through a persona-driven framework. The findings reveal both promising opportunities and persistent challenges associated with TPR adoption across stakeholder groups - caregivers, medical professionals, and older adults - underscoring the complexity of integrating socially assistive technologies into eldercare systems.

5.1 Alignment with existing literature

The literature review confirmed that TPRs offer substantial potential to enhance social connectedness, facilitate remote monitoring, and promote autonomy among older adults (Moyle et al., 2014; Beraldo et al., 2018; Koceski & Koceska, 2016). These benefits were echoed in the interview findings, particularly among tech-positive caregivers and enthusiastic elderly users, who viewed TPRs as a way to reduce social isolation and improve quality of life. Similarly, the emphasis on usability, privacy, and reliability in the literature was reinforced by all interviewee groups, suggesting that these concerns remain central to successful deployment (Hung et al., 2022; Sorrentino et al., 2020).

The literature also highlighted that user engagement and acceptance are closely tied to interface simplicity and adaptability (Casiddu et al., 2015), a point echoed in the personas developed in this study. Whether due to sensory impairments, cognitive decline, or limited digital literacy, many elderly participants favoured voice-controlled interfaces, large buttons, and customizable features that could be tailored to their needs.

5.2 Divergence across stakeholder groups

While all groups acknowledged the potential value of TPRs, attitudes toward implementation varied markedly.

Caregivers showed a spectrum of views, ranging from optimism about workload relief and enhanced communication to deep skepticism, particularly among those committed to traditional, human-centered care models. Many stressed the importance of hands-on training, and some were concerned about the emotional and ethical implications of replacing human interaction with robotic presence.

Medical professionals were more divided. Younger participants, especially those working across education and hospital settings, tended to be more open to co-developing solutions with older adults. In contrast, older clinicians expressed strong skepticism, often citing cognitive limitations, ethical concerns, or lack of trust in automation.

Elderly individuals displayed perhaps the greatest variation. While some were eager to engage with TPRs to ease household burdens, enhance safety, or stay socially connected, others expressed concerns over data privacy, technical complexity, and dependency fears. Notably, those living in rural or socially isolated contexts expressed a higher willingness to adopt TPRs if the tools aligned with their lifestyle needs and provided practical assistance.

This divergence suggests that TPR integration cannot follow a one-size-fits-all approach. Instead, design and implementation strategies must be grounded in personalization, trust-building, and contextual sensitivity.

5.3 Barriers to adoption

The study identified a consistent set of barriers across stakeholder groups:

- Technological unfamiliarity and resistance, especially among older caregivers and medical professionals.
- Ethical concerns related to privacy, autonomy, and the impersonality of robotic interactions.
- Infrastructure and financial limitations, particularly in under-resourced home and rural settings.
- Lack of training and ongoing support, leading to uncertainty and underutilization of features.

These findings align with previous work (Fischedick et al., 2023; Niemelä et al., 2021) and underscore the need for robust policy frameworks, equitable funding mechanisms, and targeted educational programs that support both staff and elderly users in onboarding and integrating TPRs.

5.4 Implications for design and policy

The persona-driven analysis adds depth to the growing body of user-centered robotics research. By capturing individual variations in motivations, challenges, and feature preferences, the study supports several design and policy recommendations:

- Design for inclusivity. Interfaces must accommodate diverse abilities, from tech-savvy older adults to those with visual, cognitive, or motor impairments.
- Support autonomy without replacing human care: TPRs should augment, not replace, human relationships. This principle is particularly important in building trust among skeptical users.
- Embed empathy into interface behavior. Features such as AI-driven companionship, personalized voice options, and emotional responsiveness could enhance perceived warmth and acceptance.
- Ensure flexible training and support. Onboarding should be gradual, hands-on, and tailored. Ongoing access to technical help and family support is vital to long-term adoption.

- Develop context-aware strategies. Urban, rural, institutional, and home settings each require unique infrastructure and integration plans. A national strategy for TPR deployment in Estonia should reflect these contextual nuances.

6 Conclusion

This study demonstrates that telepresence robots hold transformative potential for supporting aging in place, enhancing caregiver efficiency, and improving social and emotional well-being among older adults. However, realizing this potential depends on deep contextual understanding, inclusive design, and ethical, user-centered implementation. Estonia's leadership in digital innovation provides a strong foundation, but attention to social care systems, equity, and stakeholder diversity will be essential in translating technological capability into meaningful impact.

The study also has limitations. The sample was geographically and demographically limited to the Estonian context, which may not generalize globally. Interviews, while in-depth, relied on hypothetical perceptions of TPRs rather than long-term use. Further quantitative validation and longitudinal field studies are needed to assess real-world impact over time.

Several avenues for future research emerge:

- Cross-cultural comparative studies to explore how sociocultural factors influence TPR acceptance.
- Longitudinal trials of TPR use in home and institutional settings to evaluate long-term effects on health, cognition, and emotional well-being.
- AI-enhanced personalization studies to tailor TPR behaviors and responses to individual needs.
- Cost-benefit analyses to inform funding decisions and sustainable implementation models.
- Co-design approaches involving older adults in iterative TPR development cycles.

Acknowledgments

This paper is based on empirical research conducted at Tallinn University as part of the LIFE (*Learning in Interdisciplinary Focused Environment*) initiative titled "Enhancing Elderly Care Through Telepresence Robots I" in spring 2024 and "Enhancing Elderly Care Through Telepresence Robots II" in autumn 2024.

References

Aaltonen, I., Niemelä, M., & Tammela, A. (2017). Please call me? Calling practices with telepresence robots for the elderly. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 55-56).

Addas, A. (2023). Telepresence robots as facilitators of physical exercise during COVID-19: a feasibility and acceptance study. *Frontiers in Public Health*, *11*, 1277479.

Almeida, L., Menezes, P., & Dias, J. (2022). Telepresence social robotics towards co-presence: A review. *Applied Sciences*, *12*(11), 5557.

Arthanat, S. (2021). Promoting information communication technology adoption and acceptance for aging-in-place: a randomized controlled trial. *Journal of Applied Gerontology*, 40 (5), 471-480.

Ballesteros, J., Tudela, A., Caro-Romero, J. R., & Urdiales, C. (2019). A cane-based low cost sensor to implement attention mechanisms in telecare robots. In *2019 International Conference on Robotics and Automation (ICRA)* (pp. 1473-1478). IEEE.

Beraldo, G., Antonello, M., Cimolato, A., Menegatti, E., & Tonin, L. (2018). Brain-computer interface meets ROS: a robotic approach to mentally drive telepresence robots. In *2018 IEEE international conference on robotics and automation (ICRA)* (pp. 4459-4464). IEEE.

Beraldo, G., De Benedictis, R., Devaram, R. R., Cesta, A., & Cortellessa, G. (2021). The Role of Intelligent Telepresence Robots for Continuously Caring Elderly People at Home. In *AIRO@ AI* IA* (pp. 8-14).

Beraldo, G., Tonin, L., Cesta, A., & Menegatti, E. (2020). Shared Approaches to Mentally Drive Telepresence Robots. In *AIRO@ AI* IA* (pp. 22-27).

Boissy, P., Corriveau, H., Michaud, F., Labonté, D., & Royer, M. P. (2007). A qualitative study of in-home robotic telepresence for home care of community-living elderly subjects. *Journal of telemedicine and telecare*, *13*(2), 79-84.

Boman, I. L., & Bartfai, A. (2015). The first step in using a robot in brain injury rehabilitation: patients' and health-care professionals' perspective. *Disability and Rehabilitation: Assistive Technology*, *10*(5), 365-370.

Casiddu, N., Cesta, A., Cortellessa, G., Orlandini, A., Porfirione, C., Divano, A., ... & Zallio, M. (2015). Robot interface design: The giraff telepresence robot for social interaction. In *Ambient Assisted Living: Italian Forum 2014* (pp. 499-509). Springer International Publishing.

Cesta, A., Cortellessa, G., Orlandini, A., & Tiberio, L. (2012). Addressing the Long-term Evaluation of a Telepresence Robot for the Elderly. In *ICAART (1)* (pp. 652-663).

Cesta, A., Cortellessa, G., Orlandini, A., & Tiberio, L. (2013). Evaluating telepresence robots in the field. In *Agents and Artificial Intelligence: 4th International Conference, ICAART 2012, Vilamoura, Portugal, February 6-8, 2012. Revised Selected Papers 4* (pp. 433-448). Springer Berlin Heidelberg.

Cesta, A., Cortellessa, G., Orlandini, A., & Tiberio, L. (2016). Long-term evaluation of a telepresence robot for the elderly: methodology and ecological case study. *International Journal of Social Robotics*, *8*, 421-441.

Chen, Y. S., Lu, J. M., & Hsu, Y. L. (2013). Design and evaluation of a telepresence robot for interpersonal communication with older adults. In *Inclusive Society: Health and Wellbeing in the Community, and Care at Home: 11th International Conference on Smart Homes and Health Telematics, ICOST 2013, Singapore, June 19-21, 2013. Proceedings 11* (pp. 298-303). Springer Berlin Heidelberg.

Chu, L., Fung, H. H., Tse, D. C., Tsang, V. H., Zhang, H., & Mai, C. (2021). Obtaining information from different sources matters during the COVID-19 pandemic. *The Gerontologist*, *61*(2), 187-195.

Cortellessa, G., Fracasso, F., Sorrentino, A., Orlandini, A., Bernardi, G., Coraci, L., ... & Cesta, A. (2017). Enhancing the interactive services of a telepresence robot for AAL: Developments and a psycho-physiological assessment. In *Ambient Assisted Living: Italian Forum 2016 7* (pp. 337-357). Springer International Publishing.

De Benedictis, R., Beraldo, G., Devaram, R. R., Cesta, A., & Cortellessa, G. (2021, December). Enhancing telepresence robots with AI: Combining services to personalize and react. In *International Conference of the Italian Association for Artificial Intelligence* (pp. 32-45). Cham: Springer International Publishing.

EC (2021). European Commission's digital Economy and society index (DESI). Retrieved from https://digital-strategy.ec.europa.eu/en/policies/desi

European Commission (2022). European Commission's digital economy and society index (DESI). Retrieved from https://digital-strategy.ec.europa.eu/en/policies/desi

Fiorini, L., Mancioppi, G., Becchimanzi, C., Sorrentino, A., Pistolesi, M., Tosi, F., & Cavallo, F. (2020). Multidimensional evaluation of telepresence robot: results from a field trial. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 1211-1216). IEEE.

Fiorini, L., Rovini, E., Russo, S., Toccafondi, L., D'Onofrio, G., Cornacchia Loizzo, F. G., ... & Cavallo, F. (2022a). On the use of assistive technology during the COVID-19 outbreak: results and lessons learned from pilot studies. *Sensors*, *22*(17), 6631.

Fischedick, S. B., Richter, K., Wengefeld, T., Seichter, D., Scheidig, A., Doering, N., ... & Gross, H. M. (2023, September). Bridging Distance with a Collaborative Telepresence Robot for Older Adults–Report on Progress in the CO-HUMANICS Project. In *ISR Europe 2023; 56th International Symposium on Robotics* (pp. 346-353). VDE.

Fraune, M. R., Komatsu, T., Preusse, H. R., Langlois, D. K., Au, R. H., Ling, K., ... & Tsui, K. M. (2022). Socially facilitative robots for older adults to alleviate social isolation: A participatory design workshop approach in the US and Japan. *Frontiers in psychology*, *13*, 904019.

González-Jiménez, J., Galindo, C., & Ruiz-Sarmiento, J. R. (2012). Technical improvements of the Giraff telepresence robot based on users' evaluation. In *2012 IEEE RO-MAN: The 21st IEEE international symposium on robot and human interactive communication* (pp. 827-832). IEEE.

Hiyama, A., Kosugi, A., Fukuda, K., Kobayashi, M., & Hirose, M. (2017). Facilitating remote communication between senior communities with telepresence robots. In *Human Aspects of IT for the Aged Population. Applications, Services and Contexts: Third International Conference, ITAP 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part II 3* (pp. 501-515). Springer International Publishing.

Hung, L., Wong, J., Smith, C., Berndt, A., Gregorio, M., Horne, N., ... & Young, E. (2022). Facilitators and barriers to using telepresence robots in aged care settings: a scoping review. *Journal of Rehabilitation and Assistive Technologies Engineering*, *9*, 20556683211072385.

Kiselev, A., Kristoffersson, A., & Loutfi, A. (2015). Combining semi-autonomous navigation with manned behaviour in a cooperative driving system for Mobile robotic telepresence. In *Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part IV 13* (pp. 17-28). Springer International Publishing.

Koceska, N., Koceski, S., Beomonte Zobel, P., Trajkovik, V., & Garcia, N. (2019). A telemedicine robot system for assisted and independent living. *Sensors*, *19*(4), 834.

Koceski, S., & Koceska, N. (2016). Evaluation of an assistive telepresence robot for elderly healthcare. *Journal of Medical Systems*, *40*, 1-7.

Leppiman, A., Riivits-Arkonsuo, I. & Pohjola, A. (2021). Old-age digital exclusion as a policy challenge in Estonia and Finland. In Walsh, K.,

Scharf, T., Van Regenmortel, S. and Wanka, A. (Eds), Social Exclusion in Later Life, Springer International Publishing, Cham, 28, 409-419.

Loza-Matovelle, D., Verdugo, A., Zalama, E., & Gómez-García-Bermejo, J. (2019). An architecture for the integration of robots and sensors for the care of the elderly in an ambient assisted living environment. *Robotics*, *8*(3), 76.

Mascret, N., & Temprado, J. J. (2023). Acceptance of a Mobile Telepresence Robot, before Use, to Remotely Supervise Older Adults' Adapted Physical Activity. *International Journal of Environmental Research and Public Health*, *20*(4), 3012.

Michaud, F., Boissy, P., Labonte, D., Corriveau, H., Grant, A., Lauria, M., ... & Royer, M. P. (2007). Telepresence Robot for Home Care Assistance. In *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics* (pp. 50-55).

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. and The PRISMA Group (2009). Preferred reporting items for systematic reviews and Meta-Analyses: the PRISMA statement. *PLoS Medicine*, *6*(7), e1000097, doi: 10.1371/journal.pmed.1000097.

Moyle, W., Jones, C., & Sung, B. (2020). Telepresence robots: Encouraging interactive communication between family carers and people with dementia. *Australasian Journal on Ageing*, *39*(1), e127-e133.

Moyle, W., Jones, C., Cooke, M., O'Dwyer, S., Sung, B., & Drummond, S. (2014). Connecting the person with dementia and family: a feasibility study of a telepresence robot. *BMC Geriatrics*, *14*(1), 1-11.

Moyle, W., Jones, C., Dwan, T., Ownsworth, T., & Sung, B. (2019). Using telepresence for social connection: views of older people with dementia, families, and health professionals from a mixed methods pilot study. *Aging & Mental Health*, *23*(12), 1643-1650.

Narasimha, S., Madathil, K. C., Agnisarman, S., Rogers, H., Welch, B., Ashok, A., ... & McElligott, J. (2017). Designing telemedicine systems for geriatric patients: a review of the usability studies. *Telemedicine and e-Health*, *23*(6), 459-472.

Niemelä, M., Van Aerschot, L., Tammela, A., Aaltonen, I., & Lammi, H. (2021). Towards ethical guidelines of using telepresence robots in residential care. *International Journal of Social Robotics*, *13*, 431-439.

Nimrod, G. (2020). Aging well in the digital age: technology in processes of selective optimization with compensation. The Journals of Gerontology: Series B, 75 (9), 2008-2017.

Ojaperv, K., & Virkus, S. (2023). Pregnancy-related health information behaviour of Estonian women. *Global Knowledge, Memory and Communication*, 72(3), 284–314. https://doi.org/10.1108/GKMC-05-2021-0086

Paimre, M., Virkus, S., & Osula, K. (2024). How Technology, Health Information Seeking, and Socioeconomic Factors Are Associated With Coronavirus Disease 2019 Vaccination Readiness in Estonians Over 50 Years?. *Health Education & Behavior,* 51(4), 502-511.

Paimre, M., Virkus, S., & Osula, K. (2023). Health information behavior and related factors among Estonians aged≥ 50 years during the COVID-19 pandemic. *Journal of Documentation*, *79*(5), 1164-1181.

Ragno, L., Borboni, A., Vannetti, F., Amici, C., & Cusano, N. (2023). Application of social robots in healthcare: Review on characteristics, requirements, technical solutions. *Sensors*, *23*(15), 6820.

Reis, A., Xavier, R., Barroso, I., Monteiro, M. J., Paredes, H., & Barroso, J. (2018). The usage of telepresence robots to support the elderly. In*2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW)* (pp. 1-6). IEEE.

Shin, M. H., McLaren, J., Ramsey, A., Sullivan, J. L., & Moo, L. (2022). Improving a mobile telepresence robot for people with Alzheimer disease and related dementias: semi-structured interviews with stakeholders. *JMIR Aging*, *5*(2), e32322.

Smith, C., Gregorio, M., & Hung, L. (2021). Facilitators and barriers to using telepresence robots in aged care settings: a scoping review protocol. *BMJ Open*, *11*(9), e051769.

Sorrentino, A., Radi, L., Cavallo, F., Becchimanzi, C., Pistolesi, M., Tosi, F., ... & Fiorini, L. (2020). Design and Development of a Telepresence and Monitoring Service to Empower the Older Adults. In *Italian Forum of Ambient Assisted Living* (pp. 18-36). Cham: Springer International Publishing.

Tiberio, L., Cesta, A., & Olivetti Belardinelli, M. (2013). Psychophysiological methods to evaluate user's response in human robot interaction: a review and feasibility study. *Robotics*, *2*(2), 92-121.

Tiberio, L., Cesta, A., Cortellessa, G., Padua, L., & Pellegrino, A. R. (2012). Assessing affective response of older users to a telepresence robot using a combination of psychophysiological measures. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication* (pp. 833-838). IEEE.

Virkus, S. (2025). Telepresence Robots in the GLAM sector: Bridging Divides for an Inclusive and Equitable Smart Society. In B. Holland (Ed.) Handbook of trends and innovations concerning library and information science. De Gruyter, https://doi.org/10.1515/9783111443003-007

Virkus, S., Leoste, J., Marmor, K., Kasuk, T., & Talisainen, A. (2023). Telepresence robots from the perspective of psychology and educational sciences. *Information and Learning Sciences*, 124(1/2), 48–69. https://doi.org/10.1108/ILS-09-2022-0106

WHO (2015). World report on ageing and health, World Health Organization. Retrieved from https://apps.who.int/iris/handle/10665/186463

Winterstein, K., Keller, L., Huffstadt, K., & Müller, N. H. (2021, July). Acceptance of social and telepresence robot assistance in German households. In *International Conference on Human-Computer Interaction* (pp. 326-339). Cham: Springer International Publishing.

Zhang, G., & Hansen, J. P. (2022). Telepresence robots for people with special needs: A systematic review. *International Journal of Human–Computer Interaction*, *38*(17), 1651-1667.

Zhang, M., Duan, P., Zhang, Z., & Esche, S. (2018). Development of telepresence teaching robots with social capabilities. In *ASME International Mechanical Engineering Congress and Exposition* (Vol. 52064, p. V005T07A017). American Society of Mechanical Engineers.

Dynamic Lane Reversal Review

Mihael Radovan
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
mradovan@foi.hr

***Abstract**. Traditional traffic control methods based on fixed rules often fall short in addressing the variable nature of urban traffic patterns. Dynamic lane reversal systems, which enable reversible traffic lanes that change direction based on real-time conditions, have been implemented worldwide with varying levels of automation and technological integration. This review covers the evolution of lane reversal techniques over the past two decades, highlighting advancements in simulation modeling, artificial intelligence, and real-time control systems. By analyzing 35 relevant research papers published between 2005 and 2025, the paper identifies key simulation tools, methodologies, and use cases, including evacuation scenarios and autonomous vehicle integration. The findings reveal significant improvements in traffic efficiency and safety through dynamic lane reversal while underscoring the need for further research on AI-driven approaches and practical deployments in smart city environments. The review concludes by outlining research gaps and future directions for enhancing adaptive traffic management strategies through innovative technologies.*

Keywords. Lane Reversal, Tidal Flow, Reversible Traffic Lanes, Road traffic management, Artificial Intelligence, Road traffic simulation software

1 Introduction

A complex problem, such as road traffic control, has been studied for many decades. Traditional traffic control methods relied on fixed rules and algorithms, which are inflexible and inefficient due to changing traffic conditions. Traffic flow is affected by a variety of factors, including the number of vehicles on the road, the size and speed of vehicles, weather conditions, time of the day, day of the week etc.

In modern times, especially big and industrial cities experience the phenomenon of daily migration of the working population, which lives outside of these cities. During morning hours, a huge number of working people drive from suburban to the city center and cause very heavy traffic, and the situation is similar in the afternoon in opposite direction. Some cities have already found a solution for roads with more than two lanes. They've customized a system where a middle lane, or multiple lanes, can switch directions based on certain conditions. This allows the lanes to be used for traffic going in one direction at one time, and the opposite direction at another time. In his 1972 work, Lathrop (Lathrop, 1972) described this type of solution (Fig. 1).

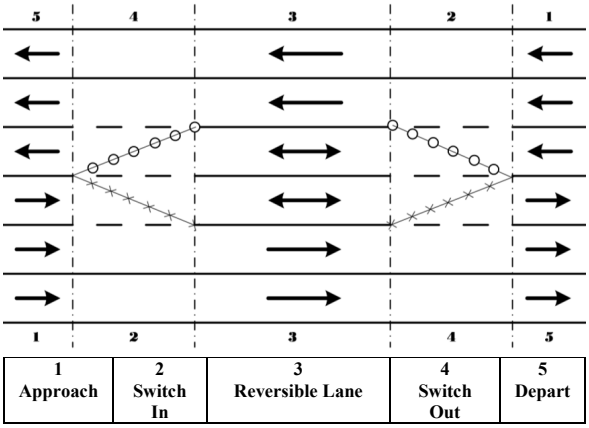


Figure 1. Traffic lane reversal (source: Lathrop 1972)

The Golden Gate, almost 3 km world famous bridge in San Francisco (USA) carries more than 110 000 vehicles per day in both directions over six lanes, according Goldan Gate Bridge Highway and Transportation District[1]. To improve traffic flow, a median barrier between the lanes is moved to make lane distribution by 4-2, 2-4 or 3-3 pattern: on weekday mornings southbound 4 lanes, northbound 2 lanes. Afternoons opposite – southbound 2 lanes, northbound 4 lanes. Weekends and during off-peak periods, each direction has 3 lanes. Until 2015, the median barrier was made by small plastic pylons. Because of inefficiency and uncertainty, in January 2015 it was replaced with the new system – moveable median barrier with a zipper truck.

The Sydney Harbour Bridge (Australia) has 8 lanes, where three lanes in the middle are reversible during peaks. The direction of reversible lanes is indicated by overhead gantries, showing green arrow pointing down to open lane in one direction on one side and static red

[1] www.goldengate.org/bridge/bridge-operations/

„X“ on the other side, indicating the lane is closed for other direction. Flashing red „X“ means the lane is closing but is not yet open for the traffic in opposite direction. This solution has been in use since 1986, when replaced manually moving lane markers operation.

The Lions Gate Bridge (Vancouver, Canada) has three lanes, where middle lane is reversal and controlled by traffic lights.

Around 1.6 km long section on the North Road in Cardiff (UK) has transformation from double two-lane to a three-lane section, where the middle lane is reversal, depending on the time of the date.

There are more use cases in the world, mostly in the USA, Canada, Australia and Europe, and there are implemented different management systems. In the USA they have been used since the 1920s. But most of these systems have one thing in common: they are mostly setup with a fixed time scheme, based on historical data, or controlled manually by control center.

Conventional road traffic management with fixed rules can combine just deterministic factors – time, weather conditions, day in the week – and adjust the working mode (Almeida, et al., 2023). These changes are also determined and fixed, only in some special cases, additionally equipped with car detectors embedded in the road or mounted on the roadside units (RSU) extend functionality based on the presence and number of the cars on the road.

Artificial intelligence (AI) offers a new approach to traffic control that can be more adaptable and responsive to changing conditions. By using AI to determine density of traffic, recognize number and types of vehicles, recognize or predict traffic and weather conditions, optimize traffic signal timing, reroute vehicles and inform drivers about conditions and changes on the route, it is possible to create a more efficient and safer transportation system. As AI technology develops continuously, it is expected to get even more innovative and effective ways to use AI to improve road transportation (Sharma, Liu, Honggang, & Shelley, 2017).

To better understand road traffic problems in urban areas, and to evaluate developed solutions for these problems, it is necessary to use simulation models and simulation software tools. Road traffic simulation software is a powerful tool that can be used to study the behavior of traffic systems (Ejercito, Nebrija, Feria, & Lara-Figueroa, 2017). It can be used for next purposes:

- predict traffic flow under different conditions, such as different traffic volumes, weather conditions, and road construction projects, and this can be used to plan for future traffic demand and to identify potential bottlenecks;
- identify bottlenecks in the traffic network, what can be used to improve traffic flow by implementing some traffic management strategies;
- adding lanes, changing traffic signal timing, or installing traffic-calming devices;
- test the effectiveness of different traffic management strategies, to choose the most effective strategy for a particular situation.

While the issue has been examined extensively in professional and scientific circles for many decades (notably exemplified by Lathrop's 1972 study (Lathrop, 1972)), this paper offers an analysis of the field's progression over the past 20 years, specifically from 2005 to 2024. It will help to identify the most used road traffic simulation environments and technologies for dynamic lane reversal as well as research gaps in the field, which could be applied in the next research. Next section describes methodology used to get and analyze materials and data regarding selected topic – dynamic lane reversal. Section III shows detailed analysis of selected papers – use cases covered with these papers, methodologies and techniques used in these research papers and their results. Section IV discusses the state of the art in the field, research gaps and potential for next research. Section V is the conclusion of the paper. In the end of the paper, there is list of used sources for this research.

2 Methodology

Essential part of any research is a literature review. In this research was used scoping review (Peters, i dr., 2021) - a type of literature review that aims to map the existing literature on a particular topic, with the goal to provide historical content, summary findings and identify gaps. Detected research gaps open avenues for future research which consequently, if pursued, can contribute to advancing the field of study.

To carry out a literature review, Rowley et al. proposed a five-step methodology (Rowley & Slack, 2004). This methodology was adopted to 7-step procedure (Kamble, Gunasekaran, & Gawankar, 2018) (Astarita, Giofre, Guido, & Vitale, 2020). This methodology was frequently used in systematic literature reviews, and in this paper was improved into the 8-step procedure, as presented in Fig. 2.

2.1 Keyword candidates

For the first, there were selected few keywords as keyword candidates, which best cover topic “dynamically reversible traffic lanes” - “dynamic lane reversal” is commonly used word in American English, as well as “tidal flow” in British English. So, three basic keywords are “lane reversal”, “reversible lanes” and “tidal flow”.

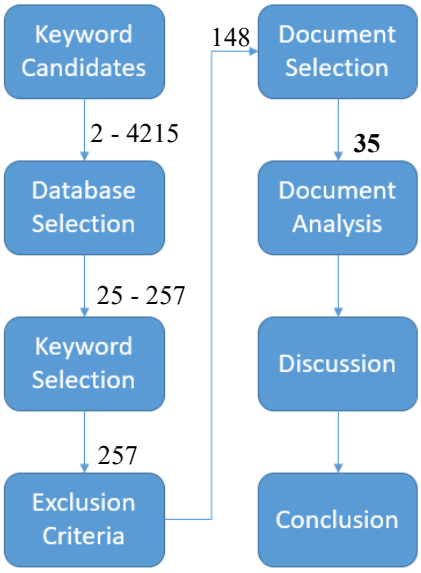


Figure 2. 8-steps procedure for the review

Primary scope of the paper is dynamic lane reversal, but the number of papers on this keyword is too low, then it was reduced to "lane reversal". Parallel, number of papers on synonym keyword, "tidal flow", was too big, because it refers not just to road traffic. On the contrary, keyword "tidal flow lanes" gives extremely low number of papers. Compromising solution is combined keyword "tidal flow" AND lanes". There are also keyword "reversible lanes". In the end, the list of keywords and keyword combinations was used for first step in the procedure:

1. "dynamic lane reversal"
2. "lane reversal"
3. "reversible lanes"
4. "tidal flow lanes"
5. "tidal flow"
6. "tidal flow" AND lanes
7. ("lane reversal" OR "dynamic lane reversal) OR "reversible lanes" OR ("tidal flow" AND lanes)

These queries were used for the search within article title, abstract and keywords in selected citation databases in few combinations, to get set of the data enough big to make further steps.

2.2 Database selection

The most used citation databases in the field of engineering and computer science, that includes peer-reviewed journals, are Scopus[2], Web of Science (WoS)[3] and IEEE Xplore[4], whereby most of journals indexed in IEEE Xplore are already covered with Scopus and WoS. The first step of the procedure was executed in Scopus and WoS, with results presented in Table 1.

[2] Elsevier Scopus: www.scopus.com
[3] Clarivate Web of Science: www.webofscience.com

Table 1. Results in the database on queries

Nr.	Keyword / Query	Scopus	WoS	I3E
1.	"dynamic lane reversal"	16	11	2
2.	"lane reversal"	58	29	11
3.	"reversible lanes"	180	32	7
4.	"tidal flow lanes"	6	2	2
5.	"tidal flow"	4215	860	111
6.	"tidal flow" AND lanes	24	7	10
7.	"lane reversal" OR "reversible lanes" OR ("tidal flow" AND lanes)	257	62	25

As shown in the Table 1, keyword Nr. 5 gives too high a number in all three databases, is not reliable, and it was excluded from the next analysis steps. The number of papers given for each keyword are significantly higher in Scopus than in WoS and much higher than in the IEEE Xplore database. Ration between given results for all 6 keywords, excluding keyword Nr. 5, is presented in Fig. 3.

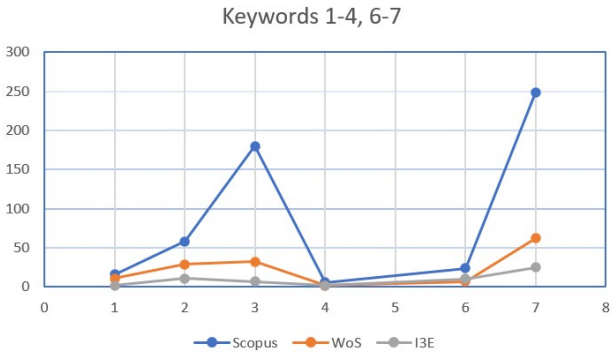


Figure 3. Result ration of results from Table 1

Regarding Astarita et al., WoS includes about 12.000 ISI indexed journals, while Scopus includes more than 25.000 peer-reviewed journals of the most important publishers (Astarita, Giofre, Guido, & Vitale, 2020) (Yong-Hak, n.d.), and for that reason, WoS and IEEE were excluded from further consideration, and Scopus was selected as source database for next steps in the procedure.

2.3 Keyword selection

In section 2.1 is shown list of 7 keyword candidates. Table 1 in section 2.2 shows results for each of three databases, and Scopus was selected as the relevant one. The keyword Nr. 5 was excluded, and in the Table 1 in red line. Keyword Nr. 2 comprehends the keyword Nr.

[4] IEEE Xplore: ieeexplore.ieee.org/Xplore/home.jsp

1, as well as keyword Nr. 6 comprehends keyword Nr. 4. Keyword Nr. 3 is also relevant. These 3 keywords are in the green line in the Table 1, and in the end, all together covered with the keyword Nr. 7 – yellow row in the Table 1.

2.4 Excluding criteria

Keyword Nr. 7 from Table 1 gave 257 results in the Scopus database. Limiting document type to article, conference paper and review, source type to journal and conference proceeding and language to English, result is decreased to 212 papers. There are a lot of very old papers, from the year before 2000, so the search was limited to papers published between the years 2005 and 2025, and result was decreased to 148.

The final query in the Scopus database is:
TITLE-ABS-KEY ("lane reversal" OR "reversible lanes" OR ("tidal flow" AND lanes)) AND PUBYEAR > 2004 AND PUBYEAR < 2024 AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "re")) AND (LIMIT-TO (SRCTYPE , "j") OR LIMIT-TO (SRCTYPE , "p")) AND (LIMIT-TO (LANGUAGE , "English")).

2.5 Document selection

After database selection and keyword selection, search with selected keyword in Scopus database gave us result list of 257 papers. Next step was fine query tuning, after that we got the list of 148 papers. These steps are explained in topics 2.2, 2.3 and 2.4. The list of 130 papers was filtered based on their relevance to the topic, by manually reading the title and abstract. The final list contains 35 papers published from 2006 until 2025 and is presented in Table 2.

Fig. 4 shows a complete 5-step process of paper selection, where numbers present the number of papers given as a result after each step. The papers in the Table 2 are published in last 20 years, and all are related with the topic “lane reversal”. The papers in the table are sorted by years ascending. Older papers present use cases with conventional management, new implementation of modern IT technology, and the newest also possibility of using artificial intelligence algorithms. Thus, it is possible to get an insight into the development of technologies used in the topic field through the timeline.

Some of these papers are used to describe particular use experience with road traffic simulation systems, some describe theoretical investigation of the problem (study) and some mathematical models of solution. These three groups are colored depending on purpose: urban traffic optimization – green, traffic management in case of evacuation – yellow, application of autonomous vehicles – orange, general purpose – gray. Papers presented the real implementation of lane reversal just for special occasions, such as sport or business happenings or in case of natural disasters, e.g. hurricanes, and on the table are colored yellow**.** Papers showing simulation, study or model description of real environment in some cities are colored green. Papers describing cases with autonomous vehicles are colored light orange, and papers describing general purpose are colored grey (s. Legend below). The third column is marked with the initials of the simulation tool used in the paper (s. Table 3).

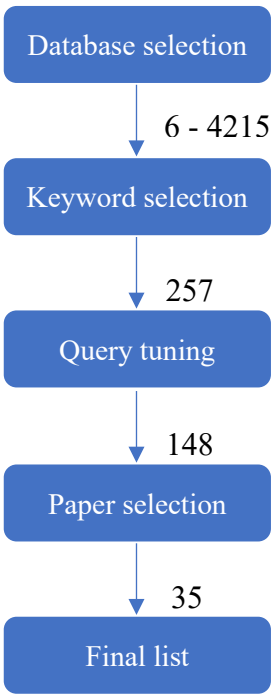


Figure 4. Document selection process

Table 2. Final list of selected papers

Nr.	Reference	Simul.	Study	Model
1	Handke2006 (Handke, 2006)	VS		
2	Logan2006 (Logan, McLeod, & Jordan, 2006)			
3	Wolshon2006 (Wolshon & Lambert, 2006)			
4	Shen2007 (Shen & Sokolowski, 2007)	VS		
5	Williams2007 (Williams, Tagliaferri, Meinhold, Hummer, & Rouphail, 2007)	CS		
6	Kim2008 (Kim, Shekhar, & Min, 2008)			
7	Brown2009 (Brown, White, Van Slyke, & Benson, 2009)			

8	Wu2009 (Wu, Sun, Gao, & Zhang, 2009)			
9	Xie2010 (Xie, Lin, & Travis Waller, 2010)	VT		
10	Lambert2010 (Lambert & Wolshon, 2010)			
11	Bede2010 (Bede, Szabó, & Péter, Optimalization of road traffic with the applied of reversible direction lanes, 2010)			
12	Edara2010 (Edara, Sharma, & McGhee, 2010)	VS		
13	Nassiri2010 (Nassiri, Edrissi, & Alibabai, 2010)	DS		
14	Delanoy2011 (Delanoy, Gaede, & Wong, 2011)			
15	Bede2011 (Bede & Péter, The mathematical modeling of reversible lane system, 2011)			
16	Cheng2011 (Cheng, Li, & Liu, 2011)	VS		
17	Liu2011 (Liu, Yang, Ke, & Wang, 2011)	VS		
18	Hausknecht2011 (Hausknecht, Au, Stone, Fajardo, & Waller, 2011)			
19	Dey2011 (Dey, Ma, & Aden, 2011)			
20	Zhao2014 (Zhao, Ma, Liu, & Yang, 2014)	VS		
21	Bede2014 (Bede & Torok, Theoretical investigation of traffic equilibrium on bridges, 2014)			
22	Frejo2016 (Frejo, Papamichail, Papageorgiou, & Camacho, 2016)	ML		
23	Ampountolas2018 (Ampountolas, Furtado, & Carlson, 2018)	AS		
24	Ding2017 (Ding, Zhang, Yuan, Zheng, & Guo, 2018)	VS		
25	Lobont2019 (Lobont & Pînzariu, 2019)	SY		
26	Kotagi2019 (Kotagi & Asaithambi, 2019)	ML		
27	Mao2020 (Mao, et al., 2020)	VS		
28	Chu2020 (Chu, Lam, & Li, 2020)	ML		
29	He2021 (He, Zhu, Zhang, & Ma, 2021)	VS		
30	Liu2022 (Liu, Ayng, Ke, & Wang, 2022)	VS		
31	Striewski2022 (Striewski, Thomsen, & Tomforde, 2022)	AS		
32	Wollenstein2022 (Wollenstein-Betech, Paschalidis, & Cassandras, 2022)			
33	Fu2022 (Fu, Tian, & Sun, 2022)	VS		
34	Katzilieris2023 (Katzilieris, Kampitakis, & Vlahogianni, 2023)	SM		
35	Li2023 (Li, Li, & Han, 2023)	CA		

Purpose:

City
Evacuation
Autonomous Vehicles
General

Table 3. Simulation tools used in papers

Initial	Tool name	Usage
VS	Vissim[5]	11
ML	MatLab[6]	3
AS	AimSum[7]	2
CA	Cellular Automaton[8]	1
CS	CorSim[9]	1
DS	DynaSmart-P[10]	1
SM	Sumo[11]	1
SY	Synchro[12]	1
VT	Vista[13]	1

Table 3 presents simulation tools used in 22 papers from the list of selected papers, with focus on traffic simulation. The column Usage shows representation of tool in research papers.

[5] https://www.ptvgroup.com/en/products/ptv-vissim
[6] https://www.mathworks.com/products/matlab.html
[7] https://www.aimsun.com
[8] https://udel.edu/~mm/traffic/ca.html
[9] https://mctrans.ce.ufl.edu/tsis-corsim/
[10] umventures.org/technologies/dynasmart-p
[11] https://eclipse.dev/sumo/
[12] www.cubic.com/transportation/products/intelligent-transportation-solutions/intelligent-systems/synchro-studio#
[13] https://vista.csail.mit.edu/

3 Documents analysis

In this section will be described scope of selected papers presented in the Table 2, and their relevance for the current research. Although the lane reversal (tidal flow) has been in use for a very long time, probably since beginning of the 20th century, for this review are relevant just paper published in last 20 years, after 2004.

Handke et al. (Handke, 2006) describe a simulation of a semi-automatic lane reversal traffic management system in Hannover (Germany), with about 1,000 dynamic components over a 16 km distance, operated by just six people. The use of the Path Flow Estimator for dynamic origin-destination matrix adjustment and mesoscopic simulation adds sophistication. The system's reliance on hardware integration and the complexity of interfacing subsystems highlights potential challenges in scalability and long-term maintenance. Furthermore, while initial results are promising, more extensive evaluation and comparative studies would strengthen confidence in the system's effectiveness and generalizability to other urban contexts.

Logan et al. (Logan, McLeod, & Jordan, 2006) described the successful implementation of a manual lane reversal system in the City of Calgary (Canada). Analysis of the implemented system showed improvements, decreasing and stopped delays, which is an excellent example of how existing infrastructure can be optimized. It provides a practical demonstration of how lane reversal can efficiently optimize urban traffic flow using existing infrastructure. However, the reliance on manual operations during the initial phase raises questions about long-term scalability and automation in comparable urban projects.

In the survey conducted in 2003 in the USA, Wolshon et al. (Wolshon & Lambert, 2006) identified a total of 45 discontinued reversible lane systems. The survey showed that lane reversal was mostly used for next purposes: congestion mitigation, planned special events, emergency management, roadway construction, transit and homeland security, where the first three were dominant. The survey also showed the most common types of roadways for the lane reversal: freeway, arterial roadway, bridge and tunnel. Despite the long history and varied applications of reversible lanes, there is a significant lack of established standards and guidelines for their planning, design, and operation.

Department of Transportation in the state Virginia (USA) performed an evaluation study of the Hampton Roads evacuation routes to determine which route would be the best choice in the case of evacuation (Shen & Sokolowski, 2007). They used the VISSIM microscopic simulation tool to model two-phase evacuation, ramp open-close, as well as a lane reversal scheme on six interstate routes in the region. The simulation results showed a mild impact on overall evacuation on one of the simulated routes.

Similar research was conducted in the USA state of Nord Carolina (Williams, Tagliaferri, Meinhold, Hummer, & Rouphail, 2007), concerning how efficiently would the beginning of the lane reversal operate, how would the open interchanges operate under contraflow conditions and how efficiently would the merge/crossover contraflow termination operate. The research successfully simulated a lane reversal plan for hurricane evacuation, demonstrating its effectiveness. However, a key finding and point of concern is the potential for extensive traffic queue formation at the end of the contraflow route.

Evacuation is a situation where residents in a dangerous area are removed to safe places as quickly as possible, and there are three models for the evacuation situation: microscopic simulation models traffic flow at the single-vehicle level, mesoscopic simulation models traffic by groups of entities, and macroscopic simulation describes traffic flow at a high level of aggregation (Kim, Shekhar, & Min, 2008). The key contribution is presenting what the authors believe to be the first macroscopic approaches for contraflow network reconfiguration, which incorporates key factors such as road capacity constraints and congestion. The study evaluates these approaches analytically and experimentally using real-world data, showing promising results.

The Houston-Galveston Area Council, together with other agencies responsible for safe and adequate transportation flow in the metropolitan area, identified the need for a tool to test policies and operational decisions for safe, secure, and efficient evacuations. Such a tool was developed using a dynamic traffic assignment model (Brown, White, Van Slyke, & Benson, 2009). A key critical point is that this model is a planning tool designed for system-wide performance evaluation and is not intended for real-time traffic management.

Wu et al (Wu, Sun, Gao, & Zhang, 2009) provide a valuable study on optimizing traffic flow models by integrating reversible lanes with an advanced traveler information system. A potential limitation, however, is that the findings are based on a theoretical model and simulation, not on a practical, real-world application. While the paper's findings are promising, it would be beneficial to see how the model performs in a real-world scenario with real-time data and driver behavior. This would provide more robust evidence of the model's effectiveness and feasibility.

Evacuation models may be categorized into two paradigms: simulation-based models and optimization-based models. Specific simulation software packages have been developed since 1980, but they were limited to evaluating a small set of pre-specified plans, mainly because of the high computational cost at that time. Optimization-based methods can identify optimal planning solutions for evacuation. Xie et al. presented an integrated approach combining both streams of research through a simulation-based dynamic traffic assignment (DTA) model, using lane reversal on road

sections and crossing elimination at intersections (Xie, Lin, & Travis Waller, 2010). Its findings are based on a theoretical model and simulation, which raises questions about the practical feasibility and safety of implementing the proposed strategies in a real-world, dynamic traffic environment.

Lambert et al. (Lambert & Wolshon, 2010) analyzed traffic flow on reversible traffic roads. The goal of this research was to collect and analyze data on traffic flow characteristics and to determine the difference between normal and reversible segments under different conditions: roads, cities, and conditions (peak period, off-peak period, game day). Data (speed, volume) were collected at two locations with double lanes, using a pneumatic tube recording system. To determine the difference between speed and volume in corresponding time blocks, a mixed model analysis of variance (ANOVA Model III) was used. The analysis clearly showed better traffic conditions using reversible lanes. This research is crucial because the lack of reliable data has previously made it difficult to confidently model or gauge the true benefits of these systems.

The previous mentioned study was analyzed and compared with another similar study (City of Tucson, 2004) (Bede, Szabó, & Péter, 2010), which showed improvement in travel time, total delay, total stops, fuel consumption, and emissions of HC, CO, and NO. The study highlights a critical limitation, noting that this type of dynamic routing "changes the system's structure in an optimal direction" but raises many unresolved practical and safety questions.

Edara et al. (Edara, Sharma, & McGhee, 2010) conducted a literature review of traffic modeling studies regarding the evaluation of hurricane evacuation plans in the USA. Widely used tools are simulations: microscopic, for traffic operations at intersections, highways, and roadways, for time estimates, bottleneck identification, and disaggregate performance measures such as link speed, delay, or queue length; mesoscopic, for evaluating advanced traffic control strategies, evacuation time estimates, and bottleneck identification on large evacuation road networks; macroscopic, for evacuation model forecasts at state, county, and street resolutions. The paper identified three major tasks in simulation development: coding the simulation network and assigning traffic, establishing study scenarios and simulation setup and establishing additional simulation assumptions. It developed a large-scale traffic simulation model for hurricane evacuation, providing valuable insights into network performance and bottlenecks. However, as the findings are based solely on a simulation, their real-world applicability could be limited by unpredictable factors not accounted for in the model.

Contraflow is another term used for lane reversal in some studies (Nassiri, Edrissi, & Alibabai, 2010). This paper presents a simulation model tested for the network of the City of Fort Worth, Texas, especially for evacuation purposes in cases of natural or human-made disasters. The Logit choice model was used for contraflow prediction with online and offline scenarios. Contraflow is modeled through a bi-level programming approach. The results show an improvement of throughput by up to 16% using contraflow compared with the non-contraflow case. The paper addresses a gap in existing literature, noting that previous studies had not thoroughly addressed the online contraflow problem.

Delanoy et al. (Delanoy, Gaede, & Wong, 2011) conducted a study on an improved lane reversal system in Calgary, Canada - the same one already studied five years ago . The system was improved with video equipment, LED traffic signs, dual-state prism signs, and a traffic management operation center.

Bede and Peter (Bede & Péter, 2011) presented a descriptive mathematical network model of a reversible lane system - a positive non-linear dynamic system. The results of model examinations during morning and afternoon peaks show that improvement of vehicle density decreases peaks by activating reversible lanes. The paper describes techniques applied to be useful in working with very large dynamic network loading and a mesoscopic simulation model. Although a macroscopic model loses the properties of single-vehicle traffic flow, it effectively deals with road density and the capacity of a transportation network. While it highlights a significant reduction in travel time and waiting time, it also notes the necessity of further analyzing the risk factors associated with the application of such systems.

(Cheng, Li, & Liu, 2011) studied traffic and intersection organization methods for reversible lanes by simulating road networks of Kunming (China) with 11 intersections and proved improvement in total delay (8.5%), capacity (7.4%), and average delay (14.8%). The paper describes techniques applied to be useful in working with very large dynamic network loading and a mesoscopic simulation model. A critical point is that the paper's findings, while promising, are based on a simulation rather than a real-world implementation.

A similar case study and simulation showed positive effects of the reversible lane system, simulated on four intersections along a three-lane urban arterial: travel time was reduced by 38.73% and average vehicle delay decreased by 50.7% using the reversible lane system on congested roads (Liu, Yang, Ke, & Wang, 2011). The paper provides a strong theoretical framework for implementing reversible lanes in urban areas, including an optimization model and specific activation criteria. However, as the research is based on a conceptual model, it does not address the practical and safety concerns of a real-world application.

In the ILP (Integer Linear Program) experiment based on macroscopic ILP traffic model, Hausknecht et al. (Hausknecht, Au, Stone, Fajardo, & Waller, 2011) analyzed the effectiveness of dynamic lane reversal and proposed a framework for lane reversal optimization using a linear programming formulation and a bi-level programming formulation. Analysis of

experiment results showed an increase in network efficiency by 72% with active contraflow. The paper provides a strong theoretical framework for implementing reversible lanes in urban areas, including an optimization model and specific activation criteria.

Although reversible lanes help optimize traffic capacity, a study conducted in Washington, D.C., on 10 roadway segments with reversible lanes, with a total length of about 17 km, suggested that more crashes are associated with reversible lane operations compared to non-reversible lane segments (Dey, Ma, & Aden, 2011). Reversible lanes can be confusing if the turning operations are not clearly signed, especially for tourists and untrained drivers, according to the authors. However, a key limitation is that the findings are based on a single, specific urban environment and only draw "preliminary conclusions," which may not be universally applicable to other cities.

Zhao et al. (Zhao, Ma, Liu, & Yang, 2014) proposed a mathematical model for a road traffic section with three four-arm intersections and two internal segments. The paper presented a simulation experiment with six different study patterns, with cycle lengths between 60 and 120 seconds, and a clearance time of 5 seconds. Analysis of experiment results suggests that the lane reversion strategy has very limited success when intersections are well optimized and is not recommended in similar cases. The simulation validation using only VISSIM with deterministic scenarios is insufficient to demonstrate real-world effectiveness, particularly regarding driver confusion and safety concerns during lane reversals that are only briefly acknowledged but not quantitatively addressed.

Bede and Peter (Bede & Torok, 2014) provided a mathematical model of traffic flows with reversible lanes from an economic point of view. In the paper, the authors presented results of an investigation of price as a traffic controlling parameter, but they also discussed vehicle dynamics, fuel consumption, and environmental impact. The numerical experiment was conducted on a simple bridge model with four lanes and bridge tolling. Theoretical analysis showed that reversible lanes could lead to decreased prices if it is possible to change lane direction while keeping road safety in mind. Mathematical modeling is superficial, lacking empirical validation, and the authors fail to demonstrate how their theoretical economic equilibrium model would translate into practical traffic management solutions for real bridge operations.

Frejo et al. (Frejo, Papamichail, Papageorgiou, & Camacho, 2016) proposed a modified macroscopic second-order traffic flow model, METANET, for reversible lanes on freeways and tested the proposed MPC (Model Predictive Control) strategy with the MATLAB simulation environment. The model was validated with real data over the Centenario Bridge in Seville, Spain, and results confirmed that the model can reproduce traffic congestion with an error between 4% and 14% for flow and speed. The proposed model and algorithms are validated solely through simulation, which may not fully capture the complexities and unpredictable nature of real-world traffic.

Ampountolas et al. (Ampountolas, Furtado, & Carlson, 2018) proposed a simple strategy for reversal lane operations, based on a triangular FD (Fundamental Diagram) policy. In the paper, they use the term “tidal flow” for reversal lane. The proposed algorithm for the FD-based switching policy was evaluated with a microsimulation study of a road segment of Aston Expressway in Birmingham, UK. For this purpose, the microscopic simulation framework AIMSUN was used. Simulation outputs were compared with real data on the road, and the results showed an increase in motorway throughput and decreased total travel time by 48.4%. This simple and practical real-time control strategy for road lane reversal, validated through simulation, should be yet tested in real-world performance.

Regarding Ding et al. (Ding, Zhang, Yuan, Zheng, & Guo, 2018), the lane reversal strategy can be widely used on city road segments with intersections, but with some constraints: the intersection signal control should have multiphase control with three phases and special dedicated left-turn and right-turn lanes; considering safety factors, the number of road crossings should not be less than three; a reversible lane segment should have enough time and space conditions to set up direction change. This study presented the VISSIM simulation of a road segment with an intersection between two roads in Xuancheng city in China, using simulation data before and after setting the reversible lane strategy. The result comparison showed queue length improvement of 17.7% and travel time reduction by 20.3%. But this solution suffers from limited scope: analyzing only a single intersection without addressing scalability, real-world validation, or potential safety concerns that are critical for such traffic management interventions.

Lobont and Pinzariu (Lobont & Pînzariu, 2019) analyzed a road segment between two intersections, on one of the main transit arteries in the city of Sibiu, Romania. Real data were collected from actual traffic conditions and implemented with the Synchro simulation framework. Observation of traffic flow without lane reversal showed daily congestion between the two intersections. Implementation of lane reversal solutions in the simulation environment showed only partly satisfactory results, as the improvement was significant only at intersection B, while the value at intersection A was worse. However, inconclusive results and lack of alternative solutions undermine the paper's practical contribution to traffic engineering.

Kotagi and Asaithambi (Kotagi & Asaithambi, 2019) conducted an extensive study with microscopic simulation of lane reversal implementation in two Indian cities: Bangalore and Kollam. The simulation environment was developed with MATLAB, using Object-Oriented Programming (OOP) concepts. The objectives of the study were: to identify and model

movements of different kinds of vehicles on urban roads; to develop a simulation model for bi-directional mixed traffic; and to evaluate lane reversal operation using the developed model. Real traffic data were collected using video-graphic technique, on a road segment without the effect of intersections, bus-stops, parked vehicles, and pedestrians, during morning and afternoon peak hours in both directions. The simulated road segment had six lanes, with ratios 3:3 (without lane reversal) and 4:2 (with lane reversal). The simulation model was validated using collected datasets, and it was proven that the simulation model can be used for simulation. Comparison of simulation output showed improvement by 20.5% in the morning direction and 19.2% in the opposite direction using lane reversal. There was also an increase of 568 veh/h in the morning peak and 396 veh/h during the evening peak. These results showed promise in reducing traffic congestion on urban roads in Indian cities. However, the model's validation is limited to two particular cities, and the reversible lane evaluation lacks critical safety analysis and practical implementation considerations such as signaling, enforcement mechanisms, and transition zones that would be essential for real-world deployment.

Mao et al. (Mao, et al., 2020) proposed real-time dynamic lane reversal in CVIS (Intelligent Cooperative Vehicle Infrastructure System), simulated with VISSIM microscopic simulation software. Simulation was conducted on an eight-lane road example, with two inner reversible lanes, and three phases of lane direction change: opening, entering, and closing. The dataset was collected on a road segment in the Chinese city of Huai'an during afternoon peak hours. Comparison of simulation outputs for traditional time-controlled lane change mode and proposed real-time lane reversal showed improvement of 13.5% in traffic flow and reduction of average delay by 27.4% using real-time lane reversal. The practical feasibility of the proposed "light curtain wall" technology and its integration with existing infrastructure remains unproven, and the study relies heavily on simulation without extensive real-world validation.

A dynamic lane reversal strategy, combined with a distributed algorithm, can significantly improve the travel time of connected and autonomous vehicles (CAV), as shown by MATLAB simulation of the proposed model (Chu, Lam, & Li, 2020). The study's reliance on simulations with idealized assumptions, such as constant vehicle speeds, no traffic lights, and a fully CAV-dominated environment limits its immediate real-world applicability, especially in mixed-traffic scenarios.

He et al. (He, Zhu, Zhang, & Ma, 2021) presented a simulation of a lane reversal control system at intersections based on VISSIM simulation software extended with a Python interface. The simulation environment was tested with real traffic flow data on a two-way 10-lane road segment. The result of the analysis showed a decrease in average queue length at intersections by 38.6%, average delay time by 20.07%, and average carbon emission by 17.3% compared with the system without a variable lane reversal system. The approach relies on simplified traffic scenarios and does not address the challenges of real-time implementation, such as driver compliance, infrastructure costs, or integration with adaptive signal control systems.

Liu et al. (Liu, Ayng, Ke, & Wang, 2022) presented the Predictive Empowered Assignment scheme for Reversible Lane (PEARL) - an advanced traffic flow prediction module with a bi-level optimization model. The model was tested with a simulation experiment using the VISSIM simulator. The test data were collected on a 16.8 km long segment of Guangfo Express Road in China. The results of the experiment showed total throughput improvement by 5.76% in one direction, 2.04% in the other direction, and a total increase in both directions by 4.47%. The paper assumes idealized conditions, such as perfect driver compliance and homogeneous lane properties and its performance gains, while promising, are modest given the heavy traffic loads, raising questions about scalability and robustness in more complex or mixed-traffic environments.

Striewski et al. (Striewski, Thomsen, & Tomforde, 2022) used the road traffic simulation software Aimsun Next to develop and evaluate four tidal-flow-management algorithms: ratio, speed, density, and occupancy. A 1 km road segment with 6 lanes in both directions served as the experimental setup. Three models were tested: without intersections, with one intersection and with two intersections. The occupancy algorithm performed best, reducing total travel time to 81% (no intersections), 71% (one intersection), and 68% (two intersections) of baseline values. The study is based on simulations, which may not fully address the complex, real-world implementation challenges and safety considerations.

Wollenstein-Betech et al. (Wollenstein-Betech, Paschalidis, & Cassandras, 2022) proposed multiple lane-reversal solutions and validated them on Eastern Massachusetts' transportation network. Their case study demonstrated a 4.7% travel time improvement. The paper makes a significant contribution by proposing a global optimization approach to the joint lane reversal and traffic routing problem, which has been shown to be NP-hard. The effectiveness of the proposed methods is demonstrated through a mathematical model and not validated in a large-scale real-world or simulated dynamic environment.

Fu et al. (Fu, Tian, & Sun, 2022) identified priority roads for dynamic lane reversal in CAV (connected and autonomous vehicles) environments by developing a multi-objective optimization model (Cell Transmission Model), integrating NSGA-II genetic algorithm with VISSIM simulations and conducting regression analysis. Results showed lane reversal performs best on multilane roads with high traffic flow and balanced bidirectional demand (small flow discrepancies). The reliance on idealized assumptions

such 100% CAV penetration and simplified lane-changing behaviors, limits the immediate real-world applicability, and the safety implications of high-frequency lane reversals remain underaddressed.

Katzilieris et al. (Katzilieris, Kampitakis, & Vlahogianni, 2023) presented microscopic simulation model for dynamic lane reversal built within SUMO (Simulation of Urban Mobility) road traffic simulator. The model was built within imaginary 4-way intersection – 500 m long 2-lane artery with one reversible lane in the middle, and secondary streets with static 2 lanes. Simulation used reinforcement learning (LR) algorithm to optimize direction change in the middle lane. Results of such model showed average travel time improvement up to 19.9% and average decreasing of CO_2 emissions on the intersections up to 15.1%. A critical limitation is that the model's effectiveness is demonstrated on an "artificial intersection" in a simulated environment, raising questions about its practical application in real-world, large-scale urban networks.

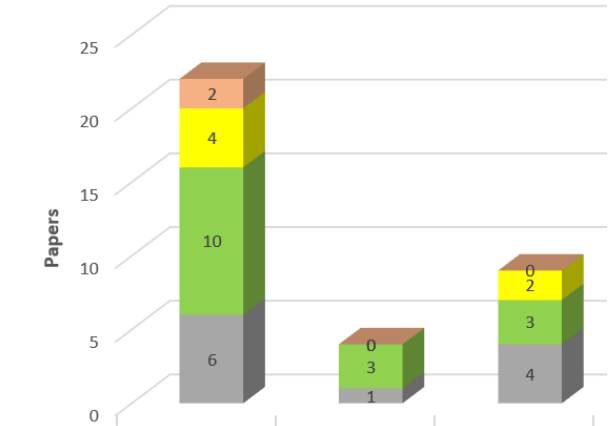
Li et al. (Li, Li, & Han, 2023) proposed dynamic lane reversal strategy which dynamically changes direction on reversible lane depending on current traffic density. A two way 6-lane road segment is divided 3:3 without lane reversal, and 4:2 or 2:4 with the lane reversal strategy. Two crucial questions were in focus of the research: how to establish the criteria to trigger entering lane reversal mode; and how to determine optimal duration for the contraflow. The simulation based on cellular automaton (CA) model showed optimal direction change when density reaches 0.37 and goes back to normal direction when density falls to 0.22. The study relies on a simplified cellular automaton model and does not address practical challenges such as driver compliance, infrastructure costs, or the integration of real-world traffic variability and connected vehicle technologies.

4 Discussion

The previous chapter gives a short overview of each paper from the list presented in Table 2, altogether 35 papers. Upon analyzing the papers, it was found that individual studies commonly employed one of three approaches: simulation models, mathematical models, or observational studies. Consequently, the papers were categorized accordingly. From the perspective of their research objectives, themes such as urban traffic optimization, traffic management during evacuations, application of autonomous vehicles, and general traffic control purposes were identified (s. Table 2).

From this list of 35 papers, 22 papers (62,86%) use some kind of simulation as a research tool to reach research objectives. Other papers have a goal to make a study of the topic or make a description of the strategy using mathematic model.

From another point of view, use case of some specific cities were described in 16 papers (45.71%) – almost a half of the papers, and mostly with simulation as a research tool (10 papers). Complete distribution is presented by Fig. 5.



	Simulation	Study	Model
AV	2	0	0
Evacuation	4	0	2
City	10	3	3
General	6	1	4

Figure 5. Selected paper distribution

Traffic simulation software tools are the best solutions to evaluate models of specific road segment and to optimize control strategy on it. They have been widely used for more than 30 years. As presented in Table 2 and Table 3, 50% of simulation focused papers from the list are using VISSIM simulation tool. 3 papers are using MATLAB, 2 papers AIMSUM, and other tools (Sumo, Corsim etc.) are used in 1 paper each.

Until 2010, they were widely used to model lane reversal strategy for evacuation purposes in cases of natural disasters, and later research was more focused to model and evaluate automatic control strategy of existing, manually controlled lane reversal segments, or to model and evaluate some new segments with huge congestion.

The review of dynamic lane reversal systems over the past two decades reveals significant progress in both the conceptual understanding and practical implementation of reversible lanes as a traffic management strategy. Traditional reversible lane systems, managed manually or by fixed schedules, have demonstrated measurable benefits, such as reductions in travel time, stops, delays, fuel consumption, and vehicle emissions. However, these conventional approaches are limited by their inflexibility and inability to respond dynamically to real-time traffic fluctuations.

Recent advancements in information technology, particularly the integration of artificial intelligence (AI), machine learning (ML), and sophisticated traffic simulation tools, have opened new avenues for optimizing lane reversal operations. Modern simulation platforms enable accurate modeling of complex, multimodal urban traffic networks and

facilitate the evaluation of innovative, adaptive control strategies. This transition from static to dynamic, data-driven management holds the promise of even greater improvements in efficiency, safety, and sustainability.

Despite these advancements, the literature review highlights several research gaps. There is a need for more real-world deployments of AI-driven dynamic lane reversal systems and comprehensive studies evaluating their long-term impacts on urban mobility. Additionally, the integration of connected and autonomous vehicles into these systems presents both challenges and opportunities that warrant further investigation.

In summary, dynamic lane reversal remains a vital and evolving tool in the quest for smarter, more responsive urban transportation systems. Continued research, especially leveraging AI and real-time data, will be crucial in realizing the full potential of reversible lanes, ultimately contributing to more sustainable and efficient cities.

5 Conclusion

Dynamic Lane Reversal (DLR) control is one of the most cost-effective strategies to increase capacity of road traffic, improve traffic flow, decrease congestion, delay time and travel time on existing road infrastructure, without physical interventions. Well optimized DLR can enhance traffic flow by up to 40% compared to conventional reversible lanes control systems. It can significantly decrease total queuing delays, especially on multilane road segments with balanced directional flows. It effectively reduces road congestion and improves overall system performance through optimal control strategies.

This strategy has been in use since the late 1920s, first using physical barriers and ramps, then traffic signals and time-based regulation, and with recent advancement of information and automotive technologies, focus shifts to artificial intelligence as well as to autonomous driving.

Based on these findings, the suggested next steps are to create a prototype for intelligent lane management that leverages both real-time and historical data from smart cameras, traffic sensors, and weather sensors, enhanced by artificial intelligence algorithms. Initially, this system should be tested within a simulation platform that accurately replicates the chosen real-world setting as a digital twin. Demonstrating success in simulation, while adhering to traffic safety standards, will be essential before proceeding to trials in an actual environment, which should involve collaboration with municipal and government agencies. Identifying effective ways to optimize road traffic management through artificial intelligence will present a challenge for future research.

References

Almeida, A., Fonseca, J., Rasinhas, P., Costa, C., Paiva, D., Silva, G., . . . Sargento, S. (2023). Safe Roads: Traffic Management and Road Safety Platform for Smart Cities. *9th IEEE World Forum on Internet of Things.* Aveiro: IEEE Xplore. doi:10.1109/WF-IoT58464.2023.10539505

Ampountolas, K., Furtado, H., & Carlson, R. (2018). Motorway Tidal Flow Lane Control. Elsevier B.V. doi:10.1016/j.ifacol.2018.07.046

Astarita, V., Giofre, V. P., Guido, G., & Vitale, A. (2020). A Review of use of the traffic simulation for the evaluation of traffic safety levels: can we use simulation to predict crashes? *23rd EURO Working Group on Transportation Meeting, EWGT 2020* (pp. 244-251). Pahpos, Cyprus: Transportation Research Procedia 52.

Bede, Z., & Péter, T. (2011). The mathematical modeling of reversible lane system. *Periodica Polytechnica Transportation Engineering.* Budapest University of Technology and Economics. doi:10.3311/pp.tr.2011-1.02

Bede, Z., & Torok, A. (2014). Theoretical investigation of traffic equilibrium on bridges. *Transport and Telecommunication*. doi:10.2478/ttj-2014-0013

Bede, Z., Szabó, G., & Péter, T. (2010). Optimalization of road traffic with the applied of reversible direction lanes. *Periodica Polytechnica Transportation Engineering*. doi:10.3311/pp.tr.2010-1.01

Brown, C., White, W., Van Slyke, C., & Benson, J. (2009). Development of a strategic hurricane evacuation-dynamic traffic assignment model for the Houston, Texas, region. *Transportation Research Record*. doi:10.3141/2137-06

Cheng, W., Li, X., & Liu, S. (2011). Research on the traffic control and organization coordinate applications of reversible lanes. *Proceedings of 2011 3rd International Conference on Awareness Science and Technology, iCAST 2011.*

Chu, K.-F., Lam, A., & Li, V. (2020). Dynamic Lane Reversal Routing and Scheduling for Connected and Autonomous Vehicles: Formulation and Distributed Algorithm. *IEEE Transactions on Intelligent Transportation Systems*.

City of Tucson. (2004, 04). Grant Road Reversible Lane facts. *Department of Transportation publication*. Retrieved from http://dot.tucsonas.gov/news

Delanoy, C., Gaede, T., & Wong, Y. (2011). The City of Calgary's reversible lane control systems. *2011 Conference and Exhibition of the Transportation Association of Canada - Transportation Successes:*

Let's Build on Them, TAC/ATC 2011. Transportation Association of Canada (TAC).

Dey, S., Ma, J., & Aden, Y. (2011). Reversible lane operation for arterial roadways: The Washington, D, USA experience. *ITE Journal (Institute of Transportation Engineers)*.

Ding, H., Zhang, Y., Yuan, H.-Y., Zheng, X.-Y., & Guo, F. (2018). Reversible lane segment design for urban signalized intersection: A case study. *CICTP 2017: Transportation Reform and Change - Equity, Inclusiveness, Sharing, and Innovation - Proceedings of the 17th COTA International Conference of Transportation Professionals* (pp. 2648-2656). American Society of Civil Engineers (ASCE). doi:10.1061/9780784480915.278

Edara, P., Sharma, S., & McGhee, C. (2010). Development of a large-scale traffic simulation model for hurricane evacuation-methodology and lessons learned. *Natural Hazards Review*. doi:10.1061/(ASCE)NH.1527-6996.0000015

Ejercito, P. M., Nebrija, K. G., Feria, R. P., & Lara-Figueroa, L. L. (2017). Traffic Simulation Software Review. *8th International Conference on Information, Intelligence, Systems and Applications* (pp. 1-4). Larnaca: IEEE Xplore. doi:10.1109/IISA.2017.8316415

Frejo, J., Papamichail, I., Papageorgiou, M., & Camacho, E. (2016). Macroscopic Modeling and Control of Reversible Lanes on Freeways. *IEEE Transactions on Intelligent Transportation Systems*. doi:10.1109/TITS.2015.2493127

Fu, Q., Tian, Y., & Sun, J. (2022). Modeling and simulation of dynamic lane reversal using a cell transmission model. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*.

Handke, N. (2006). Traffic management during large events the tidal flow system in Hannover. (p. 13th World Congress on Intelligent Transport Systems and Services). Intelligent Transport Systems (ITS).

Hausknecht, M., Au, T.-C., Stone, P., Fajardo, D., & Waller, T. (2011). Dynamic lane reversal in traffic management. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC.*

He, J., Zhu, Y., Zhang, J., & Ma, X. (2021). Reversible lane control system with low emission load based on VISSIM simulator. *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, ICBAIE 2021.* Institute of Electrical and Electronics Engineers Inc.

Kamble, S. S., Gunasekaran, A., & Gawankar, S. A. (2018). Sustainable Industry 4.0 framework: A systematic literature reviewidentifying the current trends and future perspectives. *Elsevier*, 408-425. doi:https://doi.org/10.1016/j.psep.2018.05.009

Katzilieris, K., Kampitakis, E., & Vlahogianni, E. (2023). Dynamic Lane Reversal: A reinforcement learning approach. *2023 8th International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2023.* Institute of Electrical and Electronics Engineers Inc.

Kim, S., Shekhar, S., & Min, M. (2008). Contraflow transportation network reconfiguration for evacuation route planning. *IEEE Transactions on Knowledge and Data Engineering.* doi:10.1109/TKDE.2007.190722

Kotagi, P., & Asaithambi, G. (2019). Microsimulation approach for evaluation of reversible lane operation on urban undivided roads in mixed traffic. *Transportmetrica A: Transport Science*.

Lambert, L., & Wolshon, B. (2010). Characterization and comparison of traffic flow on reversible roadways. *Journal of Advanced Transportation.* doi:10.1002/atr.114

Lathrop, W. H. (1972). Reversible roadway controls. *Traffic Quarterly, 26*(1), 133-147.

Li, W., Li, J., & Han, D. (2023). Dynamic Lane Reversal Strategy in Intelligent Transportation Systems in Smart Cities. *Sensors*.

Liu, C., Ayng, H., Ke, R., & Wang, Y. (2022, 12). Toward a Dynamic Reversible Lane Management Strategy by Empowering Learning-Based Predictive Assignment Scheme. *IEEE Transactions on intelligent transportation systems, 23*(12). doi:10.1109/TITS.2022.3202487

Liu, C., Yang, H., Ke, R., & Wang, Y. (2011). Toward a Dynamic Reversible Lane Management Strategy by Empowering Learning-Based Predictive Assignment Scheme. *IEEE Transactions on Intelligent Transportation Systems*.

Lobont, L., & Pînzariu, A. (2019). Effects of reversible lane implementation, a case study simulation. *MATEC Web of Conferences.* EDP Sciences.

Logan, M., McLeod, T., & Jordan, C. (2006). 5th avenue connector contra-flow facility: From concept to completion. *TAC/ATC 2006 - 2006 Annual Conference and Exhibition of the Transportation Association of Canada: Transportation Without Boundaries.* Transportation Association of Canada (TAC).

Mao, L., Li, W., Hu, P., Zhou, G., Zhang, H., & Dai, J. (2020). Design of Real-Time Dynamic Reversible Lane in Intelligent Cooperative Vehicle Infrastructure System. *Journal of Advanced Transportation*.

Nassiri, H., Edrissi, A., & Alibabai, H. (2010). Estimation of the logit model for the online contraflow problem. *Transport, 25*(4), 443-441. doi:10.3846/transport.2010.53

Peters, M. D., Marnie, C., Colquhoun, H., Garritty, M. C., Hempel, S., & Horsley, T. (2021). Scoping reviews: reinforcing and advancing the methodology and application. *Systematic Reviews*. doi:https://doi.org/10.1186/s13643-021-01821-3

Rowley, J., & Slack, F. (2004). Conducting a literature review. *Management Research News, 27*, 31-39. doi:http://dx.doi.org/10.1108/01409170410784185

Sharma, P., Liu, H., Honggang, W., & Shelley, Z. (2017). Securing wireless communications of connected vehicles with artificial intelligence. *2017 IEEE International Symposium on Technologies for Homeland Security, HST 2017.* Walthram: Institute of Electrical and Electronics Engineers Inc. doi:10.1109/THS.2017.7943477

Shen, Y., & Sokolowski, J. (2007). Hampton roads hurricane evacuation arterial routes preliminary study. *Business and Industry Symposium, BIS 2007 - Proceedings of the 2007 Spring Simulation Multiconference, SpringSim 2007.* Association for Computing Machinery, Inc. doi:10.1145/1404803.1404820

Striewski, S., Thomsen, I., & Tomforde, S. (2022). Adaptive Approaches for Tidal-Flow Lanes in Urban-Road Networks. *Future Transportation*.

Tranfield, D., Danyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence - Informed Management Knowledge by Means of Systematic Review. *British Journal of Management, 14*, 207-222. doi:10.1111/1467-8551.00375

Williams, B., Tagliaferri, A., Meinhold, S., Hummer, J., & Rouphail, N. (2007). Simulation and analysis of freeway lane reversal for coastal hurricane evacuation. *Journal of Urban Planning and Development*. doi:10.1061/(ASCE)0733-9488(2007)133:1(61)

Wollenstein-Betech, S., Paschalidis, I., & Cassandras, C. (2022). Optimizing lane reversals in transportation networks to reduce traffic congestion: A global optimization approach. *Transportation Research Part C: Emerging Technologies*.

Wolshon, B., & Lambert, L. (2006). Reversible lane systems: Synthesis of practice. *Journal of Transportation Engineering*. doi:10.1061/(ASCE)0733-947X(2006)132:12(933)

Wu, J., Sun, H., Gao, Z., & Zhang, H. (2009). Reversible lane-based traffic network. *Engineering Optimization, 41*, 87-97. doi:10.1080/03052150802368799

Xie, C., Lin, D.-Y., & Travis Waller, S. (2010). A dynamic evacuation network optimization problem with lane reversal and crossing elimination strategies. *Transportation Research Part E: Logistics and Transportation Review*. doi:10.1016/j.tre.2009.11.004

Yong-Hak, J. (n.d.). *Web of Science facts sheet*. (Thomson Reuters) Retrieved from http://wokinfo.com/media/pdf/WoSFS_08_7050.pdf

Zhao, J., Ma, W., Liu, Y., & Yang, X. (2014). Integrated design and operation of urban arterials with reversible lanes. *Transportmetrica B: Transport Dynamics 2014. 2*, pp. 130-150. Taylor and Francis Ltd. doi:10.1080/21680566.2014.908751

Trends in Software Engineering

Boris Tomaš
Impact Analysis of Croatian Language as the Publication Language on the Research Paper Visibility and Citation Potential in Computer Science and Engineering / Analiza utjecaja hrvatskog jezika kao jezika publikacije na vidljivost i citiranost znanstvenih radova u području računalnih znanosti i inženjerstva

Marija Kuštelega
Dark Patterns in the Public Sector: A Study of Cookie Consent Interfaces on Croatian Websites

Pablo Hidalgo, Adrián Domínguez, Daniel Rodriguez, Marko Mijač
A preliminary Study on the Use of Generative AI in Software Engineering Education

Marko Horvat, Barbara Kralj, Gordan Gledec
A Comparative Study of Vibe Coding with ChatGPT and Gemini in Front-end Web Development

Roberta Capuano, Henry Muccini, Zlatko Stapić
Implementing Education 4.0 in Software Engineering Education: the JCC experience

Impact Analysis of Croatian Language as the Publication Language on the Research Paper Visibility and Citation Potential in Computer Science and Engineering

Analiza utjecaja hrvatskog jezika kao jezika publikacije na vidljivost i citiranost znanstvenih radova u području računalnih znanosti i inženjerstva

Boris Tomaš
Trinity College Dublin
School of Computer Science and Statistics, CONNECT Centre
Dublin 2, Ireland D02 W272
boris.tomas@tcd.ie

Abstract. *This paper presents an extensive bibliometric analysis of the impact of the language of publication on the visibility and citation of scientific papers in computer science and engineering, with a particular focus on Croatian. Using data from the OpenAlex database, papers were analysed by language and country of origin. The results show that English-language papers achieve several-fold higher citation rates and global visibility, while Croatian-language papers are limited to a regional audience and have negligible international impact. The analysis suggests that publishing in Croatian is not effective for researchers who aim to contribute to global science, especially in computer science and engineering. Possible causes are discussed, including the language barrier, limited accessibility, and potentially lower quality of papers published in Croatian.*

Keywords. language of publication, citations, computer science, engineering, bibliometrics, OpenAlex, Croatian language

Sažetak. *Ovaj rad donosi opsežnu bibliometrijsku analizu utjecaja jezika publikacije na vidljivost i citiranost znanstvenih radova u području računalnih znanosti i inženjerstva, s posebnim naglaskom na hrvatski jezik. Korištenjem podataka iz baze OpenAlex, analizirani su radovi prema jeziku i zemlji porijekla. Rezultati pokazuju da radovi na engleskom jeziku ostvaruju višestruko veću citiranost i globalnu vidljivost, dok su radovi na hrvatskom jeziku ograničeni na regionalnu publiku i imaju zanemariv međunarodni utjecaj. Analiza sugerira da objavljivanje na hrvatskom jeziku nije učinkovito za znanstvenike koji žele doprinijeti svjetskoj znanosti, osobito u računalnim znanostima i inženjerstvu. U radu se raspravlja o mogućim uzrocima, uključujući jezičnu barijeru, ograničenu dostupnost i potencijalno nižu kvalitetu radova na hrvatskom jeziku.*

Ključne riječi. jezik publikacije, citiranost, računalne znanosti, inženjerstvo, bibliometrija, OpenAlex, hrvatski jezik

1 Introduction

Computer science and engineering represent one of the most dynamic and internationalised areas of contemporary science (Douglas et al., 2010; Garousi & Mäntylä, 2016; Kelly, 2015; Rajkumar et al., 2024). Owing to intensive global collaboration, rapid knowledge exchange, and constant technological progress, the international visibility and citation of scientific papers are key indicators of the impact and relevance of researchers and institutions. In this context, the language of publication becomes a decisive factor for the global reach of research results. Although a significant num-

1 Uvod

Računalne znanosti i inženjerstvo predstavljaju jedno od najdinamičnijih i najinternacionaliziranijih područja suvremene znanosti (Douglas i dr., 2010; Garousi & Mäntylä, 2016; Kelly, 2015; Rajkumar i dr., 2024). Zbog intenzivne globalne suradnje, brze razmjene znanja i stalnog tehnološkog napretka, međunarodna vidljivost i citiranost znanstvenih radova ključni su pokazatelji utjecaja i relevantnosti istraživača i institucija. U tom kontekstu, jezik publikacije postaje odlučujući faktor za globalni doseg rezultata istraživanja. Iako se u Hrvatskoj i dalje objavljuje značajan broj radova

ber of papers in Croatia are still published in Croatian, particularly in domestic journals and conference proceedings, numerous bibliometric analyses indicate the limited international impact of such papers (Garousi & Mäntylä, 2016; Nurzhanov et al., 2021). The aim of this paper is to quantitatively analyse the relationship between language of publication and citation in computer science and engineering, using the latest data from the OpenAlex[1] database of scholarly works.

1.1 Description of the OpenAlex database and reasons for its use

OpenAlex is an open, global bibliographic database that indexes scientific papers, journals, authors, and institutions worldwide (OurResearch, 2024; Priem et al., 2022). Created in 2022 as a project of the nonprofit organisation OurResearch, it succeeded the Microsoft Academic Graph, and today contains more than 260 million works from all scientific fields, including computer science and engineering. Unlike commercial databases such as Web of Science and Scopus, OpenAlex is completely free, open, and transparent; all data and program code are available under an open licence. The database allows searching, analysis, and data export via a web interface, REST API, or by downloading periodic snapshots of the entire database. OpenAlex aggregates data from numerous sources (Crossref, ORCID, arXiv, PubMed, and others), covering a wide spectrum of publication types (articles, conference papers, books, or dissertations) and provides rich metadata, including citations, links between authors, institutions, topics, and sources of funding. In this study OpenAlex was used for several key reasons:

- **Openness and accessibility**: Unlike closed commercial databases, OpenAlex enables free access to and export of data without licensing restrictions, which allows full transparency and reproducibility of the analysis.
- **Wide coverage**: OpenAlex indexes a considerably larger number of works, including publications in different languages and from less represented regions, which is important for analysing national and regional patterns in science.
- **Rich and connected metadata**: The database supports linking works with authors, institutions, topics, and countries, which is crucial for advanced bibliometric analyses and visualisations of scientific output and collaboration. OpenAlex provides detailed classification by scientific fields, including computer science and engineering, and enables filtering and analysis by language, country, and publication type, which is necessary for this research.

[1] openalex.org

na hrvatskom jeziku, posebice u domaćim časopisima i zbornicima konferencija, rezultati brojnih bibliometrijskih analiza ukazuju na ograničen međunarodni utjecaj takvih radova (Garousi & Mäntylä, 2016; Nurzhanov i dr., 2021). Cilj ovog rada je kvantitativno analizirati odnos između jezika publikacije i citiranosti u području računalnih znanosti i inženjerstva, koristeći najnovije podatke iz OpenAlex[1] baze znanstvenih radova.

1.1 Opis baze podataka OpenAlex i razlozi korištenja

OpenAlex je otvorena, globalna bibliografska baza podataka koja indeksira znanstvene radove, časopise, autore i institucije iz cijeloga svijeta (OurResearch, 2024; Priem i dr., 2022). Nastala je 2022. godine kao projekt neprofitne organizacije OurResearch, nasljeđujući Microsoft Academic Graph, te danas sadrži više od 260 milijuna radova iz svih znanstvenih područja, uključujući računalne znanosti i inženjerstvo. Za razliku od komercijalnih baza poput Web of Science i Scopus, OpenAlex je potpuno besplatan, otvoren i transparentan; svi podaci i programski kod dostupni su pod otvorenom licencom. Baza omogućuje pretraživanje, analizu i izvoz podataka putem web sučelja, REST API-ja ili preuzimanjem periodičnih snimki cijele baze. OpenAlex agregira podatke iz brojnih izvora (Crossref, ORCID, arXiv, PubMed i drugi), pokrivajući širok spektar publikacija (članci, konferencijski radovi, knjige ili disertacije) i pruža bogate meta-podatke, uključujući citate, povezanost autora, institucija, tema i izvora financija. U ovom istraživanju OpenAlex je korišten iz nekoliko ključnih razloga:

- **Otvorenost i dostupnost**: Za razliku od zatvorenih komercijalnih baza, OpenAlex omogućuje slobodan pristup i izvoz podataka bez licencijskih ograničenja, što omogućuje potpunu transparentnost i ponovljivost analize.
- **Široka pokrivenost**: OpenAlex indeksira znatno veći broj radova, uključujući publikacije na različitim jezicima i iz manje zastupljenih regija, što je važno za analizu nacionalnih i regionalnih obrazaca u znanosti.
- **Bogati i povezani meta-podaci**: Baza omogućuje povezivanje radova s autorima, institucijama, temama i zemljama, što je ključno za napredne bibliometrijske analize i vizualizacije znanstvene produkcije i suradnje. OpenAlex pruža detaljnu klasifikaciju po znanstvenim područjima, uključujući računalne znanosti i inženjerstvo, te omogućuje filtriranje i analizu prema jeziku, zemlji i tipu publikacije, što je nužno za ovo istraživanje.

[1] openalex.org

Thanks to these characteristics, OpenAlex enabled a comprehensive, open, and reproducible bibliometric analysis with an emphasis on the impact of publication language and geographical origin of works.

1.2 Structure of the paper

The paper is organised as follows: the *Introduction* presents the motivation, aims, and relevance of the topic and describes the OpenAlex database. The *Review of previous research* summarises relevant literature and prior findings on the impact of language of publication on citation in computer science and engineering. The *Methodology* describes the data sources, data collection and processing, and the analytical methods used. The *Results* section presents the main findings, including the number of papers and citations by language, the share of Croatian-language works, the geographical origin of papers, and trends over time. The *Discussion* analyses and interprets the results, discusses possible causes and limitations, and provides recommendations. The *Conclusion* summarises the main conclusions and suggests directions for future research.

2 Review of previous research

Computer science and engineering are characterised by an exceptionally high level of internationalisation, which is supported by the fact that the leading journals, conferences, and repositories are almost exclusively in English (Douglas et al., 2010; Rajkumar et al., 2024). According to Kelly (Kelly, 2015), in computer science and software engineering a significant share of citations comes from conference papers and articles in English compared with other scientific fields. Garousi et al. (Garousi & Mäntylä, 2016) show that the language of publication significantly affects citation even within the same journals or disciplines. Papers in English have a several-fold higher probability of being cited, which is associated with greater accessibility, a broader reach, and better indexing in international databases. Similar trends are observed in engineering, where international collaboration increases impact (Brito et al., 2018; Yildirim et al., 2024).

3 Methodology

3.1 Data sources and processing

Data were retrieved from the OpenAlex database of scientific works, using an API query that enables grouping of results by language and country of origin. All works classified in computer science and engineering were analysed. For each language, the total number of works, the total number of citations, and the distribution of works by country were recorded. OpenAlex assigns a work's country based on the major-

Zahvaljujući navedenim karakteristikama, OpenAlex je omogućio provedbu opsežne, otvorene i ponovljive bibliometrijske analize s naglaskom na utjecaj jezika publikacije i zemljopisnog porijekla radova.

1.2 Struktura rada

Rad je organiziran na sljedeći način: *Uvod*, donosi motivaciju, ciljeve i relevantnost teme te opisuje korištenu bazu podataka OpenAlex. *Pregled dosadašnjih istraživanja* donosi pregled relevantne literature i dosadašnjih rezultata o utjecaju jezika publikacije na citiranost u računalnim znanostima i inženjerstvu. *Metodologija* opisuje izvore podataka, način prikupljanja i obrade podataka te korištene analitičke metode. *Rezultati* prikazuju glavne nalaze analize, uključujući broj radova i citiranost po jeziku, udio hrvatskog jezika, geografski izvor radova te trendove kroz vrijeme. *Diskusija* analizira i interpretira dobivene rezultate, raspravlja o mogućim uzrocima i ograničenjima te daje preporuke. *Zaključak* sažima glavne zaključke rada i predlaže smjernice za buduća istraživanja.

2 Pregled dosadašnjih istraživanja

Računalne znanosti i inženjerstvo karakterizira izuzetno visok stupanj internacionalizacije, čemu doprinosi i činjenica da su vodeći časopisi, konferencije i repozitoriji gotovo isključivo na engleskom jeziku (Douglas i dr., 2010; Rajkumar i dr., 2024). Prema Kelly (Kelly, 2015), u računalnim znanostima i softverskom inženjerstvu značajan broj citata dolazi iz konferencijskih radova i članaka na engleskom jeziku u odnosu na druga polja znanosti. Garousi i sur. (Garousi & Mäntylä, 2016) pokazuju da jezik publikacije značajno utječe na citiranost čak i unutar istih časopisa ili disciplina. Radovi na engleskom jeziku imaju višestruko veću vjerojatnost da budu citirani, što je povezano s većom dostupnošću, širim dosegom i boljom indeksiranošću u međunarodnim bazama. Slični trendovi uočeni su i u inženjerstvu, gdje međunarodna suradnja povećavaju utjecaj (Brito i dr., 2018; Yildirim i dr., 2024).

3 Metodologija

3.1 Izvori podataka i obrada

Podaci su preuzeti iz baze znanstvenih radova OpenAlex, koristeći API upit koji omogućuje grupiranje radova po jeziku i zemlji porijekla. Analizirani su svi radovi iz područja računalnih znanosti i inženjerstva. Za svaki jezik zabilježen je ukupan broj radova, ukupan broj citata te distribucija radova po zemljama. OpenAlex dodjeljuje zemlju rada na temelju porijekla većine autora ili na temelju zemlje porijekla prvog autora rada.

ity of authors' affiliations or, alternatively, the first author's country of origin.

3.2 Examples of code used for the analysis

For the purposes of this research, bespoke tools and scripts were used to download, process, and analyse data from OpenAlex database. The code snippets shown provide transparent insight into the methods used and facilitate the reproducibility and extensibility of the analysis. Below are the key parts of the code used to retrieve data on the language of publication, number of works, citations, and geographical origin, as well as to perform relevant statistical analyses.

The result of executing Code 1 is a file with all languages in the OpenAlex database, including the number of works written in the corresponding language. These data are available in the open repository here: *github.com/boristomas/ R1Podaci/blob/main/jezici.csv*

The result of executing Code 2 is a file with data on each language, including citations and fields. These data are available in the open repository here: *github.com/boristomas/R1Podaci/blob/main/podaci.csv* In both code snippets, field identifiers 17 and 22 denote computer science and engineering.

4 Results

4.1 Number of papers by language in computer science and engineering

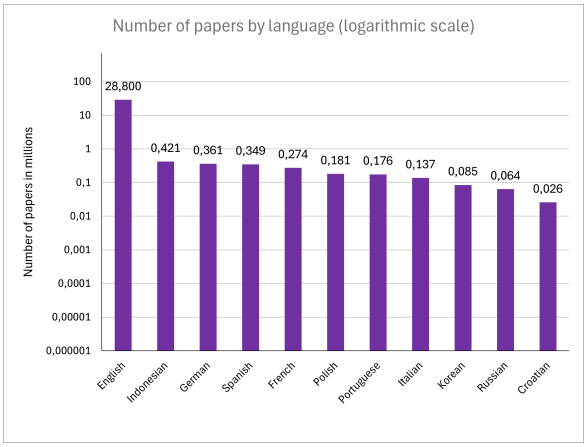


Figure 1. Number of papers by language (top 10 languages and Croatian) in computer science and engineering.

Fig. 1 shows the number of papers by language for the ten most represented languages in computer science and engineering, including Croatian as the eleventh (last column). It is clearly visible that English absolutely dominates, with more than 28 million papers, while all other languages trail significantly, including Croatian (25,810 papers). Croatian ranks 11th by number

3.2 Primjeri kôda korištenih za analizu

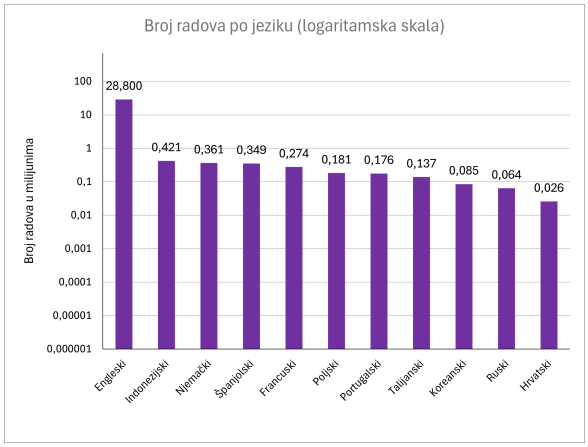
Za potrebe ovog istraživanja korišteni su vlastiti programski alati i skripte za preuzimanje, obradu i analizu podataka iz baze znanstvenih radova OpenAlex. Prikazani primjeri kôda omogućuju transparentan uvid u korištene metode te olakšavaju ponovljivost i proširivost analize. U nastavku su detaljno opisani ključni dijelovi kôda koji su korišteni za dohvat podataka o jeziku publikacije, broju radova, citiranosti i zemljopisnom podrijetlu radova, kao i za izvođenje relevantnih statističkih analiza.

Rezultat izvršavanja kôda 1 je datoteka sa svim jezicima u OpenAlex bazi podataka uključujući i broj radova napisanih na odgovarajućem jeziku. Ovi podaci su dostupni na otvorenom repozitoriju ovdje: *github.com/boristomas/R1Podaci/blob/main/jezici.csv*

Rezultat izvršavanja Kôda 2 je datoteka s podacima o svakom jeziku o citiranosti i područjima. Ovi podaci su dostupni na otvorenom repozitoriju ovdje: *github.com/boristomas/R1Podaci/blob/main/podaci.csv* U oba kôda, oznaka polja 17 i 22 su za računalne znanosti i inženjerstvo.

4 Rezultati

4.1 Broj radova po jeziku u računalnim znanostima i inženjerstvu



Slika 1. Broj radova po jeziku (top 10 jezika i hrvatski jezik) u računalnim znanostima i inženjerstvu.

Na slici 1 prikazan je broj radova po jeziku za deset najzastupljenijih jezika u području računalnih znanosti i inženjerstva, uključujući i hrvatski jezik kao jedanaesti (zadnji stupac). Jasno je vidljivo da engleski jezik apsolutno prevladava, s više od 28 milijuna radova, dok su svi ostali jezici značajno iza, uključujući i hrvatski jezik (25810 radova). Hrvatski jezik je 11. po broju

of papers among all languages in the database (55). For readability, a logarithmic scale was used; absolute counts are printed above each column.

4.2 Citation trends by language

The analysis of average citation by language, shown in Fig 2, further confirms the dominant position of English. English has an average of 11.2 citations per paper, whereas Croatian has only 0.42, more than 26 times lower. A higher citation rate is notable for French-language papers, which is likely a result of the relatively small proportion of proficient English speakers in the Francophone world. According to (First, 2025), France ranks only 49th in English proficiency, while Croatia is in a high 5th place.

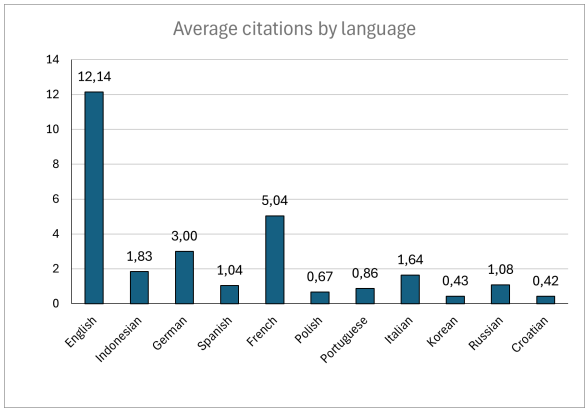


Figure 2. Average citations per paper by language (top 10 languages and Croatian) in computer science and engineering.

4.3 Share of Croatian in the total number of papers and citations

The share of Croatian-language papers in the total number of papers is just 0.08% (25810 papers), while the share of citations is only 0.003% (10838 citations) (Fig 3).

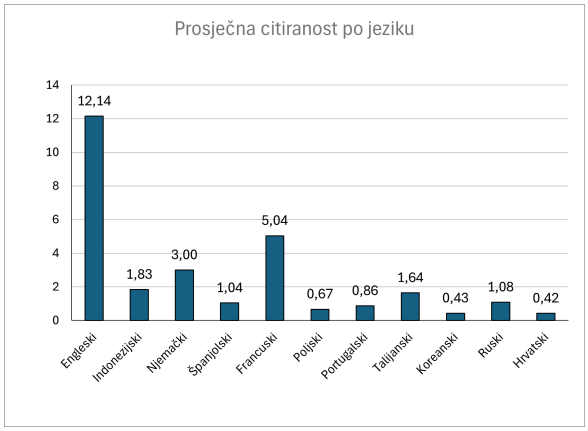
4.4 Countries of origin of Croatian-language papers

Table 1 shows the top 20 countries from which Croatian-language papers originate. Although most works are from Croatia, a significant number come from Serbia, Bosnia and Herzegovina, and Slovenia, but also from countries outside the region such as the USA, Germany, the UK, and China.

radova od svih jezika u bazi podataka (55). Zbog lakše čitljivosti korištena je logaritamska skala, broj radova je izražen kao apsolutni broj iznad svakog stupca.

4.2 Trendovi citiranosti po jeziku

Analiza prosječne citiranosti po jeziku, prikazana na slici 2, dodatno potvrđuje dominantnu poziciju engleskog jezika. Engleski jezik ima prosječnu citiranost od 11.2 po radu, dok hrvatski ima samo 0.42, što je više od 26 puta manje. Primjetna je veća citiranost radova na francuskom jeziku što je vjerojatno rezultat malog udjela stanovnika koji znaju engleski jezik u francuskom govornom području. Prema (First, 2025) Francuska je tek 49. zemlja po broju i kvaliteti govornika na engleskom jeziku, a Hrvatska je na visokom 5. mjestu.



Slika 2. Prosječna citiranost po jeziku (top 10 jezika i hrvatski jezik) u računalnim znanostima i inženjerstvu.

4.3 Udio hrvatskog jezika u ukupnom broju radova i citata

Udio radova na hrvatskom jeziku u ukupnom broju radova iznosi samo 0.08% (25810 radova), dok udio citata iznosi tek 0.003% (10838 citata) (Slika 3).

4.4 Zemlje izvora radova na hrvatskom jeziku

Tablica 1 prikazuje prvih 20 zemalja iz kojih dolaze radovi na hrvatskom jeziku. Iako je većina radova iz Hrvatske, značajan broj dolazi iz Srbije, Bosne i Hercegovine, Slovenije, ali i iz zemalja izvan regije poput SAD-a, Njemačke, UK-a i Kine.

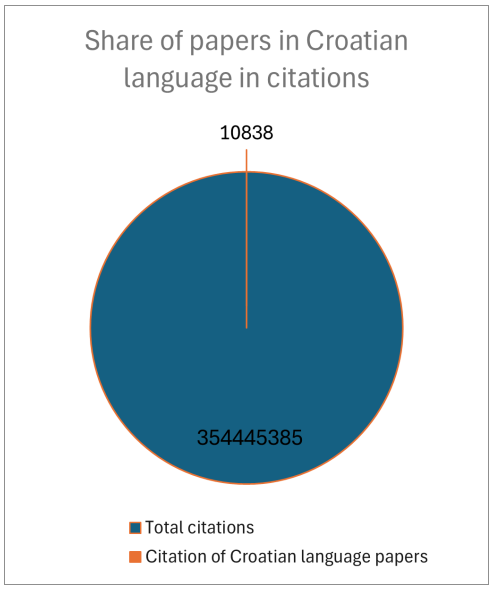
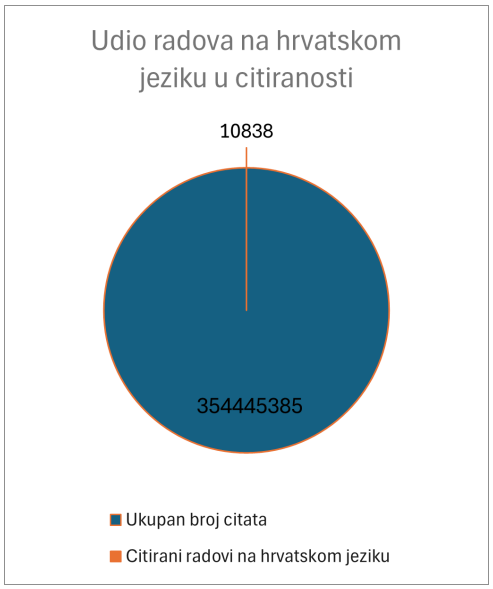


Figure 3. Share of Croatian-language papers in the total number of citations.



Slika 3. Udio radova na hrvatskom jeziku u ukupnom broju citata.

Table 1. Top 20 countries of origin of Croatian-language papers

Country	Number of papers
Croatia	*2,495*
Serbia	683
Bosnia and Herzegovina	221
Slovenia	171
USA	91
Czechia	60
Poland	53
Slovakia	53
Türkiye	52
Germany	46
UK	43
Ukraine	31
Austria	28
China	26
Hungary	25
Spain	24
Peru	19
France	18
Italy	18
Montenegro	18
Russia	18

Tablica 1. Prvih 20 zemalja iz kojih dolaze radovi na hrvatskom jeziku

Zemlja	Broj radova
Hrvatska	*2,495*
Srbija	683
Bosna i Hercegovina	221
Slovenija	171
SAD	91
Češka	60
Poljska	53
Slovačka	53
Turska	52
Njemačka	46
UK	43
Ukrajina	31
Austrija	28
Kina	26
Mađarska	25
Španjolska	24
Peru	19
Francuska	18
Italija	18
Crna Gora	18
Rusija	18

4.5 Trends in publications and citations over time

Over the last 15 years, the number of English-language papers in computer science and engineering has been continuously growing, while the number of Croatian-language papers is stagnating or declining (Fig 4). Citation rates for English-language papers also show an upward trend, while Croatian-language papers remain

4.5 Analiza trendova publikacija i citiranosti kroz vrijeme

Kroz posljednjih 15 godina, broj radova na engleskom jeziku u računalnim znanostima i inženjerstvu kontinuirano raste, dok broj radova na hrvatskom jeziku stagnira ili opada (Slika 4). Citiranost radova na engleskom jeziku također pokazuje uzlazni trend, dok radovi na hrvatskom jeziku ostaju na vrlo niskim vrijednostima.

very low. These trends further emphasise the importance of the choice of publication language for international visibility and impact.

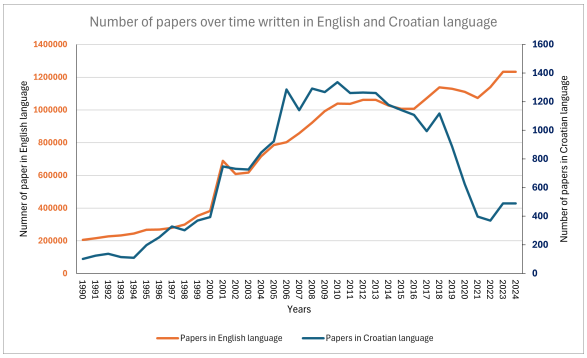
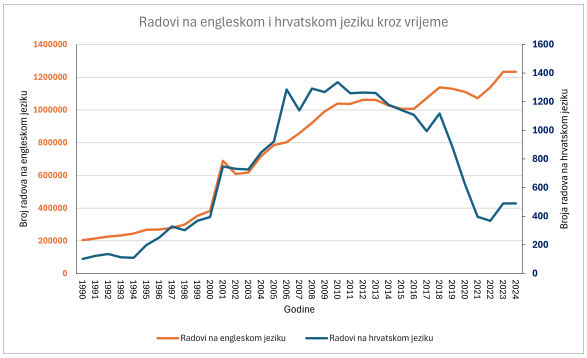


Figure 4. Papers published in English and Croatian over time.

Ovi trendovi dodatno naglašavaju važnost odabira jezika publikacije za međunarodnu vidljivost i utjecaj.



Slika 4. Objavljeni radovi na engleskom i hrvatskom jeziku kroz vrijeme.

5 Discussion

The results clearly show that English is not only dominant in terms of the number of papers and citations, but also in the international dispersion of authors and institutions. The data show an interesting and significant contribution of Indonesian, which stems from OpenAlex's strategy to minimise geo-discrimination. It is also noteworthy that Chinese (Mandarin) does not dominate by number of works. The reason lies in the fact that the Chinese government invests in and financially supports publication of papers in English (MoChridhe, n.d.).
China has recognised the importance of English in science and, through such measures, strives to ensure and maintain its leading scientific position. Croatian, although with a smaller total number of works, shows an interesting regional and diaspora component. A significant number of Croatian-language papers come from Serbia, Bosnia and Herzegovina, and Slovenia, indicating a shared scientific space across the former Yugoslavia. Works from countries outside the Croatian language's regional environment (for example, China) are an interesting phenomenon that may be the result of university promotion policies. However, the average citation of Croatian-language papers remains low, suggesting limited international visibility. This limitation may be due to the smaller number of journals indexed in international databases, the language barrier, and perceptions regarding the importance of national languages in science. The analysis suggests that publishing in Croatian is less effective for researchers aiming to contribute to global science, especially in computer science and engineering. This analysis did not include a qualitative assessment of the papers, but it can be assumed that works published in Croatian are of less general interest, as indicated by lower citation. However, a further reason may simply be that Croatian-language papers are accessible to a smaller population

5 Diskusija

Rezultati jasno pokazuju da je engleski jezik ne samo dominantan po broju radova i citata, već i po međunarodnoj disperziji autora i institucija. Analiza podataka pokazuje zanimljiv i značajan doprinos indonezijskog jezika, razlog tome leži u strategiji OpenAlex organizacije da minimizira geo-diskriminaciju. Također je interesantna činjenica da kineski jezik (mandarinski) ne dominira u broju radova. Razlog tome leži u činjenici da kineska vlada ulaže i financijski podupire objavu radova na engleskom jeziku (MoChridhe, bez datuma). Kina je prepoznala važnost engleskog jezika u znanosti te svojim potezima nastoji osigurati i zadržati vodeću poziciju u znanosti. Hrvatski jezik, iako s manjim ukupnim brojem radova, pokazuje zanimljivu regionalnu i dijasporsku komponentu. Značajan broj radova na hrvatskom jeziku dolazi iz Srbije, Bosne i Hercegovine i Slovenije, što ukazuje na postojanje zajedničkog znanstvenog prostora na području bivše Jugoslavije. Radovi iz zemalja koje su izvan regionalnog okruženja hrvatskog jezik, poput na primjer Kina su interesantna pojava koja je možda rezultat sveučilišnih politika uvjetovanja za napredovanja. Međutim, prosječna citiranost radova na hrvatskom jeziku ostaje niska, što sugerira da je njihova međunarodna vidljivost ograničena. Ova ograničenost može biti posljedica manjeg broja časopisa indeksiranih u međunarodnim bazama, jezične barijere, ali i percepcije o važnosti nacionalnih jezika u znanosti. Analiza sugerira da objavljivanje na hrvatskom jeziku manje učinkovito za znanstvenike koji žele doprinijeti svjetskoj znanosti, osobito u računalnim znanostima i inženjerstvu. Analiza nije uključivala kvalitativnu procjenu radova, no može se pretpostaviti da su radovi objavljeni na hrvatskom jeziku manje interesantni, što je pokazano i manjom citiranošću. Međutim, razlog manje citiranosti može ležati i u činjenici da su radovi na hrvatskom jeziku dos-

of researchers who understand Croatian, which reduces the likelihood of being cited. Limited availability and the language barrier are key factors influencing this result, consistent with other studies that examine differences in citation practices between English and Croatian authors (Varga & Gradečak-Erdeljić, 2017). A potentially significant source of error in this research is the limited dataset within the OpenAlex database.

6 Conclusion

The analysis clearly shows that, in computer science and engineering, publishing in English is essential in order to achieve international visibility and scientific impact. Croatian-language papers remain restricted to the regional scientific community and have a limited contribution to global science. It is recommended that researchers in these fields publish their results in English, particularly if they aim to contribute to the global scientific community. The national language may make sense for professional, educational, or popular science works, but not for scientific publications that seek international relevance. A particularly unfavourable effect on the internationalisation of Croatian science is the requirement, in some Croatian universities' promotion criteria, to publish in Croatian. Such requirements significantly hinder the career plans of visiting scientists who do not speak Croatian but wish to spend part of their scientific career in the Republic of Croatia. Future research should analyse the quality of language in Croatian-language papers produced outside the regional environment; one possible hypothesis is that such papers are originally written in another language and machine-translated into Croatian. Furthermore, future research should analyse in more detail the impact of international collaboration and the diaspora on Croatian-language scientific production, as well as possibilities for improving the visibility of domestic journals through open access and inclusion in international databases.

tupni manjoj populaciji znanstvenika koji razumiju hrvatski jezik pa samim time je manje vjerojatno da će biti citirani. Ograničenost dostupnosti i jezična barijera ključni su faktori koji utječu na ovaj rezultat, što je u skladu s nalazima drugih studija koje analiziraju razlike u citatnim praksama između engleskih i hrvatskih autora (Varga & Gradečak-Erdeljić, 2017). Potencijalno značajan izvor pogreške u ovom istraživanju može biti i ograničen set podataka u OpenAlex bazi podataka.

6 Zaključak

Analiza rada jasno pokazuje da je u području računalnih znanosti i inženjerstva objavljivanje na engleskom jeziku nužno za postizanje međunarodne vidljivosti i znanstvenog utjecaja. Radovi na hrvatskom jeziku ostaju ograničeni na regionalnu znanstvenu zajednicu te imaju ograničen doprinos svjetskoj znanosti. Preporučuje se da znanstvenici iz ovih područja svoje rezultate objavljuju na engleskom jeziku, osobito ako im je cilj doprinijeti globalnoj znanstvenoj zajednici. Nacionalni jezik može imati smisla za stručne, nastavne ili popularno-znanstvene radove, ali ne i za znanstvene publikacije koje ciljaju na međunarodnu relevantnost. Naročito nepovoljan efekt na internacionalizaciju hrvatske znanosti ima uvjetovanje napredovanja znanstvenika na sveučilištima u Republici Hrvatskoj objavom radova na hrvatskom jeziku. To uvjetovanje značajno otežava karijerni plan gostujućih znanstvenika koji ne znaju hrvatski jezik, a žele dio znanstvene karijere provesti u Republici Hrvatskoj. Buduća istraživanja bi trebala analizirati kvalitetu jezika kod radova na hrvatskom jeziku koji su napravljeni izvan regionalnog okruženja; moguća pretpostavka je da se radi o radovima koji su napisani originalnim jezikom te su strojno prevedeni na hrvatski jezik. Nadalje, buduća istraživanja trebala bi i detaljnije analizirati utjecaj međunarodne suradnje i dijaspore na znanstvenu produkciju na hrvatskom jeziku, kao i mogućnosti za unaprijeđenje vidljivosti domaćih časopisa kroz otvoreni pristup i uključivanje u međunarodne baze podataka.

References / Literatura

Brito, C., Ciampi, M., Santos, H., Vasconcelos, R., & Amaral, L. (2018, June). *Internationalization and Its Impact in Engineering Programs* (tech. rep.). EasyChair. https://doi.org/10.29007/jx9r

Douglas, S., Farley, A., Lo, G., Proskurowski, A., & Young, M. (2010). Internationalization of computer science education. *Proceedings of the 41st ACM technical symposium on Computer science education*, 411–415. https://doi.org/10.1145/1734263.1734404

First, E. (2025). EF EPI - EF English Proficiency Index. A Ranking of 116 Countries and Regions by English Skills. https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2024/ef-epi-2024-english.pdf

Garousi, V., & Mäntylä, M. V. (2016). Citations, research topics and active countries in software engineering: A bibliometrics study. *Computer Science Review*, *19*, 56–77. https://doi.org/10.1016/j.cosrev.2015.12.002

Kelly, M. (2015). Citation Patterns of Engineering, Statistics, and Computer Science Researchers: An Internal and External Citation Analysis across Multiple Engineering Subfields. *College Research Libraries*, *76*, 859–882. https://doi.org/10.5860/crl.76.7.859

MoChridhe, R. (n.d.). The Hidden Language Policy of China's Research Evaluation Reform. https://blog.westminster.ac.uk/contemporarychina/the-hidden-language-policy-of-chinas-research-evaluation-reform/

Nurzhanov, C., Pidlisnyuk, V., Naizabayeva, L., & Satymbekov, M. (2021). Research and trends in computer science and educational technology during 2016–2020: Results of a content analysis. *World Journal on Educational Technology: Current Issues*, *13*, 115–128. https://doi.org/10.18844/wjet.v13i1.5421

OurResearch. (2024). OpenAlex Documentation [Available at: https://docs.openalex.org/].

Priem, J., Piwowar, H., Orr, R., et al. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts [Available at: https://openalex.org/about].

Rajkumar, S., Sujay, D., Sangeetha, R., Sudhakar, G., Muralikrishna, K., & Kumaraswamy, D. (2024). An Overview of Computer Science and Engineering and Its Latest Technologies. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, *2*, 683–698. https://doi.org/10.47392/IRJAEH.2024.0099

Varga, M., & Gradečak-Erdeljić, T. (2017). English and Croatian citation practices in research articles in applied linguistics: a corpus-based study. *Kalbotyra*, *70*, 153–183. https://doi.org/10.15388/Klbt.2017.11199

Yildirim, G., Pan, X., & Rahman, A. (2024). Fluid Mechanics Education in Engineering: Challenges and Trends from a Bibliometric Analysis. *Proceedings of the 3rd International Conference on Advancements in Engineering Education*. https://doi.org/10.71427/iCAEED2025/24

Appendix / Prilozi

Appendix 1 / Prilog 1

Code 1. Analysis of languages and countries of origin / **Kod 1:** Analiza jezika i zemalja porijekla

```
import requests
import pandas as pd
from urllib.parse import urlparse
API_URL = "https://api.openalex.org/works?group_by=
language&per_page=200&filter=type:types/
article,primary_topic.field.id:
fields/17|fields/22"
def extract_language_code(key_url):
    if not key_url:
        return "unknown"
    parsed = urlparse(key_url)
    return parsed.path.split('/')[-1] if '/' in parsed.path else parsed.path
def fetch_language_groups(api_url):
    response = requests.get(api_url)
    response.raise_for_status()
    data = response.json()
    groups = data.get('group_by', [])
    results = []
    for group in groups:
        lang_code = extract_language_code(
         group.get('key'))
        lang_name = group.get('key_display_name', 'Unknown')
        count = group.get('count', 0)
        results.append({
            'language_name': lang_name,
            'language_code': lang_code,
            'count': count
        })
    return results
def main():
    language_data = fetch_language_groups(API_URL)
   df = pd.DataFrame(language_data)
    print(df.to_string(index=False))
```

Appendix 2 / Prilog 2

Code 2. Data analysis / **Kod 2:** Analiza podataka

```
import pandas as pd
import time
import requests

languages_df = pd.read_csv('jezici.csv', sep=';')
def get_works_and_citations(language_code):
    base_url = "https://api.openalex.org/works"
    params = {
        "filter": f"type:types/article,primary_topic.field.id:fields/17|fields/22,
                    language:languages/{language_code}",
        "sort": "publication_year:desc",
        "per_page": 200,
        "apc_sum": "false",
        "cited_by_count_sum": "true"
    }
    response = requests.get(base_url, params=params)
    if response.status_code != 200:
        print(f"Error fetching data for {language_code}: {response.status_code}")
        return 0, 0
    data = response.json()
    total_works = data.get('meta', {}).get('count', 0)
    total_citations = data.get('meta', {}).get('cited_by_count_sum', 0)
    return total_works, total_citations
def get_country_distribution(language_code):
    base_url = "https://api.openalex.org/works"
    params = {
        "group_by": "authorships.countries",
        "filter": f"type:types/article,primary_topic.field.id:fields/17|fields/22,
                    language:languages/{language_code}",
        "per_page": 200
    }
    response = requests.get(base_url, params=params)
    if response.status_code != 200:
        print(f"Error fetching data for {language_code}: {response.status_code}")
        return {}
    data = response.json()
    country_groups = data.get('group_by', [])
    country_distribution = {group['key_display_name']: group['count'] for group in
        country_groups}
    return country_distribution
results = []
for idx, row in languages_df.iterrows():
    lang_name = row['language_name']
    lang_code = row['language_code']
    if not isinstance(lang_code, str) or lang_code.lower() in ['none', '', 'nan']:
        print(f"Skipping {lang_name}: invalid language code.")
        continue
    print(f"Processing {lang_name} ({lang_code})...")
    works, citations = get_works_and_citations(lang_code)
    countries = get_country_distribution(lang_code)
    results.append({
        "language_name": lang_name,
        "language_code": lang_code,
        "total_works": works,
        "total_citations": citations,
        "countries": countries
    })
    time.sleep(1)
results_df = pd.DataFrame(results)
results_df.to_csv("podaci.csv", index=False)
```

Dark Patterns in the Public Sector: A Study of Cookie Consent Interfaces on Croatian Websites

Marija Kuštelega
University of Zagreb Faculty of Organization and Informatics
Department of Information Systems Development
Pavlinska 2, 42000 Varaždin, Croatia
marija.kustelega@foi.unizg.hr

Abstract. *With the implementation of the General Data Protection Regulation (GDPR), it is now required to obtain explicit user consent before using any data about them. The way cookie consents are designed can directly affect an individual's ability to make an informed privacy decision. Because of this, a lot of websites use dark patterns, a term described as a visual pattern used to trick users into choosing options that are not in their best interests. To evaluate the presence of dark patterns in the public sector, a sample of the 100 most popular Croatian public sector websites, based on the Tranco list, was analyzed. This analysis found that over 96% of websites with cookie consent used some form of dark patterns. Although most of them have all the options listed that they should have according to the legal obligations, not all options are easily accessible, such as the ability to reject the consent with a single click. Providing users with easily understandable information, making it simple to accept or reject cookies, and allowing users to easily change their preferences are all components of good cookie consent design.*

Keywords. cookie consent, dark patterns, privacy

1 Introduction

Gartner hype-cycle for emerging technologies for 2024 has ranked delivering human-centric security and privacy as one of the four main themes (Gartner, 2024). This demonstrates how the focus is shifting from quick solution delivery to safe solution implementation, highlighting the importance of incorporating security and privacy measures into user interface design. One of the related initiatives is usable privacy and security research, focused on designing usable privacy interfaces to help users make better privacy decisions (Distler et al., 2021). Usable privacy researchers deal with various segments of privacy-related topics in the field of human-computer interaction, one of which is the provision of usable privacy notices (Fischer-Hübner and Karegar, 2024). Because dark patterns directly contradict these core principles, they have thus emerged as a highly interesting research topic.

Dark patterns, first introduced by Henry Brignull in 2010, attempted to explain how websites deceive users through their user interface (Alharbi, Albesher, and Wahsheh, 2023). It defines that: “Deceptive design patterns (also known as ‘dark patterns’) are tricks used in websites and apps that make you do things that you didn’t mean to, like buying or signing up for something” (Brignull et al., 2023).

Because dark patterns draw attention away from ethics, transparency, and fair business practices, they pose a serious threat to e-commerce websites in particular. Moreover, the European Directive 2005/29/EC on unfair commercial practices has recognized the dark pattern as a potential threat in the form of manipulative practices (European Commission, 2022). Other organizations, such as the Organization for Economic Co-operation and Development (OECD), also deal with this topic by trying to create a definition and taxonomy of dark patterns in order to be able to recognize the characteristics of dark commercial patterns (OECD, 2022).

However, other areas besides retail are also affected; more than 90% of 243 e-government websites have dark patterns in their user interfaces (Alharbi et al., 2023). Generally, dark patterns are common in European Union (EU) countries, where at least one dark pattern was present in 73 of 75 websites and apps studied (European Commission, 2022). A newer study conducted in June 2024 by the International Consumer Protection and Enforcement Network (ICPEN) showed that 76% of 642 websites and apps worldwide had some form of dark pattern, while nearly 67% used multiple combinations of dark patterns (ICPEN, 2024). Such data show that the situation has improved recently, but there is still a strong presence of dark patterns globally.

Usually, research was focused on popular domains such as e-commerce websites and news outlets (Mathur et al., 2019; Soe et al., 2020). There is a lack of research focused on public sector websites.

Public sector websites are expected to protect user privacy in order to uphold the integrity and sovereignty of public institutions in relation to private sector companies (Sørensen and Kosta, 2019). The aim of this research was to examine the extent to which public sector websites in Croatia employ dark patterns (manipulative design practices) to obtain user consent. The novelty of this research lies in the fact that no prior study has specifically examined public sector websites in this context. Research questions were as follows:

- Which categories of dark patterns are most commonly used in cookie consent interfaces present on public sector websites in Croatia?
- What additional characteristics of cookie consent interfaces should be improved to enhance user control (e.g., number of options, clicks needed to reject consent)?

The structure of this paper is as follows: Section 2 explains background and related work; Section 3 describes the utilized methodology for sampling websites; Section 4 presents the results of the analyzed dark patterns; Section 5 discusses the findings and concludes the work.

2 Background

2.1 Dark Patterns Categories

Dark patterns in terms of cookie consent interfaces can be described as any design practices that can lead users to unknowingly consent to data collection, which is not in line with their privacy preferences (Kitkowska, 2023). Colin Gray has made a comprehensive classification of dark patterns, putting them into five key categories: (1) Nagging, (2) Obstruction, (3) Sneaking, (4) Interface interference, and (5) Forced Action (Gray et al., 2018). In Table 1, key categories of dark patterns are briefly described with examples.

Table 1. Classification of dark pattern categories (Gray et al., 2018)

Dark pattern	Description	Example
Nagging	The user's desired task is interrupted by unrelated tasks.	Pop-ups, audio notices
Obstruction	The interaction is being made more challenging than it should be.	Unequal paths
Sneaking	An attempt to hide, mask, or delay the disclosure of relevant information.	Delayed disclosure of certain costs
Interface interference	Any kind of interface manipulation that favors certain actions over others.	Highlighted buttons, discolored text
Forced action	Any situation where the user needs to perform a certain action in order to proceed.	Required action (cannot be canceled)

In analysis, emphasis was on analyzing five main dark patterns, following described classification. Since several authors have done research based on these categories, they were chosen as a basis for evaluating the websites. Additionally, a similar taxonomy was followed by the OECD (2022), which further strengthened the reliance on these theoretical descriptions of the dark pattern.

Dark patterns are successful due to their reliance on human psychology, which involves common psychological tricks to compel users to take specific actions, such as the default effect, status quo, and framing (Jakobi and von Grafenstein, 2023). Related work on dark patterns has shown that interface interference and obstruction were the most common categories, while nagging was less frequently observed in news outlets (Soe et al., 2020). Although interface interference patterns and obstruction were present in the shopping category, Mathur et al. (2019) found that manipulations relying on language and emotional appeal were more common.

Similar work has been done by authors who manually evaluated the presence of dark patterns in websites (Mehrnezhad, 2020; Soe et al., 2020), as well as more recent initiatives that attempt to automatically detect dark patterns (Kirkman, Vaniea, and Woods, 2023). Recent work has focused on exploring the impact of different cookie consent designs on user decision making (Berens et al., 2024; Bielova et al., 2024), putting effort into developing bright patterns as a solution to exploit psychological tricks to the user's advantage.

2.2 Cookie consent

Due to the notice and choice mechanism present in cookie consents, the use of dark patterns is very popular among them (Habib et al., 2022). According to the GDPR, the term “consent” means any freely given, specific, informed, and unambiguous indication of a user’s desire (Alharbi et al., 2023). Some dark patterns can be considered as a violation of the GDPR, although many sophisticated patterns fall into the "gray" zone, i.e., are not expressly regulated. Research conducted shortly after the GDPR was introduced in 2019 showed that 99% of 300 websites with Scandinavian or English content didn’t allow users to refuse consent with a single click (Hu and Sastry, 2019). This violates the right of an individual to have the option to choose to refuse a cookie.

A significant increase in the number of cookies was found immediately after the introduction of GDPR, with 80% of the top 100 sites in the United Kingdom having implemented some form of cookie in 2019 (Hu and Sastry, 2019). A recent study from 2023 showed that the situation remains largely unchanged; in fact, looking at the global level, 50% of websites still do not have cookie consent, although the situation is better in European countries (Alharbi et al., 2023). In particular, one study discovered that websites outside of the EU

have less well-executed cookie design interfaces, giving consumers fewer options to make well-informed privacy decisions (Hu and Sastry, 2019). According to a survey of 3,947 respondents, users have developed a default cookie acceptance habit, which makes it even more important to offer a simple way to refuse consent (Bielova et al., 2024).

3 Methodology

3.1. Selection of website sample

As an initial sample, websites from the Tranco list (ID VQ2VN) were taken (Pochat et al., 2018). The Tranco list was chosen because it uses data from five different sources to appropriately rank the most popular websites. The website rankings were taken for a period of one month (from 03 May 2025 to 01 June 2025). An additional advantage of this list is that it only studies top-level domains, excluding their subdomains, which usually have the same cookie consent design. Also, it has been used in a large number of works, and it has an ID that can be used to easily track the list of web pages, which improves the research's reproducibility.

3.2. Identifying website categories

From the Tranco list, websites with the ".hr" domain were chosen, as this stands for Croatia's country code top-level domain. Since the term public sector websites includes a number of websites, the websites belonging to the public sector are identified first. Accordingly, websites were included in the analysis if they belonged to one of the three categories listed:

1) **Government bodies and administration websites** - this included government, ministries, agencies, offices, and counties
2) **Public institutions websites** - this included institutes, agencies, centers, universities, hospitals, and museums
3) **State-owned companies websites** - organizations that are majority owned by the state (e.g. post, transport, etc.)

3.3. Procedure

After the sample of websites was chosen, each website was analyzed to determine the presence of dark patterns. Prior to the analysis, a structured form was prepared to standardize the evaluation process. This form included information such as the website category, the presence of cookie banners, the position of the cookie banner, the number of options available to users, the applicable dark pattern categories, and other relevant characteristics. To ensure consistency in the website analysis process, the identification of dark pattern categories was guided by the examples and classification framework outlined in the theoretical section of this paper, based on the work of Gray et al. (2018).

Each website was manually reviewed by visiting its homepage from a web browser in private (incognito) mode. This approach was employed to ensure that prior browsing history or stored cookies would not affect the results of the analysis, such as the presence of cookie consent banners. A detailed checklist based on the standardized form with dark pattern categories and other observed characteristics was completed for each website. During the evaluation, cookie consent screens were taken so that they could be analyzed later and utilized to clarify particular dark patterns. Furthermore, in order to ensure that only legitimate public sector websites were included in the sample, each website was manually verified to confirm that it belonged to one of the following categories: government bodies and administration websites, public institutions websites, or state-owned companies websites.

4 Results

Following a detailed website selection procedure, 100 websites from the Tranco list were selected to provide a sample of Croatian public sector websites. The analysis revealed that 75% (75 out of 100) of the examined websites had a cookie consent interface, while the remaining 25% (25 out of 100) did not have any form of cookie consent implemented. The examination of dark patterns proceeded only with those websites that had implemented a cookie consent interface.

Out of the 75 websites that had cookie consent, 96% (72 of them) had some form of dark pattern, while 4% (3 of them) had no dark pattern present. Fig. 1 shows the overall frequency of dark pattern categories identified in cookie consent notices. The most frequently observed dark pattern categories were interface interference and obstruction, each identified on 28 websites, followed by forced action on 25 websites, and nagging, which appeared on 4 websites.

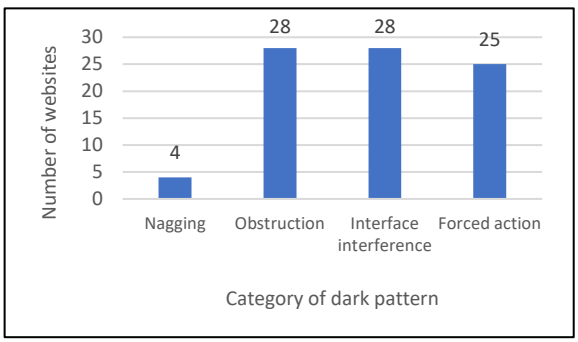


Figure 1. Frequency of dark pattern categories observed in cookie consent interfaces on Croatian public sector websites (N=75)

The interface interference category comprises four key subcategories: (1) aesthetic manipulation, (2) false hierarchy, (3) preselection, and (4) hidden information (Gray et al., 2018). Fig. 2 shows the distribution of specific subcategories within the interface interference category (N = 28), where aesthetic manipulation by colour was the most frequent subcategory (20 websites), to a lesser extent were represented false hierarchy (8 websites), hidden information (7 websites), and preselection (4 websites) subcategory.

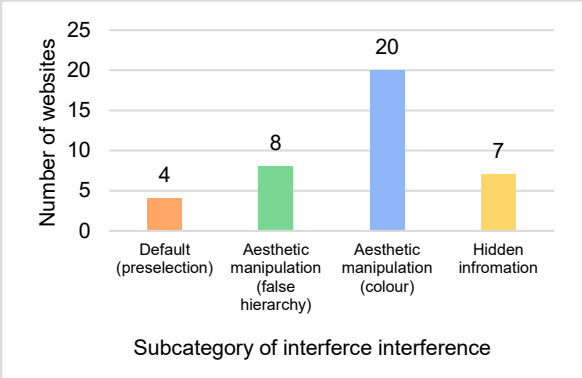


Figure 2. Frequency of interface interference subcategories observed in cookie consent interfaces on Croatian public sector websites (N =28)

4.1 Website Category

The analysis also examined the number of dark patterns present within each category of public sector websites. Fig. 3 illustrates the distribution of dark pattern categories across various types of public sector websites.

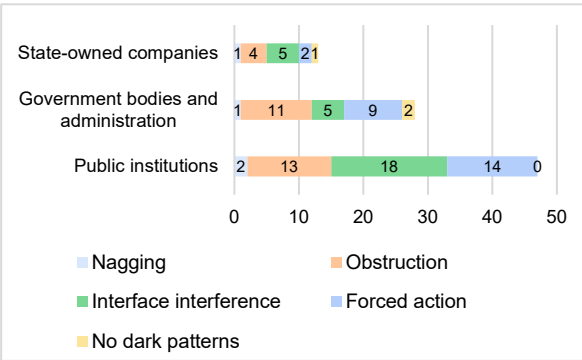


Figure 3. Frequency of dark pattern categories across different types of Croatian public sector websites (N = 75)

As shown in Fig. 3, nagging, forced action, interface interference, and obstruction were present in all three categories. The most frequent pattern in state-owned companies and public institution websites was interface interference, while obstruction was mostly present in government bodies and administration websites (in 11 examined websites).

4.2 Cookie Banner Position

The analysis also examined the positional placement of cookie banners on each website. Fig. 4 shows that cookie banners were predominantly positioned in the footer of the website (63 of them), less frequently in the middle (8 of them), and rarely in the header of the websites (4 of them). Additionally, several websites did not have any cookie banner implemented (25 of them).

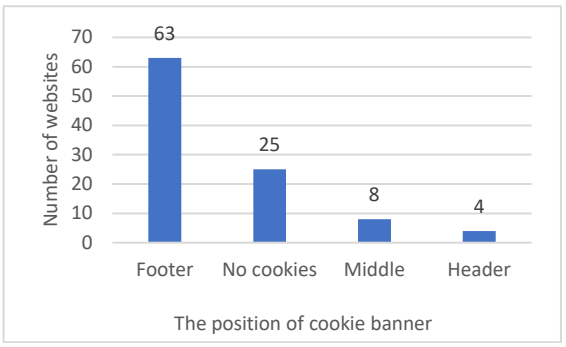


Figure 4. Frequency of cookie banner positions observed in Croatian public sector website (N=100)

4.3 Number of Options in Cookie Consent

To evaluate the extent of user autonomy in managing consent, the analysis examined the number of options available to users within cookie consent interfaces, on websites where cookies were present (N = 75). Fig. 5 shows that websites offering two options were the most prevalent (47%, 35 of them), these typically included options such as “Accept” and “Manage settings”, allowing limited user control. Furthermore, websites that provided only a single option were also represented (29%, 22 of them), most commonly through some variation of an "I agree", "Accept", or "Continue" button. Lastly, in a somewhat smaller proportion of cases (24%, 18 of them), three options were offered, typically including “Accept,” “Reject,” and “Customize” button, thereby allowing users greater control over their cookie preferences.

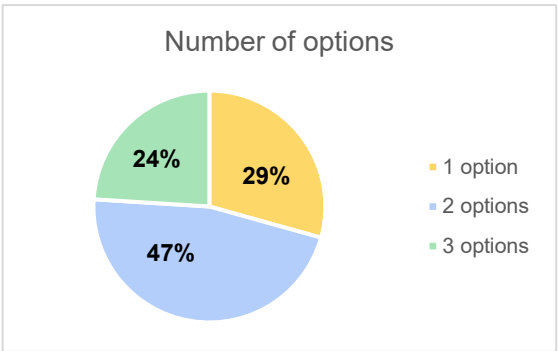


Figure 5. Number of options available to the users in cookie consent interfaces (N =75)

4.4 Number of Clicks to Reject Cookie

Finally, the number of clicks required to reject a cookie was analyzed. Fig. 6 shows that the majority of websites (74.67%, 56 websites) do not explicitly provide the option to reject cookies. Additionally, in 22.67% of cases (17 websites), users could refuse consent with a single click, while in 2.67% of cases (2 websites), rejecting cookies required two or more clicks.

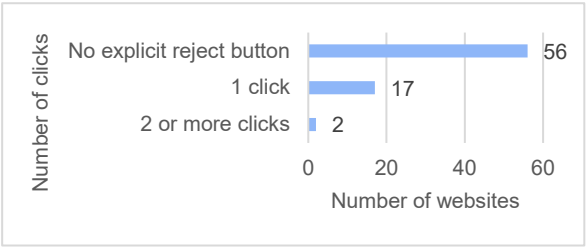


Figure 6. Number of user clicks required to reject cookie observed in cookie consent interfaces on Croatian public sector websites (N = 75)

4.5 Analysis of Dark Patterns Categories

As the categories of dark patterns are often intertwined and there is no clear distinction between them, this section details the criteria used to flag a website to a particular dark pattern. The description of each dark pattern was accompanied by a screen of the cookie consent to illustrate its representation in the listed dark pattern categories.

Nagging: The nagging category usually does not have a strict division; one most frequent forms are pop-ups that are hard to cancel. For example, before accepting cookie consent, a pop-up window appears asking whether the user wants to receive notifications about privacy policy updates, which the user must resolve in order to proceed. This was classified as a nagging pattern since the user's access to the page is interrupted by these options. Fig. 7 illustrates an example of a nagging pattern.

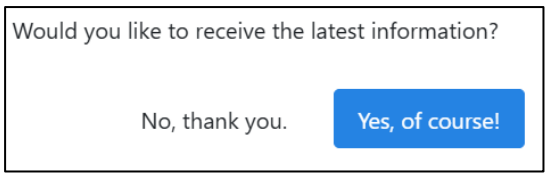


Figure 7. Example of a nagging pattern, a pop-up that interrupts the user's task

The second case, somewhat rarer, is the one in which the privacy settings are changed twice. First, cookie consent is accepted, and then immediately after that, another simpler form of cookie consent appears. This is classified as such a pattern because it burdens the user with an unnecessary number of interactions and information.

Obstruction: The obstruction category includes cases in which the process of accepting cookies must proceed in several steps in order to make it difficult to perform a certain action. In particular, this category included all cases of unequal path, for example, where in the initial view it was not possible to reject all cookies in one click, instead the user had to go to the privacy settings in order to change them. This was the main criterion for selecting cookie consent in this category, so it would not overlap with other dark patterns categories. Fig. 8 shows an example of an obstruction pattern where it is easier for the user to click on "I accept everything" button than go to the separate screen to change privacy settings.

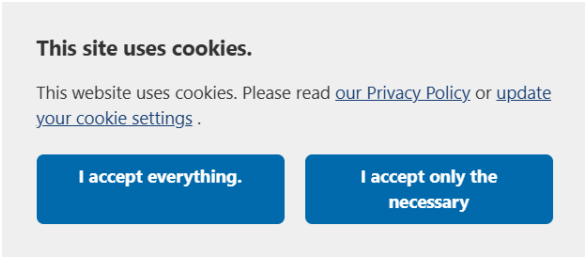


Figure 8. Example of obstruction pattern, only two options present

Interface interference: The interface interference was one of the main categories that included visual manipulations with the user interface that could affect the user's decision-making. Preselection is a common interface interference subcategory that is based on a default set or marked options that users usually accept in order to quickly perform an action. These subcategories included options in which checkboxes or other options were predefined, except for the strictly necessary cookies collection option. For example, the "I agree" checkbox can be previously selected for unnecessary cookies like marketing cookies, statistical cookies, and so on.

Fig. 9 shows an example of a preselection dark pattern where analytic cookies were selected by default.

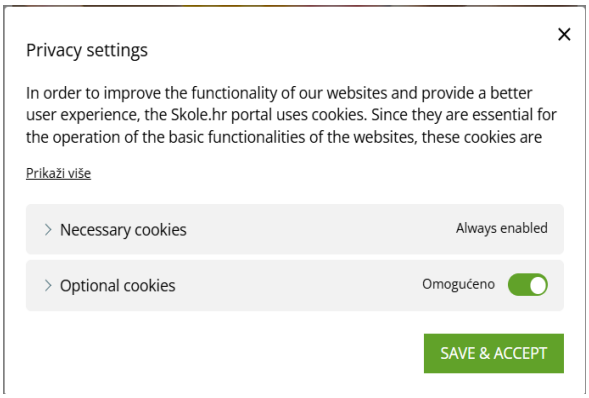


Figure 9. Example of preselection pattern, optional cookies marked by default

The second subcategory is aesthetic manipulation, which included options in which one button was visually emphasized relative to others, typically through the use of color. For example, when the accept button was prominently highlighted while the remaining options appeared grayed out. Furthermore, the false hierarchy subcategory comprised cases in which two buttons were highlighted, while the third option was weakly visible or expressed (e.g., the "Accept" and "Reject" buttons were highlighted, while the "Customize" option was weakly visible). The final subcategory, hidden information, encompassed options that were not easily visible, such as changing the privacy settings hidden within the text, displayed in a smaller font size, or displayed in the form of a link. Fig. 10 shows an example of interface interference subcategory named aesthetic manipulation by colour, where the "Allow" button is highlighted.

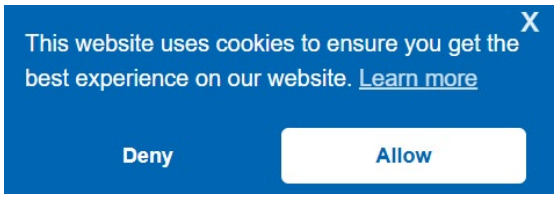


Figure 10. Example of aesthetic manipulation by colour

Fig. 11 shows a false hierarchy pattern where not all options are equally displayed (by size), and hidden information pattern in which the "Cookie settings" button is less noticeable.

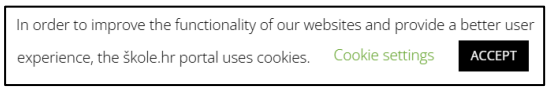


Figure 11. Example of false hierarchy and hidden information dark pattern

Forced action: The forced action category included cookie consent interfaces that compelled users to take action, or where users were forced to accept cookies due to a limited choice. A typical example of this option is where the user is only informed that cookies are being used, by accepting an option such as "I agree", "Continue", "Accept cookies", or "Close" button (Fig. 12).

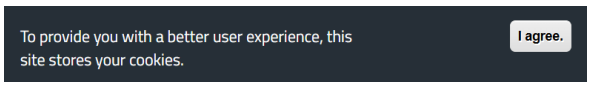


Figure 12. Example of forced action, only "I agree" button present

5 Discussion and Conclusion

In this research, a detailed analysis of 100 websites was carried out, with the aim of analyzing the presence of dark patterns in Croatian public sector websites. With regard to the first research question, the most common category of dark patterns observed was interface interference and obstruction patterns.

However, interface interference dark pattern was present due to its reliance on a strong psychological effect, and the large number of subtypes that fell into that category. Additionally, interface interference patterns were common in e-government websites, where dark patterns are often seen as highlighted buttons and preselected options (Alharbi et al., 2023). The European Commission (2022) study showed that preselection (as a subtype of interface interference) was the most popular dark pattern present in 41 out of 75 analyzed websites. However, the results present in this research show that the preselection pattern was not so common in public sector websites. This could be because personalization and data collection are less significant for public-sector websites than they are for commercial platforms such as shopping, news, or sports websites. For this reason, the preselection option was not predominantly used on public sector websites, as the goal was not to trick users into unknowingly accepting analytical or other optional cookies.

Regarding second research question, the number of options and clicks to refuse consent were examined separately. The most common cookie banner positions were located at the footer of the website. Previous research also shows that the footer and middle of a website are the most common places where cookies are placed (Mehrnezhad, 2020; Soe et al., 2020). Although it cannot be stated with certainty, the footer is generally a suitable place for cookie consent interfaces, as it often receives more user interaction (Utz et al., 2019).

Research conducted after GDPR was introduced revealed that certain cookie consent did not allow users to refuse consent with a single click (Hu and Sastry, 2019). As can be seen from the results of research conducted in this paper, this number has not improved significantly. This is concerning because the large number of websites makes it difficult for the user to find options for rejecting cookies. It is recommended that future regulations mandate an explicitly visible option for rejecting cookies.

The contribution of this work is evident in the dark patterns analysis on public sector websites that have not received sufficient attention in the literature. Thus, this addresses the gap in understanding what dark patterns are present on public sector websites and further areas for improvement. It has also been observed that a large number of sites do not have cookie consent, which is acceptable only in cases when they do not collect any data about users.

The limitations of this work refer to the fact that only a small sample of popular public sector websites was analyzed, to gain a more comprehensive understanding, the research should be repeated using a larger sample. In addition, the classification of dark patterns itself is rather vague, and there are no strictly defined rules about what belongs to which dark pattern category. For this reason, comparisons should be taken

with caution, because the very process of determining whether something belongs to the dark pattern category is quite subject to the researcher's assessment. Future research should focus on defining criteria, such as a checklist with detailed sub-criteria for categorizing different types of dark patterns. This would facilitate more objective assessments and allow for easier comparison of results across studies. Usable privacy research frequently includes experiments on a specific prototype or scenario to better understand the user's thinking and the impact of dark patterns (Fischer-Hübner and Karegar, 2024). Therefore, deeper research and analysis of the human aspect of privacy and security should be conducted. In addition to the future initiatives, apart from the dark pattern analysis itself, strengthening bright pattern practices should be considered to reduce this gap.

The specific implications of the research suggest that efforts should be made to minimize the appearance of interface interference and obstruction patterns, as they were observed in the majority of cases. Some of these practices include placing buttons in the same sizes and colours, as well as ensuring sufficient visibility of all options. Despite the psychological aspect of these patterns, mitigating their impact can be achieved by educating users and increasing their awareness of the good and bad cookie design practices.

Recommendations for future institutional practices can be proposed for each website category to promote ethical interface design. State-owned companies and public institutions should focus on mitigating interface interference patterns by avoiding deceptive visual manipulations and ensuring a neutral design that enables transparent user choices. Government bodies and administration websites should primarily address obstruction patterns by implementing a cookie consent interface design that minimizes the number of actions required for users to reject cookie consent. Since public sector websites serve a large audience, a general recommendation for all categories is to provide users with simple and clear options to withdraw consent. Implementing these practices is essential to strengthen the citizens' trust in public sector websites.

References

Alharbi, J. A., Albesher, A. S., & Wahsheh, H. A. (2023). An Empirical Analysis of E-Governments' Cookie Interfaces in 50 Countries. *Sustainability (Switzerland)*, *15*(2). doi:10.3390/su15021231

Berens, B. M., Bohlender, M., Dietmann, H., Krisam, C., Kulyk, O., & Volkamer, M. (2024). Cookie disclaimers: Dark patterns and lack of transparency. *Computers & Security*, *136*, 103507. doi:10.1016/j.cose.2023.103507

Bielova, N., Litvine, L., Nguyen, A., Chammat, M., Toubiana, V., & Hary, E. (2024). The effect of design patterns on (present and future) cookie consent decisions. In *33rd USENIX Security Symposium (USENIX Security 24)* (pp. 2813-2830).

Brignull, H., Leiser, M., Santos, C., & Doshi, K. (2023, April 25). *Deceptive patterns – user interfaces designed to trick you.* Retrieved December 19, 2024, from http://darkpatterns.org/

Distler, V., Fassl, M., Habib, H., Krombholz, K., Lenzini, G., Lallemand, C., Cranor, L. F., & Koenig, V. (2021). A Systematic Literature Review of Empirical Methods and Risk Representation in Usable Privacy and Security Research. In *ACM Transactions on Computer-Human Interaction* (Vol. 28, Issue 6). Association for Computing Machinery. doi:10.1145/3469845

European Commission: Directorate-General for Justice and Consumers, Lupiáñez-Villanueva, F., Boluda, A., Bogliacino, F., Liva, G., Lechardoy, L., & Rodríguez de las Heras Ballell, T. (2022). *Behavioural study on unfair commercial practices in the digital environment : dark patterns and manipulative personalisation : final report*, Publications Office of the European Union. doi: 10.2838/859030

Fischer-Hübner, S., & Karegar, F. (2024). Overview of Usable Privacy Research: Major Themes and Research Directions. *The Curious Case of Usable Privacy: Challenges, Solutions, and Prospects*, 43-102. doi:10.1007/978-3-031-54158-2_3

Gartner (2024). *Gartner 2024 Hype Cycle for Emerging Technologies Highlights Developer Productivity, Total Experience, AI and Security*. Retrieved December 19, 2024, from https://www.gartner.com/en/newsroom/press-releases/2024-08-21-gartner-2024-hype-cycle-for-emerging-technologies-highlights-developer-productivity-total-experience-ai-and-security

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (patterns) side of UX design. In *Conference on Human Factors in Computing Systems - Proceedings*, *2018-April*. doi: 10.1145/3173574.3174108

Habib, H., Li, M., Young, E., & Cranor, L. (2022, April 29). "Okay, whatever": An Evaluation of Cookie Consent Interfaces. In *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/3491102.3501985

Hu, X., & Sastry, N. (2019, June). Characterising third party cookie usage in the EU after GDPR. In *Proceedings of the 10th ACM Conference on Web Science* (pp. 137-141). doi: 10.1145/3292522.3326039

ICPEN (2024). *ICPEN Sweep finds majority of websites and mobile apps use dark patterns in the marketing of subscription services*. Retrieved December 19, 2024, from https://icpen.org/news/1360#_ftnref1

Jakobi, T., & von Grafenstein, M. (2023). What HCI Can Do for (Data Protection) Law—Beyond Design. In *Human Factors in Privacy Research,* pp. 115 - 136. doi:10.1007/978-3-031-28643-8_6

Kirkman, D., Vaniea, K., & Woods, D. W. (2023). DarkDialogs: Automated detection of 10 dark patterns on cookie dialogs. In *Proceedings - 8th IEEE European Symposium on Security and Privacy, Euro S and P 2023*, 847–867. doi: 10.1109/EuroSP57164.2023.00055

Kitkowska, A. (2023). The hows and whys of dark patterns: Categorizations and privacy. In *Human Factors in Privacy Research*, 173-198. doi: 10.1007/978-3-031-28643-8_9

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. In *Proceedings of the ACM on Human-Computer Interaction* (Vol. 3, Issue CSCW). Association for Computing Machinery. doi:10.1145/3359183

Mehrnezhad, M. (2020). A Cross-Platform Evaluation of Privacy Notices and Tracking Practices. In *Proceedings - 5th IEEE European Symposium on Security and Privacy Workshops, Euro S and PW 2020*, 97–106. doi: 10.1109/EuroSPW51379.2020.00023

Organisation for Economic Co-operation and Development (2022). Dark commercial patterns. *OECD Digital Economy Papers*. No. 336. OECD Publishing Paris. doi: 10.1787/44f5e846-en

Pochat, V. L., Van Goethem, T., Tajalizadehkhoob, S., Korczyński, M., & Joosen, W. (2018). Tranco: A research-oriented top sites ranking hardened against manipulation. In *Proceedings of the 26th Annual Network and Distributed System Security Symposium* (NDSS 2019). doi: 10.14722/ndss.2019.23386

Soe, T. H., Nordberg, O. E., Guribye, F., & Slavkovik, M. (2020, October). Circumvention by design-dark patterns in cookie consent for online news outlets. In *Proceedings of the 11th nordic conference on human-computer interaction: Shaping experiences, shaping society* (pp. 1-12). doi:10.1145/3419249.3420132

Sørensen, J., & Kosta, S. (2019, May). Before and after GDPR: The changes in third party presence at public and private european websites. In *The World Wide Web Conference* (pp. 1590-1600). doi: 10.1145/3308558.3313524

Utz, C., Degeling, M., Fahl, S., Schaub, F., & Holz, T. (2019, November). (Un) informed consent: Studying GDPR consent notices in the field. In *Proceedings of the 2019 ACM SIGSAC conference on computer and communications security* (pp. 973-990). doi:10.1145/3319535.3354212

A preliminary Study on the Use of Generative AI in Software Engineering Education

Pablo Hidalgo, Adrián Domínguez, Daniel Rodriguez
Department of Computer Science
University of Alcala
Alcalá de Henares, 28805, Madrid, Spain
{adrian.dominguez, daniel.rodriguezg}@uah.es
pablo.hidalgo@ufv.es

Marko Mijač
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
mmijac@foi.hr

***Abstract**. This study investigates the use of generative AI within a university-level software engineering (SE) course by analysing survey responses from second-year students. The survey explored multiple dimensions of students' interaction with AI while working on their course projects, including frequency of use, task types, prompting strategies, and perceived challenges. The results reveal widespread use of AI, especially for coding and debugging, though students also apply it across other phases of the software development lifecycle (SDLC). Despite this broad engagement, and even some use of advanced prompting, understanding of these techniques remains limited. All students reported verifying AI outputs, indicating low trust, which is further reinforced by the fact that many cited the inaccuracy of AI-generated results as their biggest challenge. Students also expressed a clear interest in improving their skills, particularly in prompt design. These findings underscore the need for structured support in AI literacy and prompting skills, as well as adapting course projects for an AI-enhanced learning context. This study provides a foundation for future research and instructional design in SE education.*

Keywords. Gen AI, software engineering, education

1 Introduction

Generative AI tools such as DeepSeek, Claude, ChatGPT, and GitHub Copilot are rapidly becoming part of the everyday toolkit for software developers, including students. In higher education, these tools offer opportunities to support coding, debugging, writing documentation, and other essential tasks in Software Engineering (SE) activities. The growing presence of these tools in educational contexts raises urgent questions about how students are engaging with them, what challenges students face, and what kinds of support students need use these tools effectively and responsibly.

As a result, the educational community has seen a surge of empirical studies investigating how generative AI is being used by students. This growing body of work is not only welcome but necessary. Given the diverse educational contexts, varying levels of AI literacy, and the rapid evolution of models, tools, and prompting strategies, no single study can offer generalizable conclusions. Continuous empirical research is needed to track how students adopt, adapt to, and are affected by these technologies, especially as institutional policies, pedagogical approaches, and the tools themselves evolve.

This paper contributes to the ongoing effort by examining how second-year undergraduate students used generative AI tools within a SE course. By analysing survey data on their usage patterns, prompting behaviours, challenges, and skill needs, we aim to provide timely insight into student practices and perceptions. These findings offer valuable implications for the design of instructional strategies, the development of AI literacy, and the adaptation of course projects to align with an AI-enhanced learning environment.

The rest of the paper is organized as follows: Section 2 provides background on generative AI in education and SE. Section 3 outlines the study's methodology, followed by survey results in Section 4. Section 5 presents a discussion of the findings, and Section 6 briefly covers the Threats to Validity. Finally, Section 7 concludes the paper with implications and directions for future research.

2 Background

In general, generative AI has transformed higher education by enabling personalized learning,

automating repetitive tasks, and fostering creative problem-solving. Generative AI tools support diverse applications, from writing assistance to code generation. However, their adoption raises concerns about academic integrity, over-reliance, and equitable access.

For instance, Ali et al. (2024) conducted a systematic review of AI applications in education, identifying key challenges such as tool reliability and ethical concerns, which align with the struggles reported by students in our study. Similarly, Fu and Weng (2024) emphasized the importance of framing responsible, human-centered AI practices, reinforcing our call for ethical integration in curricula. Lepp and Kaimre (2025) explored student perceptions of generative AI in programming education, finding that while students appreciated AI's assistance, its impact on actual learning outcomes was mixed—supporting our observation that students may rely on AI without fully understanding its mechanisms. Jin et al. (2025) provided a global overview of institutional AI adoption policies, highlighting the need for clear guidelines and structured support, which our findings also suggest. Yusuf et al. (2025) proposed a conceptual framework for pedagogical AI agents, noting their potential to enhance engagement and learning outcomes. This is particularly relevant given our students' limited awareness of AI agents and their capabilities. Finally, Terragni et al. (2025) discussed the future of AI-driven SE, emphasizing the growing role of autonomous agents and the importance of preparing students for this shift—an area where our respondents showed interest but lacked exposure.

In SE, AI tools provide now a huge support helping with code generation, debugging, and testing. There are reports that GitHub Copilot, for example, vastly reduces coding time and improves task completion rates[1]. On the other hand, there are also issues in the adoption of AI (Giannakos, 2024) as they can hinder students conceptual understanding by prioritizing task completion over a proper learning of the actual concepts. Furthermore, prompt engineering, i.e., the way precise inputs to optimize AI responses—is a critical skill for leveraging LLMs effectively, prompt patterns like few-shot prompting (providing examples to guide AI), chain-of-thought reasoning (breaking tasks into logical steps), and role-based prompting (assigning AI a specific role, e.g., "act as a senior developer") as essential for high-quality outputs. Not only that, but structured prompts can also improve code quality highlighting their importance in SE education.

Knot et al (2024) also states that students often rely on trial-and-error due to a lack of formal training, limiting their ability to tackle complex tasks like system design. Recent frameworks propose teaching prompt engineering as a core competency, similar to programming or debugging, see for example Lee and Palmer (2025). Hou et al (2024) review the use of Large Language Models in SE, highlighting their applications, optimization techniques, and challenges. They also include the number of studies using different prompt patterns.

AI agents, i.e., autonomous systems that perform multi-step tasks with minimal human intervention, are also transforming SE. Unlike LLMs, which respond to individual prompts, AI agents can autonomously write, test, and deploy code, integrating seamlessly with development environments. These agents support all SDLC phases by automating repetitive tasks. However, student awareness of AI agents is limited hindering students' ability to use agent in development projects.

Finally, there a need to for ethical integration of AI in education, including hallucination (AI generating incorrect outputs), bias in training data, and unequal access to tools. In SE, these are important, as biased AI outputs can introduce security vulnerabilities or perpetuate suboptimal design patterns. For example, an AI-generated sorting algorithm might prioritize speed over correctness if not properly verified, leading to errors in production systems. Nam and Bai (2024) emphasise the need for including ethics in the curricula to identify and mitigate AI biases, ensuring responsible use in technical disciplines.

3 Methodology

This study investigates the use of Generative AI in SE education through a survey-based approach combining quantitative and qualitative data collection. The research was guided by two questions:

- RQ1: How do students use Generative AI when performing SE tasks?
- RQ2: What challenges do students face when using Generative AI in their coursework?

To address these questions, we administered a structured survey to which 26 students enrolled in a second-year undergraduate SE course responded. The questions focused on students' use of Generative AI during their work on course-related software projects. While working on their projects, students were engaged in typical software development lifecycle (SDLC) activities, including requirements specification, software design, implementation (coding), and testing. Throughout the course, students were encouraged to explore and incorporate Generative AI tools in completing these activities and were provided with teaching materials that would help them do so. The questionnaire that was given to students comprised 10 questions, designed to capture both quantitative and qualitative data. Multiple-choice items were included to assess usage patterns and perceptions in a structured format, while open-ended

[1] https://github.blog/news-insights/research/research-quantifying-github-copilots-impact-in-the-enterprise-with-accenture/

questions provided opportunities for participants to elaborate on their experiences and perspectives in greater depth.

The survey was administered in person during a scheduled class session that was compulsive at the end of the term. Therefore, the survey was answered by almost all students taking the course, however, the participation was voluntary, and all responses were anonymous. In the future, we will explore the possibility of linking surveys to academic scores, for further statistical analysis. Quantitative responses were analysed using descriptive statistics (e.g., frequency distributions and means) to identify patterns in students' use of AI. Qualitative responses were analysed using open coding to extract themes and patterns related to student experiences and expectations.

4. Results

In this section, we provide the summarised results of the survey.

When asked how often they use AI tools for SE tasks, students responded as follows (see *Fig. 1*): "Always" – 3 students, "Frequently" – 13 students, "Sometimes" – 9 students, 'Rarely' – 1 student, and 'Never' – 0 students.

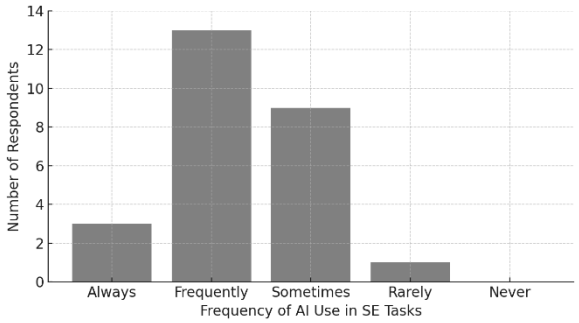


Figure 1. Frequency of AI use for SE tasks

Students were then asked to specify which SE tasks do they use AI for by choosing from multiple applicable options (see Fig. 2). The most frequently selected task was "Debugging," reported by 20 respondents. Both "Coding" and "Testing" were each selected by 16 respondents. "Requirements" was chosen by 13 respondents, while "Design" and "Documentation" were selected by 11 and 10 respondents, respectively.

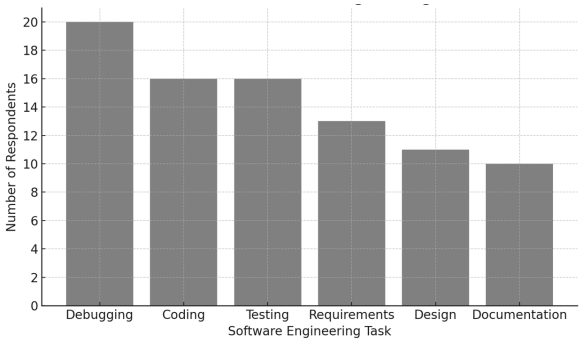


Figure 2. AI use across SE tasks

Regarding students' self-assessed competence with AI tools in the context of coding, a majority (19 respondents) considered themselves "Proficient," while 7 respondents rated their skills as "Basic."

In relation to debugging, students were asked how they typically ask AI for help, with multiple selections allowed (see Fig. 3). The most frequently reported approach was to "Paste errors," selected by 19 respondents. This was followed by "Ask for explanations" (15 respondents) and "Use step-by-step prompts" (13 respondents).

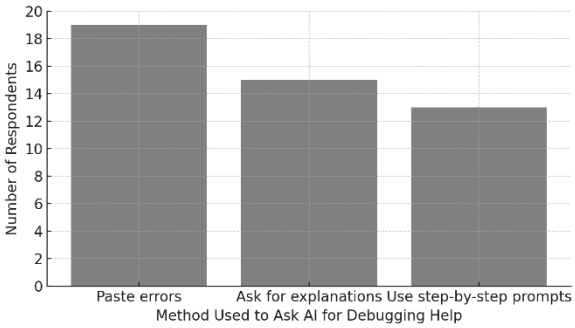


Figure 3. How students ask AI for debugging help

Students were asked whether they had experimented with advanced prompting techniques when interacting with AI tools. A total of 20 respondents indicated (see *Fig. 4*) that they had used at least one advanced strategy, while 6 stated that they relied solely on simple queries. Among those who had used advanced prompting, the most selected type was "Few-Shot" (13 respondents), followed by "Role-Based" (11), "Chain-of-Thought" (10), and "Constraints" (9).

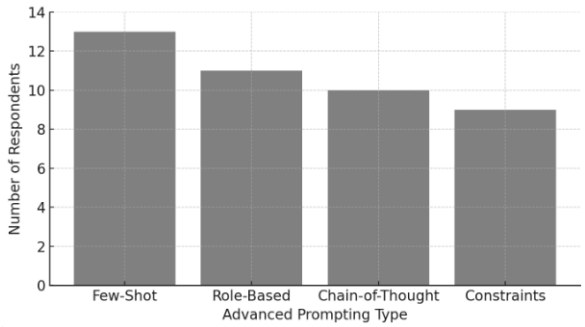


Figure 4. Use of advanced prompting techniques

In response to the question about verifying AI-generated outputs (see Fig. 5), 22 students indicated that they "always" check the results themselves. Four students reported that they do so "sometimes, for important things," while none selected the option "No, I trust it".

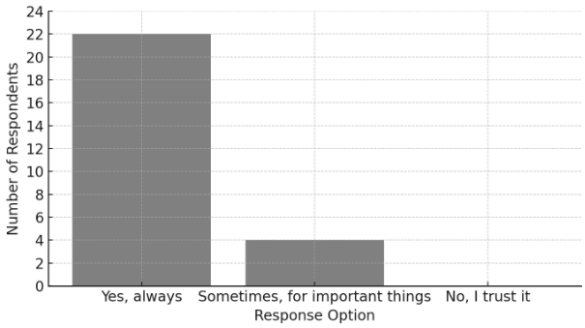


Figure 5. Verifying AI's results

When asked whether they understand what a prompt pattern is and how to use it, 14 students responded "Yes," while 10 indicated "No." It is worth noting that only 24 out of the 26 participants responded to this question.

The most commonly reported challenge students face when using AI is that it "makes mistakes," selected by 20 respondents. Other struggles were noted less frequently: 4 students pointed to ethical concerns, 3 admitted over-reliance on AI, and 2 cited other issues.

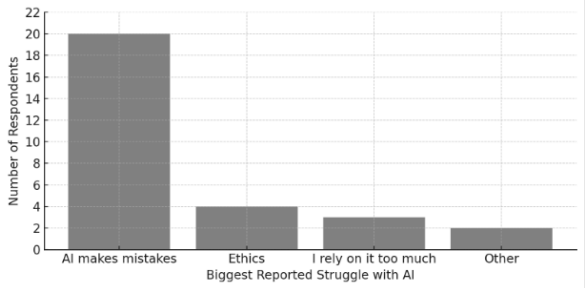


Figure 6. Students' struggles with AI

Out of 26 students, 22 provided a response to the open-ended question asking for an example where AI either helped or failed them. Eight students described positive experiences with AI assistance. Among these, 4 mentioned AI helped them with fixing errors, 2 with coding tasks, 1 with design and diagramming, and 1 with summarizing materials. In contrast, 14 students shared instances where AI failed to meet expectations. The most cited issue was incorrect code generation (7 respondents), followed by failures in error fixing (5 respondents) and diagram creation (2 respondents).

The survey results reveal several key outcomes regarding students' engagement with generative AI in SE tasks. First, AI use has become a regular part of students' workflow, with the majority reporting frequent or constant use which suggests a shift in how students approach software development, integrating AI tools as standard practice. Students used AI across a wide range of SE tasks, not just coding and debugging. However, the predominance of debugging and coding tasks also could reflect a tendency to rely on AI technical problem-solving rather than conceptual or design-oriented work.

The self-assessed competence levels show that most students feel confident in using AI for coding, which may reflect growing familiarity and confidence with these tools. Yet, it seems that fewer understand the underlying principles of prompting. This gap between practice and understanding highlights the need for formal training in prompt engineering for SE (this may not so necessary for general use, as LLMs are getting better at understanding user's use intend by the users). Finally, verification of AI outputs emerged as a challenge identified as the most common struggle. Open-ended responses showed that while some students benefited from AI assistance in coding and debugging, others encountered significant failures, especially in diagram creation. These mixed outcomes underscore the importance of teaching students how to critically evaluate and effectively prompt AI tools, in particular generating diagrams seems to be harder than generating code or test cases. Finally, students expressed a desire to improve their skills by applying AI to the full SE life cycle, indicating the necessity for educational support.

5 Discussion

In this section, we discuss our findings related to the two research questions stated previously in methodology section.

To begin with, we consider how students are currently using generative AI in the context of SE tasks (RQ1). The survey data indicate all students reported some level of AI use, with over 60% reporting that they use AI "frequently" or "always" for SE tasks. This suggests that AI has not only achieved a wide adoption among SE students but has become a regular part of their workflow rather than an occasional aid. Given this widespread adoption, SE educators may need to rethink the type, scope, and complexity of student

projects to ensure they remain appropriately challenging and educational in an AI-augmented learning environment. Regarding the use of AI in different SE tasks, while code-centric tasks dominate, students do not stop there, they are also aware of AI's potential in earlier and later phases of development process. This awareness is important as the SDLC is both increasingly and deeply infused with AI at every stage.

Survey results imply that all students verify AI-generated results, with 85% of students report doing that "always". Even if this does not reflect actual practice, it at least implies a shared awareness of the necessity of result verification. This cautious approach is especially relevant in SE tasks, where precision and correctness are critical and unchecked errors can lead to significant issues and maintenance problems. The fact that verification is universally reported also indicates that students may have moved beyond the initial naivety and inflated expectations that often accompanied the early hype around generative AI.

Students expressed strong confidence in their AI skills when coding, with over 70% identifying as proficient. They reported a variety of techniques for debugging with AI—ranging from simply pasting error messages into a chat window to more structured, step-by-step interactions. When shown prompt examples, more than 75% indicated that they had used some form of advanced prompting, with Few-shot prompting, Chain-of-thought and Role-based prompt types being the most recognized ones. However, only 58% reported knowing what prompt patterns are, highlighting a clear gap between practice and understanding—and underscoring the need for more formal, structured education in this area.

Turning to the second research question, we examine the challenges students face when using generative AI in their coursework (RQ2). A clear majority of students identified AI's tendency to produce incorrect or misleading outputs as their biggest struggle. This concern was reported far more frequently than other issues such as ethical dilemmas or over-reliance on AI. The emphasis on AI-generated mistakes aligns with previously mentioned students' cautious behaviour, such as consistently verifying outputs. While technical reliability clearly emerged as the central challenge—far outweighing other concerns—some students did report issues such as ethical dilemmas or overreliance on AI tools. Although only a small number explicitly mentioned becoming too dependent on AI, this may serve as an early warning sign, particularly given that the respondents are second-year students still developing foundational SE skills. It is also possible that the actual number is higher, as some students may not yet recognize their overreliance or may be hesitant to admit it. Similarly, the low frequency of ethical concerns in the responses may point not to their irrelevance, but rather to a lack of awareness—highlighting the need for more structured education around responsible AI use in academic and professional contexts.

Open-ended responses provided interesting insights into AI's performance and students' satisfaction with it. While some students shared positive examples of AI assistance, particularly in error correction, coding, and summarization tasks, nearly twice as many described instances where AI failed to deliver accurate or usable results. Common failure points included incorrect code, ineffective debugging assistance, and flawed diagram generation. The higher number of failure reports may suggest that such scenarios are more frequent, or at least more memorable. This raises important questions about whether these failures stem from limitations in the tools themselves or from students using them ineffectively—again pointing to the need for formal education on how to prompt and evaluate AI effectively. It may be also possible that some of these frustrations reflect unrealistic expectations or low tolerance for imperfection. As second-year students who began their academic journey in the age of generative AI, they may take the technology for granted in ways that more mature students or professionals—who worked without such tools—do not. This generational shift in expectations may shape how students perceive both the value and the limitations of AI.

When asked what AI-related skills they would like to develop, students overwhelmingly expressed a desire to improve their prompting abilities. This aligns with earlier findings indicating widespread use of prompting techniques, but only partial understanding of underlying patterns. Some students also noted a desire to apply AI more holistically across the SE process, rather than limiting its use to isolated tasks like debugging or code generation. Others expressed interest in learning how to design software with AI support or even build AI systems themselves. These responses suggest that students are not only aware of their current skill gaps, but are also willing to learn beyond surface-level usage. The strong emphasis on prompting shows the need for instructional support in teaching students how to engage with AI tools more deliberately, strategically and effectively.

To translate these findings into actionable teaching strategies, educators should consider integrating structured AI literacy modules into SE curricula. These modules should cover not only the technical use of generative AI tools, but also prompt engineering, agents, ethical considerations, and critical evaluation of AI outputs. For example, SE educators need to design assignments that require students to use different prompting techniques and reflect on the outcomes, thereby deepening their understanding of prompt patterns and their impact on AI outcomes. Moreover, educators need to adapt project-based learning assignments to include AI-supported workflows. This could involve tasks where students must use AI tools for requirements gathering, design,

coding, and testing, in CD/CI (Continuous Development/Continuous Development) followed by a critical review of the AI-generated output. Such activities would help students develop quality software with the help of AI across the whole SDLC.

6 Threats to Validity

This study presents several threats to the validity (TTV) that may affect the interpretation and generalizability of its findings. These are categorized below according to common validity frameworks in empirical SE research (Wohlin et al, 2024).

External Validity. The small sample size—26 students from a single second-year software engineering course—limits generalizability. Student behaviours may differ across institutions, academic levels, or course designs.

Internal Validity. Survey timing at the end of the course may have influenced responses due to recent experiences or exam stress. Uncontrolled factors like prior AI exposure or peer influence could also affect reported behaviours.

Construct Validity. The custom-designed, non-validated questionnaire may not reliably measure key constructs such as AI proficiency or ethical awareness. Additionally, self-reported data introduces bias, as students may misjudge their skills or respond in socially desirable ways.

Despite these TTV, this preliminary study offers some initial insights into student interactions with generative AI. These findings serve as a foundation for further research that incorporates validated instruments, broader survey samples (students and professionals), and mixed-research approaches to deepen our understanding of AI integration in SE engineering education.

7 Conclusions and future works

This preliminary study explored how second-year SE students use generative AI tools in their coursework and what challenges they face. The findings indicate that AI has already become a routine part of students' development practices, especially for tasks like debugging and coding. While students report confidence in their skills and actively experiment with advanced prompting strategies, gaps in understanding remain—a point also acknowledged by students themselves, who expressed a desire to learn more about prompting and AI use. Technical reliability remains the primary challenge to students, but ethical awareness, responsible use, and the risk of overreliance also appear to require further attention.

The results point to a clear need for structured AI literacy within SE education—especially in the areas of prompting, critical evaluation of outputs, ethical use, and awareness of potential overreliance. In parallel, student tasks and project designs may also need to evolve to remain challenging and meaningful in an environment where AI support is widespread.

As a preliminary study, this work lays the groundwork for more extensive and validated research involving larger and more diverse student cohorts and professionals. Future work will focus on designing and evaluating instructional modules that integrate generative AI into SE education. Additionally, future surveys could incorporate student performance on tasks involving AI tools as a proxy for understanding prompting patterns, thereby reducing reliance on self-reported data.

Acknowledgements

Daniel Rodriguez acknowledges "ACT-e: AI for enhanced classroom teaching" within the EUGLOH Consortium.

References

Ali, O., Murray, P.A., Momin, M., Dwivedi, Y.K. & Malik, T. (2024) The effects of artificial intelligence applications in educational settings: Challenges and strategies, *Technological Forecasting and Social Change*, 199, p. 123076. https://doi.org/10.1016/j.techfore.2023.123076

Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology, *Qualitative Research in Psychology*, 3(2), pp. 77-101.

Cigdem, S., Rumyana, N. & Giuseppe, D. (2024) Software engineering education in the era of conversational AI: current trends and future directions, *Frontiers in Artificial Intelligence*, 7. https://doi.org/10.3389/frai.2024.1436350

Fu, Y. & Weng, Z. (2024) Navigating the ethical terrain of AI in education: A systematic review on framing responsible human-centered AI practices, *Computers and Education: Artificial Intelligence*, 7, p. 100306. https://doi.org/10.1016/j.caeai.2024.100306

Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D. & Rienties, B. (2024) The promise and challenges of generative AI in education, *Behaviour & Information Technology*, pp. 1-27. https://doi.org/10.1080/0144929X.2024.2394886

Hou, X., Zhao, Y., Liu, Y., Yang, Z., Wang, K., Li, L., Luo, X., Lo, D., Grundy, J. & Wang, H. (2024) Large Language Models for Software Engineering: A Systematic Literature Review, *ACM Transactions on Software Engineering and*

Methodology, 33(8), Article 220, pp. 1-79. https://doi.org/10.1145/3695988

Jin, Y., Yan, L., Echeverria, V., Gašević, D. & Martinez-Maldonado, R. (2025) Generative AI in higher education: A global perspective of institutional adoption policies and guidelines, Computers and Education: Artificial Intelligence, 8, p. 100348. https://doi.org/10.1016/j.caeai.2024.100348

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence, 6*, 100225. https://doi.org/10.1016/j.caeai.2024.100225

Lee, D. & Palmer, E. (2025) Prompt engineering in higher education: a systematic review to help inform curricula, *International Journal of Educational Technology in Higher Education*, 22, p. 7. https://doi.org/10.1186/s41239-025-00503-7

Lepp, M. & Kaimre, J. (2025) Does generative AI help in learning programming: Students perceptions, reported use and relation to performance, Computers in Human Behavior Reports, 18, p. 100642. https://doi.org/10.1016/j.chbr.2025.100642

Nam, K., & Bai, Y. (2024). Unpacking the role of AI ethics online education for science and engineering students. *International Journal of STEM Education*, 11, Article 35. https://doi.org/10.1186/s40594-024-00493-4

Terragni, V., Vella, A., Roop, P. & Blincoe, K. (2025) The Future of AI-Driven Software Engineering, *ACM Transactions on Software Engineering and Methodology*, 34(5), Article 120, pp. 1-20. https://doi.org/10.1145/3715003

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T. & Du, Z. (2024) Artificial intelligence in education: A systematic literature review, *Expert Systems with Applications*, 252(Part A), p. 124167. https://doi.org/10.1016/j.eswa.2024.124167

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2024). Experimentation in Software Engineering (2nd ed.). Springer Berlin/Heidelberg, https://doi.org/10.1007/978-3-662-69306-3

Yusuf, H., Money, A. & Daylamani-Zad, D. (2025) Pedagogical AI conversational agents in higher education: a conceptual framework and survey of the state of the art, *Education Tech Research Dev*, 73, pp. 815-874. https://doi.org/10.1007/s11423-025-10447-4

Questionnaire

Section 1: How Do You Use AI?

1. How often do you use AI for coding/SE tasks? (Check one box)
    - Never (I don't use AI)
    - Rarely (only a few times)
    - Sometimes (for some tasks)
    - Frequently (almost every project)
    - Always (AI is part of my workflow)
2. Which SE tasks do you use AI for? (Check all that apply)
    - Requirements (e.g., "Generate user stories for a fitness app")
    - Design (e.g., "Create a UML diagram for a banking system")
    - Coding (e.g., "Write Python code for a REST API")
    - Debugging (e.g., "Explain why this Java error occurs")
    - Testing (e.g., "Generate unit tests for this function")
    - Documentation (e.g., "Summarize this code for a README file")
    - Other: _____

Section 2: How Skilled Are You with AI?

3. For CODING, which best describes you? (Check one)
    - Basic: I ask simple things (e.g., "How do I write a for-loop in Java?")
    - Proficient: I use smart prompts (e.g., "Optimize this SQL query for speed" or "Explain this algorithm, then rewrite it in C++.")
4. For DEBUGGING, how do you ask AI for help? (Check all that apply)
    - Paste errors (e.g., "Fix this error: NullPointerException")
    - Ask for explanations (e.g., "Why does this Python code give a TypeError?")
    - Use step-by-step prompts (e.g., "Analyze this stack trace and suggest fixes")
    - Other: _____
5. Have you tried ADVANCED PROMPTING? (Check all that apply)
    - Few-Shot (e.g., "Here's my code. Suggest improvements like these examples.")
    - Chain-of-Thought (e.g., "First explain the bug, then suggest fixes.")
    - Role-Based (e.g., "Act as a senior dev and review my code.")
    - Constraints (e.g., "Rewrite this function with O(1) space complexity.")
    - None (I only use simple questions)

Section 3: Challenges & Feedback

6. Do you CHECK AI's work? (Check one)
    - No, I trust it
    - Sometimes, for important things
    - Yes, always! (I test code/docs myself)

7. What's your BIGGEST struggle with AI? (Check one)
    - o AI makes mistakes
    - o I rely on it too much
    - o Ethics (e.g., plagiarism, cheating?)
    - o Other: _____
8. Give ONE example where AI helped (or failed) you:

(Example: "AI wrote a SQL query for me, but it had a syntax error.")
Your Answer:
_____

9. What AI skills do you WANT TO LEARN?

(Example: "Better debugging prompts" or "How to design with AI")
Your Answer:
_____

10. In general, do you think you know what a prompt pattern is and how to use them?
    - o Yes, I know them and use them
    - o No, I don't know about prompt patterns

A Comparative Study of Vibe Coding with ChatGPT and Gemini in Front-end Web Development

Marko Horvat, Barbara Kralj, Gordan Gledec
University of Zagreb, Faculty of Electrical Engineering and Computing
Department of Applied Computing
Unska 3, HR-10000 Zagreb, Croatia
{Marko.Horvat3, Barbara.Kralj, Gordan.Gledec}@fer.unizg.hr

Abstract. *The invention of Generative AI and Large Language Models has recently catalyzed "vibe coding" as a new paradigm of software development in which developers use natural language to state their intentions. However, there is currently a significant lack of empirical research comparing the fundamental behaviors of GenAI tools and their code quality. This paper presents such comparative study of GPT-4o and Gemini 2.5 Pro for front-end web development using everyday technologies HTML, CSS, and JavaScript. Using zero-shot and prompt-chaining strategies, we tasked the models to create three commonplace web applications of increasing complexity. The architecture and features of the generated code were evaluated using a mixed-method evaluation framework. The results show that GPT-4o and Gemini 2.5 Pro represent two different development paradigms; GPT-4o functions as a tool that generates minimal, concise code that follows user instructions, but in more complex tasks it sometimes generates errors and unwanted changes in the codebase. In contrast, Gemini 2.5 Pro operates as a proactive-enhancement agent that generates more complex, feature-rich code by anticipating the user's needs and adding advanced UI functionalities. Importantly, vibe coding is formally defined, explained, and compared to other AI-assisted programming approaches. The codebase created for this research is available at: https://github.com/mhorvat/vibecoding_frontend.*

Keywords. Software engineering, Automatic programming, Human-computer interaction, Generative AI, Large Language Models

1 Introduction

The field of software engineering is currently undergoing a massive paradigm shift, driven by the integration of Generative Artificial Intelligence (GenAI) and Large Language Models (LLMs) into the software development lifecycle (Coello, Alimam & Kouatly, 2024). These models are no longer confined to autocompletion but are emerging as proactive virtual partners in application design, code implementation, refactoring, and testing (Kotsiantis, Verykios & Tzagarakis, 2024). This evolution has very recently given rise to a term "**vibe coding**" coined by Andrej Karpathy, which describes a software development methodology centered on a developer's intuitive, natural language expression of intent to an LLM (Karpathy, 2025). This practice lowers the entry barrier for code developers, making possible for individuals with minimal formal programming knowledge to create software artifacts. As such, vibe coding signifies a fundamental transition from programming as a formal, syntactic, computer engineering task to a conversational, Human-Computer Interaction (HCI) challenge (Gunatilake et al., 2024).

Despite the widespread adoption of LLM-based coding assistants (Porter & Zingaro, 2024), because of the fast pace of development there is currently a significant lack of rigorous, comparative studies analyzing their performance and technical characteristics from a software engineering perspective (Shakya, Vadiee & Khalil, 2025) (Liang, Yang, & Myers, 2024). It remains unclear whether these tools are interchangeable or if they embody different underlying design philosophies that profoundly affect the software development process and the final product (Sergeyuk et al., 2025).

The presented research addresses this gap by evaluating the two most commonly used general-purpose LLM models, OpenAI's GPT-4o and Google's Gemini 2.5 Pro, on vibe coding of three typical front-end development tasks of increasing complexity. The trade-offs in vibe coding between the two models are analyzed and their implications discussed for novice developers, who are increasingly turning to these tools as their primary means of learning and building software, as well as senior developers who may turn to vibe coding for quick prototype building or assessment of new software technologies.

The remainder of the paper is organized as follows: the next section defines the vibe coding paradigm and explains its advantages and disadvantages in the software development process, the third section details the experimental methodology, while the fourth section

presents the experimental results. The fifth section discusses the broader implications of our findings. Finally, the last section concludes with directions for future research and recommendations for academia.

2 What is Vibe Coding?

Vibe coding is a novel software development paradigm (as of June 2025) in which a developer expresses their intentions in creating a software product using natural language in a GenAI LLM. Instead of writing precise, line-by-line code, the developer acts as a "high-level coordinator", guiding the AI agent through an iterative dialogic process of repeated code generating prompts and refinements. In vibe coding the key challenge is no longer mastering a programming language and writing software line-by-line but efficiently articulating the desired outcome, i.e., the "vibe", and critically evaluating the results produced by the AI.

It is very important to distinguish vibe coding from any AI-assisted programming because vibe coding does not imply "using AI tools to help write code" (Sapkota, Roumeliotis & Karkee, 2025). Specifically, the term vibe coding can be defined as "**generating code with AI without understanding the code that is produced**". As clearly stated by Andrej Karpathy: "vibe coding is a method of developing throwaway projects that is enjoyable, causing one to forget that the code exists" (Karpathy, 2025). This is not the same (in fact, it may be argued it is exactly the opposite) as incorporating LLM tools into a process for the documented and responsible development of production code. Table 1 lists and briefly describes the key features of vibe coding that differentiate it from AI-assisted coding and traditional programming paradigms.

Table 1. Key differences between vibe coding, AI-assisted coding, and traditional programming

	Vibe Coding	AI-assisted Coding	Traditional programming
Developer role	Coordinator guiding the AI with natural language	Coder using AI tools for assistance	Only author, writing all code manually
Developer core skills	Prompting, evaluation, and iterative refinement	Programming skills and AI tool proficiency	Mastery of syntax, algorithms, and architecture
Code understanding	Not required; the code is a "black box"	Required; the developer owns all code	Absolute; the developer is the author
Interaction method	Natural language dialogue	IDE with AI autocompletion and suggestions	Directly writing code in an IDE
Process type	Probabilistic and non-deterministic	Hybrid (deterministic + probabilistic)	Fully deterministic
Intended use	Rapid prototyping and non-production projects	Productivity boost for professional developers	All development, including production systems

Vibe coding offers an opportunity for most people to develop custom software, even though the majority does not know how or even will not learn to code in a particular programming language or software technology.

2.1 Who Benefits from Vibe Coding?

Given these special characteristics, it is important to define who the vibe coding is intended for. Although the paradigm seems *especially appealing to junior developers*, particularly those with minimal formal programming experience, it is *equally valuable for senior developers* who may lack the time or interest to thoroughly learn every emerging technology or framework. Vibe coding reduces the entry barrier by enabling users to express intent through natural language, rather than requiring them to master new syntax or APIs. This significantly accelerates development, reduces the cognitive load associated with traditional coding practices, and eliminates the "cold start" problem with personal productivity in new technologies by allowing developers to start programming immediately without having to consult manuals and technical documentation or acquiring sufficient experience with a new toolset.

2.2 Who Should Not Use Vibe Coding?

However, the simplicity that vibe coding offers can be misleading and potentially harmful particularly for inexperienced programmers. Junior developers often lack the theoretical foundations required to comprehend software architecture, detect hidden errors, and debug ineffectual components. In academic settings, schools and universities, particularly within computer engineering and computer science curricula, relying on vibe coding may cause students to produce suboptimal but sufficiently functional code without understanding the underlying logic, control flow, or data structures. Because vibe coding is inherently simple to use, students might disregard fundamental learning processes, acquiring the ability to prompt for solutions but lacking the underlying necessary analytical foundations that formal programming education provides.

Senior developers, on the other hand, who have already mastered these fundamental skills and learned programming in a traditional, structured way, are better equipped to use vibe coding productively. For them, it is a tool for rapid prototyping, exploring new technologies, and reducing boilerplate overhead. Nonetheless, as the codebase generated by LLMs grows, experienced developers may find it increasingly difficult to trace, understand, and debug such projects. The maintenance challenges of large AI-generated projects could become a significant overhead. The non-deterministic and verbose nature of LLM output may worsen the complexity, introducing inefficiencies and code maintainability issues in the long run.

Therefore, a clear recommendation emerges; *vibe coding should be used as a productivity tool for experienced developers for rapid prototyping or exploring new domains*. Vibe coding should not be used as a primary method for programming education, nor is it appropriate for developing large-scale or production-grade software systems.

2.3 Essential skills for a Vibe Coder

The essential skills for the new role of vibe coder or vibe programmer are no longer centered on the traditional computer engineering knowledge such as computer science theory, algorithms and methods, object-oriented programming, computer language syntax, programming patterns, or the actual programming language experience, but on the competence of the interaction with the LLM-based chatbot. Specifically, these new skills important for vibe coding include:

Prompt Engineering: The ability to efficiently create precise, specific, and context-rich prompts is paramount to guiding the LLM toward a desired outcome. The methodology employed in this study, which breaks down complex application development into a sequence of iterative prompts, aligns with established prompt engineering best practices like iterative refinement and task decomposition.

Critical Evaluation: The developer must possess the ability to critically assess the quality, correctness, and architectural soundness of the AI-generated code. LLMs frequently introduce subtle errors, security flaws, or inefficient patterns that may function correctly but are poorly designed or may not work at all although LLM is "confident" that the generated code is correct and bug-free.

Iterative Refinement: Crucially, success in vibe coding depends on engaging in a conversational loop of generating code, testing its output, and providing targeted feedback and corrective prompts to the AI to fix deviations and hopefully, progressively converge towards a satisfactory solution. However, this iterative and interactive dialogue does not necessarily always lead to the ultimate goal, as the LLM can sometimes get "stuck" at a certain point in the development process, showing no further progress despite repeated corrections. In such cases it is often advisable to restart the vibe coding process from the beginning rather than continue an unproductive conversational thread.

2.4 Vibe Coding as End-User Software Engineering in HCI research

The user of a vibe coding paradigm, i.e., **vibe coder**, can be defined as a *novice with minimum prior knowledge of the software development process*, or even, *person who is not a software developer*. This context positions the vibe coding in the domain of End-User Development (EUD) or End-User Software Engineering (EUSE) (Robinson et al., 2025) (Gunatilake et al., 2024), a HCI field where non-professional developers create, modify, or extend software artifacts. In this context, LLMs represent a powerful new enabling technology for EUSE, allowing users to specify complex needs in natural language rather than relying on constrained graphical interfaces or simplified scripting languages.

2.5 Vibe Coding is not completely novel: The "Zero-Code" Analogy

The ambition to give software developers with little to no formal training an ability to create functional applications is not a recent phenomenon born from the GenAI LLM revolution. The core promise of "vibe coding" - enabling creation through high-level intent rather than low-level syntax - actually follows a long history of "zero-code" and Rapid Application Development (RAD) tools that began in the late 1980s (Beynon-Davies et al., 1999).

Pioneering examples include Apple's HyperCard, a hypermedia system released in 1987 that allowed users to build interactive programs, presentations, and databases using a simple "stack of cards" metaphor. In the 1990s, the client-server era saw the rise of powerful RAD tools (Hirschberg, 1998). Sybase PowerBuilder, first released in 1991, became famous for its "DataWindow" technology, which enabled developers to create complex, data-driven forms, and reports with minimal coding. Similarly, Borland Delphi, launched in 1995, provided a visual design environment and a component-based architecture that dramatically accelerated the applications development. Other early tools already offered intuitive graphical interfaces for database creation and management, further lowering the barrier to entry for application development (Naz & Khan, 2015).

However, a crucial distinction separates these historical precursors from modern vibe coding: *the past tools were deterministic systems*. Their behavior was predictable; a given set of inputs and actions would always and reliably produce the same output. In contrast, LLM-based vibe coding is inherently probabilistic (Taulli, 2024). The output from an AI prompt is not guaranteed to be identical across repeated attempts, and the models themselves are often "black boxes". Furthermore, just as their predecessors often struggled when faced with requirements for complex or unusual functionality, LLMs can also fail to produce viable solutions for novel or highly specialized tasks (Jing et al., 2024).

Vibe coding, therefore, can be seen not as a complete breakthrough. It is only the latest incarnation in a multi-decade quest to make software development simpler and more accessible. Vibe coding is distinguished from AI-assisted programming tools and traditional programming paradigm by several distinct features, as listed in Table 1, but the most by its natural language conversational interface and non-deterministic and stochastic behavior.

3 Experimental Setup

The objective of the conducted experiment was to empirically compare the performance, generated software components, and interactive behavior of OpenAI's GPT-4o and Google's Gemini 2.5 Pro models as would be used by a novice developer for frontend web development tasks.

The experiment was intentionally constrained to foundational web technologies to isolate the LLMs' core generative capabilities without the potentially many confounding variables of newer, complex, or less popular frameworks which might have less represented in the LLMs' training set. The chosen stack included:

- HTML: For web document structure.
- CSS with Bootstrap: For frontend styling and responsive layout.
- Plain ("vanilla") JavaScript: For client-side dynamic logic, deliberately avoiding frameworks like React, Node.js or Vue.js to better assess the models' basic coding abilities.

We tasked the models to create three characteristic web frontends of increasing complexity: 1) simple paraphrasing and spellchecking tool, 2) moderately complex messenger client, and 3) complex e-commerce web shop. The first two applications required only basic HTML, CSS, and JavaScript without external graphical assets, and we provided the models with a predefined set of images and icons to be incorporated into the web shop interface. Additionally, a reference picture was provided to the models as a visual template to illustrate the intended appearance of the web shop frontend layout.

A zero-shot prompt chaining approach was used to instruct the models through the code development process. Specifically, for each application, a predefined set of sequential prompts defining web features and frontend layout was given to each model. After each prompt, the generated code was analyzed, and corrective follow-up prompts were issued as needed to fix errors or address misinterpretations. This process simulates a realistic, conversational development workflow in the vibe coding paradigm. All interactions were conducted in English, the *de facto* language for software development and the primary language for LLM training data. In the final step of the experiment, the generated code for the web shop was exchanged between the two models, with each model tasked with refactoring the other's output. A long and detailed zero-shot prompt was used for the final refactoring.

The sequence of prompts for each interface, the paraphrasing tool, the messenger client, and the web shop, is shown in Fig. 1. A custom metrics was defined containing a comprehensive set of various vibe coding features to evaluate each model's performance. The metrics is explained in Table 2.

Table 2. Model evaluation metrics used in the experiment.

Category	Feature	Description
Interaction effort	Number of instructional prompts	The number of initial prompts required to define a specific frontend task.
	Number of corrective prompts	The number of follow-up prompts needed to fix errors or deviations.
	Total number of prompts	The sum of all interactions with the model.
Code quality and size	Total Lines of Code (LOC)	The total volume of generated HTML, CSS, JS code (excluding comments).
	Number of unrequested features	Functionalities generated by the LLM that were not explicitly asked for in the prompts. Model proactivity.
Reliability	Number of errors	The number of functional or syntactic errors in the generated code.
	Number of misinterpreted prompts	Instances where the model significantly misunderstood a core requirement.
User perception	Subjective rating (1-5)	User rating of the final solution's quality, functionality, and adherence to the initial "vibe."

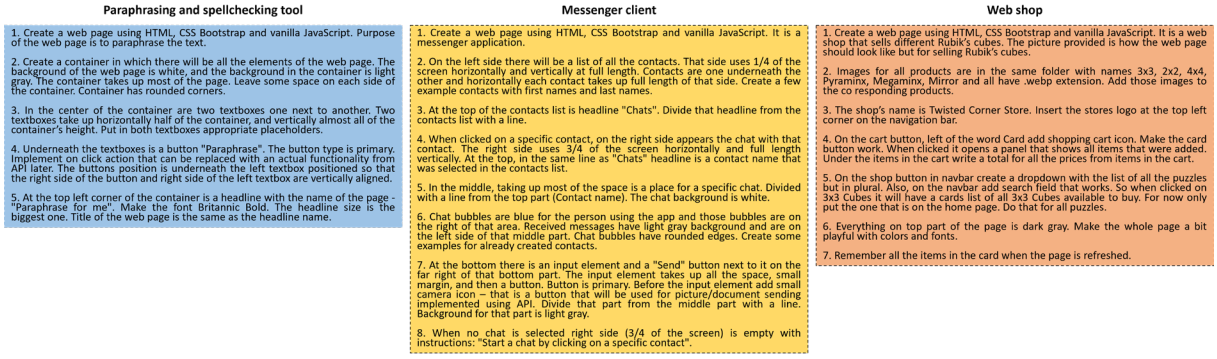


Figure 1. The sequence of prompts used to vibe code the three web frontends (paraphrasing tool, messenger client and web shop) of different complexity, the simplest on the left, and the most complex on the right.

4 Results and analysis

The final appearance of the paraphrasing frontends for both evaluated models is shown in Fig. 2.

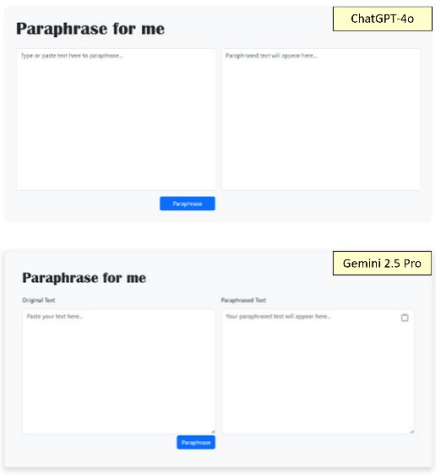


Figure 2. Comparison of paraphrasing tool frontends generated by GPT-4o (top) and Gemini 2.5 Pro (bottom).

As can be seen in Fig. 2, both versions share a similar minimalist design but there are notable differences in usability and visual clarity. Gemini's interface includes explicit field labels ("Original Text" and "Paraphrased Text"), which improve clarity and usability. Also, it incorporates a copy-to-clipboard button in the output field, providing a functional enhancement missing in the GPT-4o code. Overall, although similar Gemini's layout provides a more structured visual design with slightly better user experience (UX) and web interface quality.

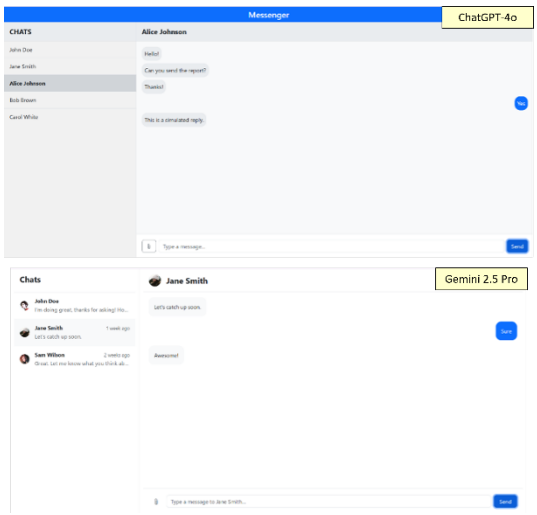


Figure 3. Comparison of messenger tool frontends generated by GPT-4o (top) and Gemini 2.5 Pro (bottom).

As with the first frontend task, in the second task Gemini 2.5 Pro's messenger frontend shows more UX maturity by including profile images, timestamped messages, and chat previews in the sidebar (Fig. 3). These features improve user orientation and make the interface feel more like a modern messaging platform. The GPT-4o version, on the other hand, offers a minimalist layout with functional but less engaging visual elements, without avatars, time context and preview content. Although both designs include quick reply buttons, and functional input areas, Gemini's user interface is richer, more intuitive to modern users.

In the last task, generating a web shop, the models were given a prompt and an image with a template for the layout of the user interface (see Fig. 4).

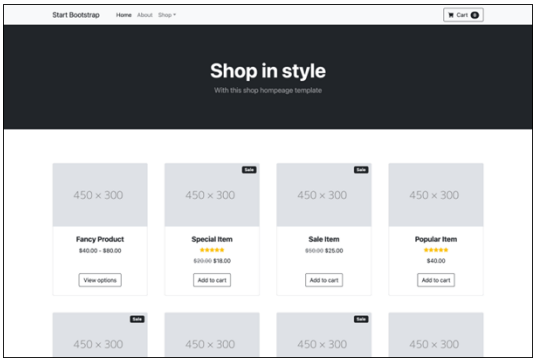


Figure 4. The webshop frontend layout template given to the models together with the prompts.

The output of the webshop vide coding task is presented in two parts. In addition to generating the homepage of the webshop homepage, both models also produced a shopping cart layout. The main layout of the front end is shown in Fig. 5, while the shopping cart is presented separately in Fig. 6.

The webshop frontend generated by Gemini appears more complete. It contains additional features such as product discount display, a sales label, a structured footer, and a more extensive product catalogue. The GPT-4o model provides a more colorful and well-rounded aesthetic with simpler product labelling and fewer UI enhancements. The Gemini vibe code for the shopping cart interface has a higher level of design maturity and functional completeness compared to GPT-4o. Gemini includes important usability improvements such as product thumbnail images, clearly defined quantity indicators, intuitive delete icons, and a prominently placed checkout button that allows for better e-commerce experience. The GPT-4o is functionally appropriate and sufficient, but lacks visual context and transactional cues, making it more suitable for early prototyping than for a user-ready deployment.

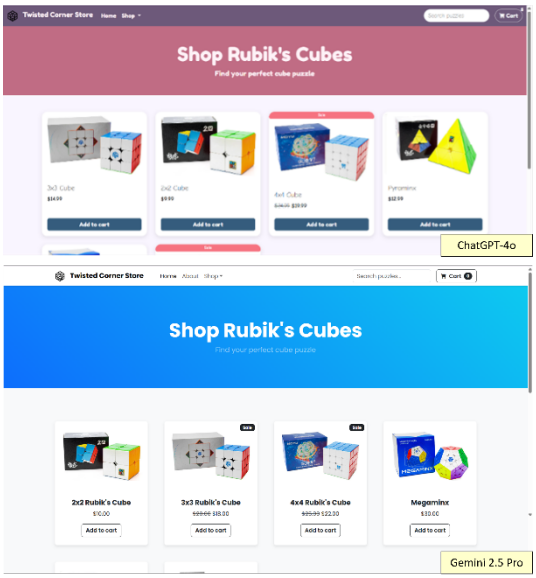


Figure 5. Comparison of webshop frontends created by GPT-4o (top) and Gemini 2.5 Pro (bottom).

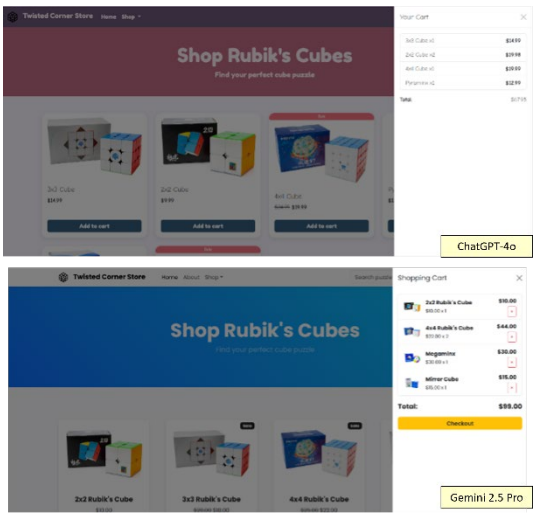


Figure 6. The shopping cart layout in webshop frontends from GPT-4o (top) and Gemini 2.5 Pro (bottom).

Regarding the webshop task, similarly to previous tasks, it can be concluded that while GPT-4o version is functional and visually appealing for a prototype, Gemini's design is more in line with modern e-commerce standards, and supports scalability, visual elements hierarchy, and distinctive marketing elements.

The analysis of the generated code revealed that the aggregated quantitative results of all three applications exhibited consistent variations in the behavior of the employed models. First, Gemini 2.5 Pro consistently produces significantly more code than GPT-4o. This gap becomes larger as task complexity increases, from a 61% larger codebase for the simple paraphrasing tool to an 89% larger one for the most complex webshop. Second, Gemini consistently generates unrequested but useful UI features (e.g., loading spinners, checkout buttons), whereas GPT-4o does not generate extra features and strictly follows the user's prompts. Third, GPT-4o makes errors that require re-entering prompts that have already been entered in the prompt sequence, sometimes several steps before. Fourth, Gemini consistently received higher subjective satisfaction ratings from the standpoint of a novice user. Finally, Gemini's proactive nature sometimes allows it to anticipate future steps, reducing the number of instructional prompts needed, as seen in the Messenger frontend task. The aggregated quantitative results are shown in Table 3.

Table 3. Quantitative performance metrics results.

Metric	Task	ChatGPT	Gemini
Number of instructional prompts	Paraphrase	5	5
	Messenger	7	4
	Web Shop	7	7
Total number of prompts	Paraphrase	14	10
	Messenger	21	13
	Web Shop	14	17
Number of corrective prompts	Paraphrase	9	5
	Messenger	14	9
	Web Shop	7	10
Total Lines of Code (LoC)	Paraphrase	84	135 (+61%)
	Messenger	120	147 (+23%)
	Web Shop	193	365 (+89%)
Unrequested Features	Paraphrase	0	2
	Messenger	0	2
	Web Shop	0	4
Avg. Subjective Rating (1-5)	Paraphrase	3.6	3.8
	Messenger	4.0	4.5
	Web Shop	3.7	4.1

Interestingly, there were no syntax errors in code from either model, in either language – HTML, CSS or JavaScript. The total number of misinterpreted prompts was 1 for GPT-4o, and 0 for Gemini 2.5 Pro. The subjective rating of the quality of the produced final output is given in Table 1.

A qualitative analysis of the code confirms already noticed differences in the architectural and stylistic decisions of each model. While Gemini's proactive approach resulted in a better and more dynamic architecture for the medium-complexity Messenger task, the same tendency proved negative in the high-complexity case webshop. Gemini 2.5 Pro hardcoded a large number of product entries directly into HTML, creating a rigid and unscalable codebase. In contrast, GPT-4o model used a more robust and traditional software engineering pattern: separating data (a JavaScript array of products) from presentation (dynamically rendered HTML). This approach is by design more manageable and scalable.

Because of the space constraints it is not possible to present the entire generated code in full detail; however, several representative elements can be highlighted.

For example, in the first scenario for the paraphrasing tool, the overall code quality produced by both agents is high. The generated code includes all requested components, is well-structured, and uses intuitive variable naming. GPT-4o implementation relies on a single main container (`paraphrase-container`) (Fig. 7) for element layout, whereas Gemini 2.5 Pro introduces additional components such as card and box-shadow, giving the application a more professional appearance (Fig. 7).

GPT-4o

```
<div class="paraphrase-container">
    <h1>Paraphrase for me</h1>
    <div class="content-row">
        ...
    </div>
</div>
```

Gemini Pro 2.5

```
<style>
.card {
    width: 100%;
    height: 100%;
    background-color: #f8f9fa; /* Light gray background */
    border: 1px solid #dee2e6;
    box-shadow: 0 0.5rem 1rem rgba(0, 0, 0, 0.15) ;
    display: flex; /* Use flexbox for the card itself */
    flex-direction: column; /* Stack card content vertically */
}
</style>
...
<div class="content-wrapper">
    <div class="card">
        <div class="card-body p-4 p-md-5">
            <h1 class="card-title mb-4">Paraphrase for me</h1>
            <div class="grid-container">
                ...
            </div>
        </div>
    </div>
</div>
```

Figure 7. The main HTML container for element layout in paraphrasing tool generated by GPT-4o (above) and Gemini 2.5 Pro (below).

In JavaScript, GPT-4o provides a minimal implementation with only the basic `onClick` action (Fig. 8), while Gemini delivers a more extensive and detailed implementation (Fig. 8). Notably, Gemini disables the "Paraphrase" button during text processing and adds a loading spinner, further improving usability. Finally, the layout in the ChatGPT-generated interface is static, whereas Gemini's interface exhibits greater flexibility and adaptability. A similar pattern emerges in the Messenger client task and the web shop task: GPT-4o versions are minimal, displaying only the necessary, while Gemini enriched the implementation with additional usability features improving both aesthetics and functionality.

The final step of the experiment was a refactoring scenario with the webshop code previously generated by the two agents. Specifically, the ChatGPT agent was assigned the code originally generated by the Gemini agent and, conversely, the Gemini agent received the code generated by ChatGPT. Along with the code, both agents were provided with a clearly defined prompt outlining how to approach the refactoring task. Upon receiving the instructions, both agents identified issues in the code and presented concise summaries of the problems, followed by detailed implementation steps for the refactoring process and best practice recommendations for future development.

GPT-4o

```
document.getElementById('paraphraseBtn').addEventListener('click', () => {
    const inputText = document.getElementById('inputText').value.trim();
    const outputTextarea = document.getElementById('outputText');
    outputTextarea.value = '';
    if (!inputText) {
        outputTextarea.value = 'Please enter some text to paraphrase.';
        return;
    }
    // Placeholder action : replace with actual API call
    console.log('Paraphrase button clicked. Input:', inputText);
    outputTextarea.value = '...paraphrased text will appear here...';
});
```

Gemini Pro 2.5

```
const paraphraseBtn = document.getElementById('paraphrase-btn');
const inputText = document.getElementById('inputText');
const outputText = document.getElementById('outputText');
const spinner = paraphraseBtn.querySelector('.spinner-border');
const copyBtn = document.getElementById('copy-btn');
paraphraseBtn.addEventListener('click', () => {
    ...
    spinner.classList.remove('d-none');
    ...
});
copyBtn.addEventListener ('click', () => {
    ...
});
...
```

Figure 8. JavaScript implementation of the button click handler in GPT-4o (above) and Gemini 2.5 Pro (below).

In the refactoring both agents reorganized the original code base into multiple files and directories to improve readability and make it easier to find components, as shown in Fig. 9. All business logic was moved to JavaScript files: 5 different files in GPT-4o refactoring, and only one JavaScript file (`main.js`) with Gemini 2.5 Pro. In addition, both agents introduced standardized conventions for naming components, adhering to the BEM (Block Element Modifier) method.

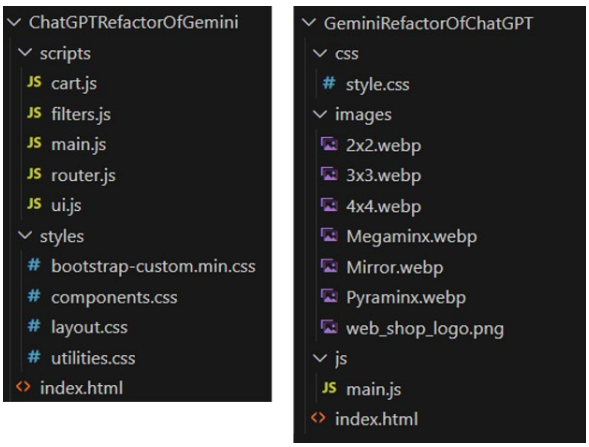


Figure 9. The file structure of the refactored code for the web shop task from GPT-4o (left) and Gemini 2.5 Pro (right).

However, despite providing informative comments and useful suggestions, the code refactored by GPT-4o performed worse than the original implementations. Moreover, the webshop refactored by GPT-4o was not functional because important elements such as the product display and the top navigation bar were

missing. The GPT-4o model assumed that the products would be located in a data directory within a `products.json` file, but failed to specify this requirement or provide an example of the expected JSON structure.

On the other hand, Gemini 2.5 Pro successfully refactored the code from GPT-4o and created an almost identical version of the original page. Despite minor shortcomings, such as the placement of the "Sale" button in the top right corner and oversized product images, the Gemini agent performed a more effective refactoring than GPT-4o. It preserved all the original functionalities and introduced subtle improvements.

The complete codebase developed for this research, consisting of HTML, JavaScript, and CSS files for all four tasks or scenarios (paraphrasing, messenger, webshop and refactoring), is freely accessible at: *https://github.com/mhorvat/vibecoding_frontend*.

5 Discussion

The key finding of the conducted experiment in front-end web development is that GPT-4o and Gemini 2.5 Pro are not simply interchangeable tools for vibe coding. They embody two different philosophies of AI-powered support that represent different levels of support for a developer.

OpenAI's ChatGPT with GPT-4o model acts as a direct and literal tool. Its behavior is predictable and controllable. It produces lean code that strictly adheres to the user's explicit instructions. This paradigm places the entire cognitive burden of design, functional specification, and architectural planning on the user.

Google Gemini 2.5 Pro works as a proactive and collaborative agent. It attempts to derive higher-level goals, anticipate the user's needs, and enrich the output with features and subtleties that go beyond specifications stated in the prompt. It shares the burden of design and makes decisions to deliver a more complete product.

For a junior developer, the choice between these two paradigms is crucial, and our results show that Gemini's approach is generally superior. The data show that Gemini consistently achieves higher subjective satisfaction ratings across all three front-end scenarios. This is because its proactive enhancements provide a more complete and professional looking application that is ready to use immediately. For a beginner, it is very motivating to see how quickly an application with many functions can be created.

The literalness of GPT-4o leads to simpler code, but it is also a significant disadvantage for junior users. Beginner developers must behave like experts, specify every design detail and anticipate web front-end features they may not even know are possible. This may be difficult for junior users and can lead to a frustrating development experience. Gemini's ability to "fill in the gaps" in the prompt, i.e., the software specification, makes it a far more effective assistant, guiding the novice to a better result.

However, Gemini's proactive nature, while beneficial, is not without risks. The "complexity-architecture inversion" observed in the webshop task, where Gemini's hard-coded HTML was architecturally inferior to dynamic data array generated by GPT-4o is a clear example of this feature. But for a developer who is more interested in rapid prototyping rather than building a large-scale enterprise-level application this is a worthwhile trade-off.

The primary goal of any developer using vibe coding should be to quickly turn an idea into a working software prototype. In this context, Gemini's delivery of a feature-rich, visually polished application, and more rapid development and validation is more valuable than the architectural simplicity of GPT-4o code.

Finally, the presented study has several limitations that must be acknowledged. First, the experiment did not include many other popular coding AI tools. An expanded experiment including even more LLMs, and frontend types is planned for future research and a more comprehensive publication. Second, the experiment was conducted with ratings from a single user which limits the generalizability of the obtained results. A formal verification should be undertaken in the future to assess the quality of the generated code more clearly. Third, the study was limited to vanilla HTML, CSS, and JavaScript. The observed paradigms may differ significantly for more complex frameworks such as Node.js, Vue.js or React. Finally, the experiment was focused solely on front-end UI development. The reported results may not apply to other application architecture components such as backend logic, database management or middleware, and should be investigated separately in the future.

6 Conclusion

The presented research provides a direct empirical comparison of OpenAI GPT-4o and Google Gemini 2.5 Pro LLMs in a vibe coding context of web frontend, showing that for junior and senior developers alike, the models are not only different, but one is demonstrably superior. Google Gemini acts as a proactive and augmenting agent, proving to be the more effective tool for both user groups. Its ability to anticipate user requirements, incorporate beneficial unsolicited features, and create more refined and comprehensive web frontend components is perfectly aligned with the goals of an inexperienced developer looking to convert a high-level "vibe" quickly and efficiently into a functional product. Also, for senior developers a proactive model such as Gemini 2.5 Pro will be more beneficial to develop "quick and dirty" code, but with a caveat of a larger codebase than with using GPT-4o. Such larger code will be more difficult to trace and

debug. However, this is a trade-off that a senior developer must contend with.

In the academic setting, vibe coding certainly has the potential to transform the educational process. Its integration into university-level programming instruction has an opportunity to increase engagement, accelerate learning, and improve overall teaching quality (Šarčević et al., 2024), especially when integrated into one of the many digital game-based learning models that support formal student assessment (Horvat et al., 2022). However, the increasing ease of automatic code generation also introduces new pedagogical challenges, particularly in academic integrity, plagiarism detection, and assessment design (Hartley, Hayak & Ko, 2024) (Mekterović, Brkić & Horvat, 2023). GenAI tools can help students with coding assignments by suggesting, hinting, and generating code snippets, encouraging better coding practices. However, using GenAI in programming education can lead to students becoming too dependent on AI-generated code and failing to understand fundamental concepts, which may negatively impact the long-term sustainability of computer engineering in higher education (Silva, 2024).

Developing robust solutions for detecting vibe-coded laboratory assignments, seminar papers, and student projects is an urgent issue that the academic community must address (Adnin et al., 2025). From an application design standpoint, vibe coding encourages the creation of responsive, mobile-friendly, and platform-consistent interfaces, allowing students and professionals to create modern frontends with minimum technical overhead (Smolić et al., 2024).

In addition to limitations of the experiment, as outlined in the previous section, which should be corrected with future investigations, this research opens several interesting avenues for further research. First, including other models such as Anthropic Claude[1], GitHub Copilot[2] or Cursor AI[3]. Second, by expanding the scope by replicating the study with more complex JavaScript frameworks and software technologies to determine if the observed paradigms remain consistent across the technology stack. Third, by using professional tools and formal code analysis techniques to determine the quality of the generated code. Fourth, the robustness of the vibe coding paradigm should be tested on different programming assignments, such as backend, databases, microservices, etc. Finally, in the future it would be interesting to investigate the impact of different prompting strategies and measure the impact of prompt structure on code quality, architectural choices, and model behavior more systematically.

Universities and curriculum designers must formulate explicit standards for the ethical and successful integration of vibe coding into coursework, ensuring that AI-assisted coding and vibe coding enhance rather than undermine the development of fundamental programming skills.

We hope our findings will have significant positive implications for tool selection and software developer education in vibe coding.

References

Adnin, R., Pandkar, A., Yao, B., Wang, D., & Das, M. (2025, April). Examining Student and Teacher Perspectives on Undisclosed Use of Generative AI in Academic Work. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).

Beynon-Davies, P., Carne, C., Mackay, H., & Tudhope, D. (1999). Rapid application development (RAD): an empirical review. *European Journal of Information Systems*, *8*(3), 211-223.

Coello, C. E. A., Alimam, M. N., & Kouatly, R. (2024). Effectiveness of chatgpt in coding: A comparative analysis of popular large language models. *Digital*, *4*(1), 114-125.

Gunatilake, H., Grundy, J., Hoda, R., & Mueller, I. (2024). The impact of human aspects on the interactions between software developers and end-users in software engineering: A systematic literature review. *Information and Software Technology*, 107489.

Hartley, K., Hayak, M., & Ko, U. H. (2024). Artificial intelligence supporting independent student learning: An evaluative case study of ChatGPT and learning to code. *Education Sciences*, *14*(2), 120.

Hirschberg, M. A. (1998). Rapid application development (rad): a brief overview. *Software Tech News*, *2*(1), 1-7.

Horvat, M., Jagušt, T., Veseli, Z. P., Malnar, K., & Čižmar, Ž. (2022, May). An overview of digital game-based learning development and evaluation models. In *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)* (pp. 717-722). IEEE.

Jing, Y., Wang, H., Chen, X., & Wang, C. (2024). What factors will affect the effectiveness of using ChatGPT to solve programming problems? A quasi-experimental study. *Humanities and Social Sciences Communications*, *11*(1), 1-12.

Karpathy, A. (2025, Feb 3). Vibe coding. Retrieved from

[1] https://www.anthropic.com/claude.
[2] https://github.com/copilot
[3] https://cursor.com/

https://x.com/karpathy/status/1886192184808149383

Kotsiantis, S., Verykios, V., & Tzagarakis, M. (2024). AI-assisted programming tasks using code embeddings and transformers. *Electronics*, *13*(4), 767.

Liang, J. T., Yang, C., & Myers, B. A. (2024, February). A large-scale survey on the usability of ai programming assistants: Successes and challenges. In *Proceedings of the 46th IEEE/ACM international conference on software engineering* (pp. 1-13).

Mekterović, I., Brkić, L., & Horvat, M. (2023). Scaling automated programming assessment systems. *Electronics*, *12*(4), 942.

Naz, R., & Khan, M. N. A. (2015). Rapid applications development techniques: A critical review. *International Journal of Software Engineering and Its Applications*, *9*(11), 163-176.

Porter, L., & Zingaro, D. (2024). *Learn AI-Assisted Python Programming: With Github Copilot and ChatGPT*. Simon and Schuster.

Robinson, D., Cabrera, C., Gordon, A. D., Lawrence, N. D., & Mennen, L. (2025). Requirements are all you need: The final frontier for end-user software engineering. *ACM Transactions on Software Engineering and Methodology*, *34*(5), 1-22.

Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2025). Vibe coding vs. agentic coding: Fundamentals and practical implications of agentic ai. *arXiv preprint arXiv:2505.19443*.

Sergeyuk, A., Golubev, Y., Bryksin, T., & Ahmed, I. (2025). Using AI-based coding assistants in practice: State of affairs, perceptions, and ways forward. *Information and Software Technology*, *178*, 107610.

Shakya, R., Vadiee, F., & Khalil, M. (2025, April). A Showdown of ChatGPT vs DeepSeek in Solving Programming Tasks. In *2025 International Conference on New Trends in Computing Sciences (ICTCS)* (pp. 413-418). IEEE.

Silva, C. A. G. D., Ramos, F. N., De Moraes, R. V., & Santos, E. L. D. (2024). ChatGPT: Challenges and benefits in software programming for higher education. *Sustainability*, *16*(3), 1245.

Smolić, E., Boras, B., Horvat, M., & Jagušt, T. (2024). Smartphone-Enabled Interaction on Large Displays—A Web-Technology-Based Approach. *Electronics*, *13*(5), 929.

Šarčević, A., Tomičić, I., Merlin, A., & Horvat, M. (2024, May). Enhancing Programming Education with Open-Source Generative AI Chatbots. In *2024 47th MIPRO ICT and Electronics Convention (MIPRO)* (pp. 2051-2056). IEEE.

Taulli, T. (2024). *AI-Assisted Programming: Better Planning, Coding, Testing, and Deployment*. " O'Reilly Media, Inc.".

Implementing Education 4.0 in Software Engineering Education: the JCC experience

Roberta Capuano, Henry Muccini
University of L'Aquila, Italy
FrAmeLab, DISIM Department
Via Vetoio, 67100 L'Aquila, Italy
{roberta.capuano, henry.muccini}@univaq.it

Zlatko Stapić
University of Zagreb Faculty of Organization and Informatics
Pavlinska 2, 42000 Varazdin, Croatia
zlatko.stapic@foi.unizg.hr

Abstract. *Education 4.0 emphasizes fostering innovation over merely imparting knowledge by seamlessly integrating advanced technologies into teaching and learning environments. This paradigm is particularly well-suited for software engineering education, aligning modern educational approaches with the evolving industry needs.*
This paper presents the Joint Creative Classroom methodology and reports the experience of the application of the Education 4.0 principle for the design, implementation and assessment of the Software Architecture Analysis and Design JCC course. This initiative aimed at fostering cross-university collaboration in higher education, involves two distinct universities, each contributing a unique course to create an innovative, interdisciplinary learning experience for students. Thus, this paper aims to provide insights into the challenges and successes of such collaborative endeavours.
The results indicate that over 90% of the students acknowledged the joint course's positive impact on enhancing their soft skills. Furthermore, the framework facilitated exposure to distributed software engineering practices.

Keywords. Education 4.0, interdisciplinary learning, industry collaboration, cross-university collaboration, global software engineering, learning design, JCC

1 Introduction

In the evolving landscape of *Global Software Engineering* (GSE), the imperative for flexible, cross-border, and *innovative teaching methods* has become increasingly evident (Iniesto et.al., 2021; Herbsleb and Moitra 2022; Beier et al., 2012). Teaching software engineering requires creating authentic contexts, fostering teamwork, handling projects with sufficient complexity, and providing ongoing student support (Angelov and de Beer, 2017). Moreover, as the *industry demands* continue to grow in complexity, educational institutions must equip their students with the skills and insights required to excel in this dynamic field (Moraes et al., 2023).

The *Education 4.0* framework is advocated as a solution to these emerging trends (Hussin, 2018; Bonfield et al., 2020; Salmon, 2019). Proposed by Harkins (2008), Education 4.0 describes innovation-producing education as opposed to knowledge-producing education. Definitions vary but usually focus on innovation, novelty, use of technology, and connections with employment and industry (Hussin, 2018; Salmon, 2019; Fisk, 2017; Mukul, 2023). In the Erasmus+ project *Accelerating the transition towards Edu 4.0 in HEIs – Teach4Edu4*, we proposed a definition of Education 4.0 that draws on ideas and descriptions in a range of past literature (Hussin, 2018; Salmon, 2019; Fisk, 2017; Jisc, 2019): Education 4.0 employs an approach to learning and teaching that emphasises the development of skills and competences necessary in a modern workplace using up-to-date technology. The skills and competences developed may relate directly to the technology, or they may be the softer skills (such as team-working and creativity) that are needed to work effectively in such an environment. The approach involves the use of technology and/or pedagogy that is innovative in the context and therefore requires flexible and creative approaches to its implementation.

In the context of the *Teach4Edu* project, we proposed the concept of *Joint Creative Classroom (JCC)* to implement *Education 4.0* principles. JCC courses meet the demand for specialized education not offered by a single university, pooling resources from multiple institutions for a well-rounded learning experience. This matches the growing collaboration possibilities opened by Erasmus Mundus[1], European's Universities[2], and Erasmus plus national and European

[1] https://erasmus-plus.ec.europa.eu/opportunities/opportunities-for-individuals/students/erasmus-mundus-joint-masters

[2] https://education.ec.europa.eu/education-levels/higher-education/european-universities-initiative

programs (Czerska-Shaw and Krzaklewska, 2022; Fernández, 2009; Beier et al., 2012). In this context, *JCC courses* also prepare students for *GSE* practices, fostering collaboration, communication, and problem-solving skills in diverse, cross-cultural environments.

This paper presents the *JCC* methodology and reports the experience of implementing a *JCC* course titled *Software Architecture Analysis and Design*, jointly created by the *University of L'Aquila (UNIVAQ)* and the *University of Zagreb (UNIZG)*. This course builds upon the foundational courses already provided by the respective universities, enabling students from related courses not only to partially participate in each other's classes but also to work together in cross-disciplinary teams on a project, fostering the exchange of knowledge and comprehension of software architecture by *osmosis* (Odom and Kelly, 2022; Lago et al., 2012). The collaborative initiative spans the entire spectrum of software development, from architectural design to implementation and analysis, providing students with a comprehensive and immersive learning experience implementing the Education 4.0 principles.

The rest of the paper is structured as follows: Section 2 presents the related works on pedagogy methodologies, Education 4.0, and their application in the context of *GSE*. Section 3 introduces the *JCC* model and how it has been applied for the design and implementation of the *Software Architecture Analysis and Design JCC course*. Section 4 show how the implementation of the *JCC* matches the Education 4.0 principles, as well as the main challenges encountered. Finally, Section 5 concludes the paper.

2 State of the Art

In this Section, we analyse the state of the art in pedagogy methodologies in the context of computer science and *Global Software Engineering (GSE)* and Education 4.0. Finally, based on the provided analysis, we discuss the motivation and scope of this paper.

Computer Science, Pedagogy and Education 4.0. In the past years, different of systematic literature reviews (SLRs) on computer science teaching and innovative pedagogical methodologies have been published.

Aničić et al. (2016) conducted a meta-analysis encompassing 155 publications from 1980 to 2014, providing insights into the current research landscape regarding the education and career trajectories of graduates in the field of Information and Communication Technology (ICT). The findings underscore the necessity for curriculum design and implementation to be adapted to align with industry requirements. As noted by the authors, existing literature highlights the imperative for innovative strategies in curriculum design and delivery, including the establishment of competency-based programs that transcend traditional semester-based frameworks, facilitating a more flexible curriculum that minimizes classroom time, and offering courses focused on both innovation management and the process of innovating itself. Effective pedagogical methods that can enhance graduate employability include experiential learning, learning through error, teamwork, and collaborative education. Additionally, approaches such as job-oriented experimental courses, problem-based or project-based learning, and work-integrated learning are advocated to cultivate a comprehensive array of skills, competencies, and knowledge that align with the principles of Education 4.0.

Garousi et al., (2019) specifically examined the alignment of software engineering education with industrial requirements. Their SLR analysed 34 articles published between 1995 and 2018 and identified eight pertinent research questions, two of which are highly relevant to our investigation: (1) *What curriculum models (bodies of knowledge) have been utilized in the design of the studies?;* and (2) *What educational recommendations are articulated in each study?* In their study, the authors revealed four thematic categories for the educational recommendations: (1) A need for greater emphasis on soft skills (20 papers), (2) An emphasis on active Infrastructure as Code (IAC) (3 papers), (3) A reduced focus on specific topics (2 papers), and (4) Additional recommendations (7 papers). To foster the development of soft skills, the authors advocated for incorporating real-world projects, establishing industry-academia collaborations in educational design, and anticipating future trends in student preparation.

Medeiros et al. (2018) reviewed 89 papers addressing introductory programming in higher education from 2010 to 2016, investigating how prior skills impact programming learning and the challenges students and educators encounter. The authors identified a significant gap in understanding problem-solving and prior knowledge, alongside the need for improved tools and methodologies for problem formulation and solution representation. In a broader review examining Artificial Intelligence (AI) and education, encompassing 425 papers published between 2000 and 2019, Guan et al. (2020) illustrated the diverse approaches employed to integrate technology into higher education; however, only a subset of these approaches has proven sustainable over time.

Overview of practices for integrating software testing within programming education based on an analysis of 195 empirical papers was performed by Scatalon et al. (2020). The study revealed varying levels of student engagement with testing practices in programming assignments, including analysis of test results from submission tools, interaction with instructor-provided tests, utilization of support mechanisms for test design (e.g., plugins for input and expected output), and instances of students crafting their tests. Nonetheless, few studies addressed the

mechanisms through which students acquired testing concepts in programming courses.

While these SLRs offer valuable insights into the intersection of computer science, programming education, and AI, none specifically address or incorporate the principles of Education 4.0. In addition, the COVID-19 pandemic, and the swift transition to online or hybrid education have significantly impacted education, necessitating a contemporary reassessment of how computer science educators embrace innovative pedagogies and Education 4.0 frameworks. This context is crucial for understanding the current state of the field and the need for innovative approaches.

Consequently, two systematic literature reviews were undertaken within the context of our *TEACH4EDU4* to investigate how computer science educators design and implement innovative methodologies and technologies consistent with Education 4.0 principles. In an initial exploratory study involving 20 studies across Europe, Rienties et al. (2021) found that most European educators employed only one or two of nine Education 4.0 elements presented by Hussin (2018).

A subsequent study (Rienties et al., 2023) examined how current pedagogical approaches in Computer Science align with the principles of Education 4.0, reviewing 66 articles incorporating an average of several key characteristics from Education 4.0. The study confirmed the existence of three distinct clusters in how computer science educators design innovative courses: 1) EDU 4.0 light, 2) project-based/hands-on learning, and 3) full EDU 4.0. In the EDU 4.0 light category, the focus was primarily on fostering independent learning, offering flexibility for learning anytime and anywhere, personalized learning, and allowing students choice in how to learn.

However, there was minimal emphasis on hands-on learning and no project-based learning. The project-based/hands-on learning cluster prioritized project-based and hands-on activities, but placed less emphasis on personalized learning, flexibility in how to learn, and learning anytime and anywhere. The third cluster, labelled full EDU 4.0, emphasized hands-on learning, independent learning, personalized learning, flexibility in learning anytime and anywhere, and student choice in how to learn.

Education in the context of Global Software Engineering. The dynamic landscape of education has witnessed a surge in initiatives aimed at fostering synergistic learning experiences, often facilitated through collaborative methodologies.

Bosnić et al. (2019) investigated the effects of Distributed Software Development Course (DSD) on the early career stages of software engineers. Results proved the beneficial effects of this initiative on professional careers. as almost 70% of the students engaged in distributed collaborations.

Report on experiences using Project-Based Learning to teach virtual, cooperative courses in global software engineering was given by Fu et al. (2018). The study provides a comparative analysis of bachelor's and master's level courses, shedding light on the challenges faced by students at various stages of their academic journey.

Marutschke et al. (2020) share experiences teaching global software engineering through distributed, virtual courses. They evaluate the effectiveness of project-based learning methods and collaborative platforms, providing best practices for similar courses. This work emphasizes the importance of effective coordination and cultural sensitivity in global software engineering education.

Bosnić and Čavrak (2019) investigate how distributed student teams in three European universities organize their project work on software products using the Scrum framework. Findings suggest that architectural choices positively impact work division, while younger students tend to create mixed-location sub-teams, leading to improved project performance.

Lago et al. (2008) outline a method for creating a collaborative course in global software development, leveraging diverse background knowledge. The authors introduce an orientation map to guide the development of a joint course, emphasizing cultural diversity.

Finally, the REXNet project in (Salah et al., 2018) involves universities in Kurdistan Region-Iraq and Oklahoma State University. Supported by IREX, it establishes distributed laboratories for collaborative experiments, modernizing education methods and promoting technological collaboration in the MENA region, particularly in Iraq.

Paper Motivation and Scope. Despite the analysis of the state of the art revealed that researchers put a lot of effort in studying the pedagogical aspects in computer science teaching and on how to apply them to promote GSE, for the best of our knowledge the adoption of the Education 4.0 framework for the design and implementation of software engineering related courses is still not investigated. Thus, the goal of this paper is to examine how the implementation of the *JCC* class aligns with the core principles of Education 4.0. Through a detailed analysis of the case study, we aim to identify which elements of Education 4.0 were addressed, how they were applied in practice, and their impact on student learning outcomes. We also seek to draw lessons from this case study to inform future educational practices within the framework of Education 4.0.

3 JCC Course Design and Implementation

This section discusses the implementation and design of the *Joint Creative Classroom* (*JCC*) course titled *Software Architecture Analysis and Design*. We show

Table 1. Education 4.0 Characteristics by Hussin

ID	Description	Addressed in our Experiment
P1	Learning can be taken place anytime anywhere.	Yes
P2	Learning will be personalized to individual student.	No
P3	Students have a choice in determining how they want to learn.	Yes
P4	Students will be exposed to more project-based learning.	Yes
P5	Students will be exposed to more hands-on learning through field experience.	Yes
P6	Students are required to apply their theoretical knowledge to numbers and use their reasoning skills to make inferences based on logic and trends from given sets of data.	No
P7	Students will be assessed differently and the conventional platforms to assess students may become irrelevant or insufficient.	Yes
P8	Students' opinion will be considered in designing and updating the curriculum.	Yes
P9	Students will become more independent in their own learning.	Yes

how the course addressed nine key Education 4.0 principles created by Hussin (2018) and reported in Table 1 by fostering global collaboration (P1) through cross-border teamwork and enabling self-paced learning (P3) via flexible structures and self-assessment tools. It emphasized project-based and hands-on learning (P4 and P5), allowing students to connect theory with practical, real-world challenges while developing critical soft skills like communication and teamwork. In addition, technology-enhanced learning environments and modern assessment methods (P7 and P8) promoted innovation and placed student feedback at the core of the educational experience. Finally, we discuss the effectiveness of the course reporting the results of pre- and post-surveys administered to participant students.

3.1 Course design

In today's increasingly globalized and technology-driven world, the *JCC* joint classes methodology offers an innovative approach to teaching software architecture. Developed within the framework of *Education 4.0*, this model emphasizes cross-university collaboration to prepare students for the complexities of modern, global, software engineering. The *JCC* course, titled *Software Architecture Analysis and Design*, was conceived as an interdisciplinary educational experience designed to merge theory and practice, as well as bridge gaps between institutions and academic disciplines. By bringing together students from two universities and engaging them in a unified learning experience, this course provides a comprehensive, immersive environment in which students develop both technical expertise and essential soft skills, such as communication, teamwork, and problem-solving, all within the context of global software engineering.

The ***Software Architecture course*** offered by *University of L'Aquila (UNIVAQ)* provides a deep dive into the foundational and advanced aspects of modern software architecture[3]. The course covers a wide range of topics central to the design and management of complex software systems. Students are introduced to *Components and Connectors*, which form the building blocks of software architecture, and learn to make critical *Architecture Design Decisions* through effective *Group Decision Making (GDM)* processes. The course explores various *Architectural Styles*, including *Cloud Architectures* and *Self-Adaptive Architectures*, equipping students to design distributed and scalable systems. Additionally, the course addresses the concept of *Technical Debt*, teaching students to recognize and manage trade-offs in system design to ensure long-term sustainability. Finally, *DevOps practices* are emphasized, highlighting collaboration between development and operations teams and providing students with hands-on experience using tools such as *Kubernetes* for container orchestration and *Kafka* for messaging systems. This course offers students the theoretical and practical skills required to design, assess, and manage complex software architectures.

The ***Software Analysis and Development course*** from *University of Zagreb (UNIZG)* focuses on the entire software development lifecycle, giving students a solid foundation in both conceptual and practical aspects of software development[4]. The course begins with an exploration of *Trends and Market Analysis*, providing insights into industry dynamics and emerging technologies. Students then learn a range of *Software Development Methodologies*, with an emphasis on Agile project management frameworks, which prepare them for managing iterative development processes. A key component of the course is its focus on *Mobile Development*, where students apply *SOLID Design Principles* to build

[3] *https://univaq.coursecatalogue.cineca.it/insegnamenti/2024/35828-1/2017/9/10013*

[4] *https://nastava.foi.hr/course/93066*

maintainable, scalable software systems with mobile, smart, wearable or IoT components. The course also covers *Non-Functional Requirements*, stressing the importance of performance, security, and usability in software design. Additionally, students learn about *Testing Automation* and *Continuous Integration*, ensuring that they develop skills in validating and integrating code efficiently. By the course's conclusion, students are well-versed in both technical design and business considerations, equipping them to face modern software development challenges.

Although the individual courses provide comprehensive coverage of their respective domains, neither course offers a holistic view of developing complex software products, which inherently involve methodological, architectural and implementation demands. Thus, the objective of the ***Software Architecture Analysis and Design,*** *the JCC course* we designed, was to integrate the distinct yet complementary content of the two described courses and to create a learning experience that blends theoretical knowledge with practical application in both domains allowing students from both universities to work together and contribute to the same project from their different perspectives[5] (P4, P5). Using *learning design tools* and relying on the *complement model*, the course was structured to ensure a cohesive educational experience despite the geographical separation of the students and the differences in the institutions' curricula. This integration also allowed students to benefit from the strengths of both institutions, providing a holistic view of software architecture and development, but also aligns with Education 4.0 characteristic of enabling students learning anytime and anywhere (P1), even with students from different country.

Special attention was devoted to planning and developing the *learning design (LD)* for the *JCC course*. Learning objectives were clearly defined to align the content from both universities. This process involved a comprehensive analysis of learning outcomes but also themes and topics that will contribute to their implementation. The modes of delivery of the course content as well as the activities that were to be performed by students or by teams were carefully designed. By collaborating with industry representatives, we selected teaching and evaluation methods that best aligned with both student needs and industry expectations (P7 and P8). This curriculum design effectively integrated practical skills with theoretical concepts, creating a well-rounded educational program that prepares students for the job market while ensuring compliance with the official accreditation standards required by both universities. In this way students will become more independent in their own learning as expected in Education 4.0 (P3).

The tool used for learning design is *Balanced Design Planning Tool* (*BDP Tool*)[6] which is research-based, constructive-alignment oriented and modular solution based on *learning outcomes* (*LOs*) and learner workload as foundations of learner-centred learning approach and principles (Divjak et al., 2022). An excerpt from the tool showing the detailed plan of one teaching topic with several teaching units and different teaching and learning activities is presented in Fig. 1, while the excerpt from the learning design analysis given by the tool is presented in Fig. 2.

Course schedules were synchronized to provide a logical progression of topics, enabling students to seamlessly move from one area of study to another. Collaborative platforms were utilized for sharing course materials and resources, while communication tools like video conferencing and email facilitated interactions between students and instructors across institutions. Weekly or bi-weekly meetings with instructors ensured that students' progress was continuously monitored, allowing for real-time feedback and adjustments (P8).

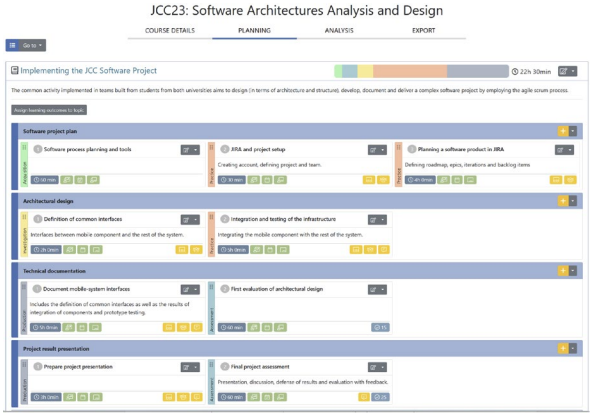


Figure 1. Excerpt from the learning design topic definition

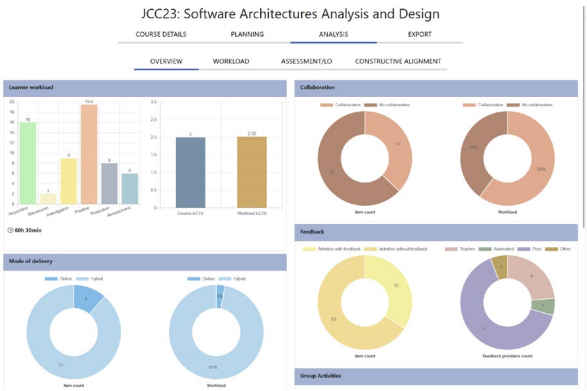


Figure 2. Excerpt from the learning design analysis

To elucidate the design choices, we will highlight which collaboration models were adopted for each dimension of the course design, namely content, attendance, and assessment, shedding light on the

[5] *https://learning-design.eu/en/preview/81f9a72eb213cf3e3d1b2ea2/details*

[6] *https://learning-design.eu/*

Table 2. Adopted Models for JCC Course Design

	Content	**Attendance**	**Assessment**
Collaboration Model	Complement	Fusion	Complement Theory/Practice
Description	Introduction to Software Architecture, Components and Connectors, and Technical Debt from UNIVAQ. Trends and Market Analysis, SCRUM, and SOLID design principles, provided by UNIZG.	Students from UNIVAQ attended part of the classes provided by UNIZG. Conversely, students from UNIZG attended part of the classes from UNIVAQ.	Each student evaluated through a self-assessment approach. Mixed teams, comprising students from both universities, undertook distinct software architecture and analysis and design projects. These projects were evaluated collectively, recognizing individual contributions while celebrating the collective achievements of teams.

comprehensive framework that underpins our *JCC* course. Table 2 presents the models selected for each dimension providing a brief description of how they have been adopted in the design of the *Software Architecture Analysis and Design* course.

3.2 Course implementation

The implementation of the *JCC* course took place over one (winter) semester, but with staggered start dates between the universities. This required careful coordination to ensure that students from both institutions could engage in the joint components of the course without disruption. The course engaged three primary roles: students, instructors, and practitioners each playing integral parts in the collaborative learning experience.

Students. The course drew participation from a diverse cohort of students, representing both *UNIVAQ* and *UNIZG*. Notably, the *Software Architecture* course at *UNIVAQ* saw an approximate enrolment of 40 students, while the *Software Analysis and Design* course at *UNIZG* garnered approximately 90 students. At the outset, both groups were introduced separately to the *JCC* course, allowing them to familiarize themselves with the content and expectations. Subsequently, students were given the opportunity to express their interest in joining the *JCC* course. Six out of the 40 students from *UNIVAQ* and 18 out of the 90 students from *UNIZG* opted to embark on this cross-university educational journey. Students were organized into mixed teams, with each team comprising members from both universities.

Instructors. The instructional team consisted of two educators, one from each participating university, who played crucial roles in managing the *JCC* course. Through a mix of weekly synchronous and asynchronous meetings, they delivered lectures and exercises, monitored the course's progress, and assessed how students responded to the collaborative approach. This constant interaction allowed for real-time adjustments, ensuring the course adapted to the students' evolving needs (P8). To accommodate differences in the start dates of the courses, the instructor from *UNIVAQ* replicated key lectures for *UNIZG* students, and vice-versa. Where possible, both educators worked together to deliver joint sessions on essential topics, ensuring a comprehensive learning experience for all students, enabling smooth integration of both student group into the *JCC* course.

Practitioners. The role of the industry mentors was not only to provide the students with suitable industry-related project assignments, but also to provide real-world insights, expertise, and guidance to students, helping bridge the gap between academic knowledge and practical industry experience (P5). Their role involved mentoring students, offering industry-specific feedback, and helping to align the project with current industry standards and best practices. Most of the company's activities were conducted online, but after the completion of the projects, students from *UNIZG* had the opportunity to present their solutions to the management and other employees on-site at the company.

Attendance. In terms of attendance, the course followed a *fusion* model. Students attended their respective courses at their home universities but also participated in selected sessions from the partner institution. For example, students from *UNIVAQ* attended modules on SCRUM and SOLID design principles delivered at *UNIZG*, while students from *UNIZG* attended sessions on Components and Connectors and Technical Debt delivered at *UNIVAQ*. This cross-institutional exposure allowed students to experience diverse teaching styles and content, enriching their subject matter understanding. It is important to note that, as the *Software Architecture* and *Software Analysis and Development* courses were designed to provide 6 *ECTS* credits[7] to students, students were awarded an additional 2 *ECTS* credits for their participation in the *JCC* course.

Projects. To give the students the possibility to work on an industry related and real-life projects, the implementation of the course included the cooperation with industry partner – *Cetitec Ltd.* It is a company with a strong focus on developing an automotive communication embedded software. The company already left a deep impact on the automotive industry,

[7] ECTS - European Credit Transfer and Accumulation System

with many vehicles on the road using the company's products. As the future sets in, modern technologies and agendas require the whole industry to even enlarge the presence of various software solutions inside the vehicles, with specific requirements targeting a communication part.

In this context, three highly demanding projects were given to the students to develop viable solutions:

- Prototype of remote control of the car and its components via mobile application.
- Prototype of monitoring and configuration of car-to-car communication via mobile application.
- Prototype of autonomous driving information and configuration via mobile application.

The primary objective of the projects assigned to the students was to facilitate the development of a demonstration highlighting the seamless incorporation of contemporary technologies commonly employed in conventional automobiles into remotely operated vehicles. This integration was complemented by the introduction of a user interface and control mechanism via a mobile application. These solutions were intended not only as showcases but also as experimental platforms, fostering the exploration of novel concepts and innovations, which could subsequently undergo further refinement and integration into actual automotive systems.

It is noteworthy to underscore that the projects were designed to be autonomous, yet capable of harmonious integration, and each team had the discretion to opt for hardware simulation or physical prototyping. Moreover, the projects required students to integrate learnings from both courses, applying SCRUM methodologies to real-world challenges such as designing software solutions for the automotive industry. The industry relevance of these projects ensured that students were not only engaging with theoretical concepts but also learning to apply them in practical, tangible ways. Such setup exposed students to more hands-on learning through field experience which also contributed to the alignment of our course with Education 4.0 requirements (P5).

Teams. A total of 24 students were organized into six teams, with each project topic being concurrently assigned to two distinct teams. These teams were encouraged to engage in a collaborative exchange of knowledge and ideas, while concurrently maintaining the expectation of delivering a unique, independently authored solution.

Assessment model. The course was assessed using a *complement theory / practice model*, blending individual evaluations with collaborative project-based assessments. Students were assessed on their theoretical knowledge through quizzes and self-assessment tools specific to their home institution's course. More importantly, the assessment also included a collaborative element, where mixed teams of students from both universities worked on joint projects. The acquisition of the knowledge was assessed continuously during the semester. The process involved the usual assessment activities performed at both universities as well as the assessment of the joint elements of the course through self-assessment and project assessment (P7).

Self-assessment. Self-assessment was executed through a sequence of concise examinations intricately synchronized with the instructional content delivered by the educators. Access to these self-assessment activities was conditional upon students' prior completion of prerequisite tasks associated with knowledge acquisition, discussion, or investigation within the relevant domain. Using the tools from the learning management system, teachers prepared the pool of questions (approximately 20 questions per lesson) and designed the assessment elements (number of questions, duration, time frame for approaching the assessment, question and answer randomization options, feedback, the use of Safe exam browser etc.). During the assessment, each student was presented with unique 5-questions quiz which students had to answer in the timeframe of no more than 10 minutes.

The findings presented in Table 3 revealed a notably elevated level of intrinsic motivation among participants engaged in knowledge acquisition through the prescribed activities. All enrolled students diligently completed the self-assessment exercises. It is noteworthy that these results exhibit a considerable enhancement when compared to the performance of traditional course participants at their respective home universities.

Table 3. Summary of Assessment Results

Assessment	Pass Rate (%)	Average Score (%)
SOLID design principles	95.83	87.0
Technical debt	91.67	83.20
Software architecture	87.50	83.6
SCRUM agile framework	83.33	87.6

Assessment of projects. The assessment framework delineated three separate evaluations: the initial assessment conducted at the commencement of the semester, subsequent to the establishment of the agile project; mentoring and the intermediate evaluation (see Fig. 3), positioned at the midpoint of the semester following the conclusion of the research and architectural design phase by the teams; and the concluding assessment, which took the form of a public project solution defence held at the end of the semester. The novel approach in the assessment which did not rely on the conventional assessment methods emphasizes the assessment model as perceived by Education 4.0 (P7).

Considering the demanding nature of the projects, characterized by their multifaceted requirements encompassing intricate architectural design, the necessity for simulation or hardware prototyping, and the imperative involvement of embedded, backend,



Figure 3. Presentation of the solution mockups during the mentoring session

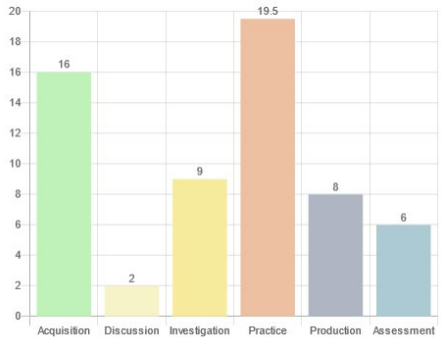


Figure 4. Distribution of student's workload in hours across different activity learning types

and mobile development, coupled with the demand for effective agile teamwork and extensive documentation, it became evident that a total of four out of the six teams succeeded in producing comprehensive solutions within the duration of the semester. Regrettably, the remaining two teams faced initial setbacks, failing to meet the requirements during the first examination period, and consequently, they were required to continue their work after the *JCC* course had concluded.

Alignment with Full Edu 4.0 learning design model. The distribution of student workload across different learning activities is shown in Fig. 4. Our course places significant emphasis on acquisition (27,67 hrs), practice (23.5 hrs), and investigation (11.17 hrs), in relation to three types of how pedagogical approaches in Computer Science align with the principles of Education 4.0 as detailed in (Rienties et al., 2023).

This model emphasizes experiential learning, self-directed study, personalized education, flexibility to learn anytime and anywhere, and providing students with the autonomy to choose their learning methods. Further information on how the *JCC* course matches with the Education 4.0 principles are provided in Section 4.

3.3 The effectiveness of the course

The effectiveness of the *JCC* course was assessed through pre- and post-surveys administered to participating students. The surveys aimed to gauge their perceptions and experiences in various aspects of the course focusing on *real-word skills development*, *soft skills development*, and the evaluation of the *assessment process*. The following subsections present a comparative analysis of the pre- and post-survey responses.

Real-world Skills Development. The results of the pre and post surveys on the course demonstrate an interesting shift in students' perspectives regarding gaining hands-on, authentic experiences, and real-world skills as depicted in Fig. 5 (P5). Before the course, a considerable number of students expressed a neutral opinion, indicating that they were unsure about their ability to gain such experiences. However, post-course surveys show that a substantial number of students have shifted their perspective positively. A considerable percentage of students have moved from the "Neutral" category to the "Agree" and "Strongly Agree" categories, reflecting a tangible increase in their confidence that they have indeed gained hands-on, authentic experiences, and real-world skills through the course.

The data strongly supports the efficacy of the course in providing students with the hands-on experiences and practical skills sought after in the field of software engineering. The shift from a more neutral stance in the pre-survey to a highly positive one in the post-survey underscores the course's ability to meet the students' expectations and aspirations for authentic learning experiences.

Soft Skill Development. The survey results demonstrate a positive shift in students' perceptions regarding the development of their soft skills, specifically in areas such as team-working and creativity as depicted in Fig. 6 (P4). Prior to the course, most students (13) expressed agreement with the statement, indicating an initial interest in improving these skills. Following the course, this sentiment was further reinforced, with an even larger number (22) of students indicating that they were able to strengthen their soft skills. Notably, a sizeable portion of students (11) were initially neutral on the matter, suggesting a potential for growth in this area.

The absence of responses in the "Strongly Disagree" and "Strongly Agree" categories both before and after the course indicates a consistent middle ground. This suggests that, while some students may not have felt strongly about their soft skill development prior to the course, the majority did see positive progress in this regard by the course's conclusion.

Innovative Assessment. The data from the pre- and post-surveys regarding the assessment methods provides interesting insights into the students'

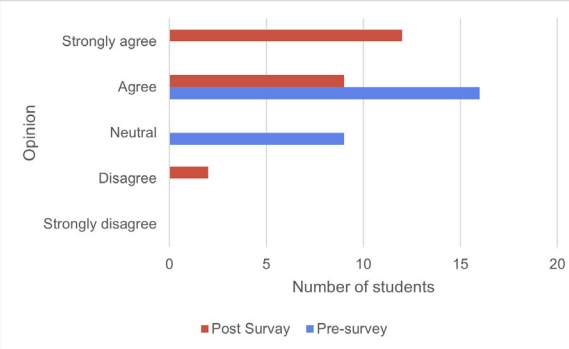


Figure 5. Hands-on Experiences and Real-world Skills

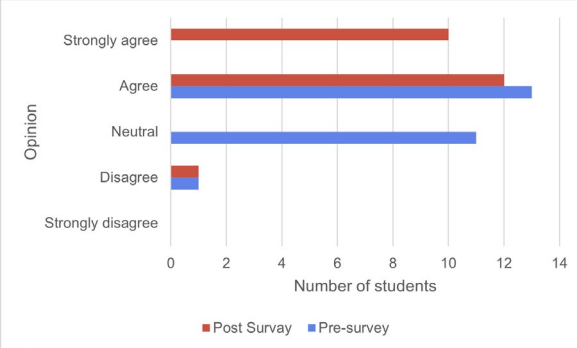


Figure 6. Development of soft skills

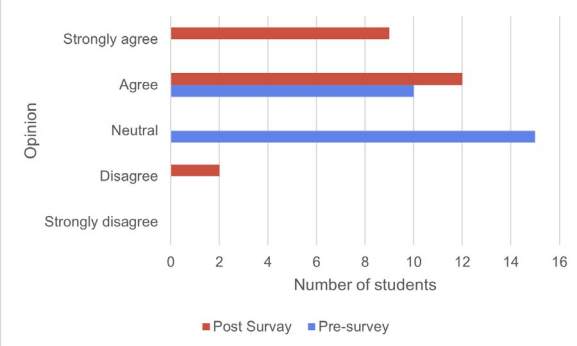


Figure 7. Innovative assessment

perspectives on innovative evaluation techniques as reported in Fig. 7 (P7). Before the course, most students were in a neutral position, with 15 expressing neither agreement nor disagreement with the statement. This suggests a certain level of uncertainty or a lack of prior exposure to innovative assessment methods.

After the course, we observe a notable shift in opinions. A considerable number of students (12) now agree that they have participated in new ways of assessment, indicating a positive change in their perception. Additionally, nine students strongly agree, demonstrating a ringing endorsement of the innovative assessment methods introduced during the course.

This data suggests that the course successfully introduced students to innovative assessment techniques, leading to a favourable shift in their perspectives.

Students Level of Satisfaction. The results of the post-survey indicate an elevated level of satisfaction and positive feedback from the students. A significant majority, 15 out of 21, expressed satisfaction with the JCC course, emphasizing its effectiveness in meeting their expectations. Even more encouraging, 16 out of 21 students expressed a willingness to recommend the course to their peers, underscoring the value they saw in the collaborative learning experience. Additionally, the unanimous agreement among all students regarding the support provided by the teacher, who acted as a facilitator, highlights the effectiveness of this teaching approach in fostering a conducive learning environment. These results affirm the success of the *JCC* course in delivering a meaningful and impactful educational experience.

Other Considerations. The continuous assessment model included mentoring by company experts and university instructors as part of the project work, as well as theoretical knowledge assessments through online tests created on the e-learning platform.

Considering *UNIZG* students, the average final grade of JCC students is 64,4% and pass rate is 100% while the average final grade of non-JCC students is 64,2% and pass rate is only 82,42%, while 17,58% failed the course and had to enrol the course next academic year.

Concerning *UNIVAQ* students, the average final grade of JCC students is 88,3%. Five of six students enrolled in the JCC course pass the exam in the first session (83%). On the other hand, non JCC students scored, in average, 93%, but only the 79% of the students completed the exam by the first session. It is important to notice that *UNIVAQ* JCC students scored worst due to synchronization problem with students from *UNIZG*.

4 Discussion

The *JCC* course, designed collaboratively between two European universities, provides an interesting case study of the practical application of Education 4.0 principles in higher education. The course emphasized interdisciplinary learning, cross-border collaboration, and the integration of advanced technologies to prepare students for the dynamic and global landscape of software engineering. In this section, we explicit how the design and implementation of the course matches the Education 4.0 core principles, as well as the encountered main challenges.

4.1 The *JCC* Course and Education 4.0 Principles

Principle P1: Cross-Institutional Collaboration. The course leveraged the fusion model, where students from two different universities participated in joint

classes and projects. This experience exposed students to diverse viewpoints, promoting the exchange of knowledge across disciplines and borders. This aligns with the idea that learning can occur anywhere (P1), as it breaks down physical barriers and prepares students for the globalized, interconnected nature of modern work environments. In addition, the cross-university collaboration not only enhanced students' domain-specific and industry-ready knowledge but also equipped them with essential soft skills such as teamwork, communication, and problem-solving within a global software engineering context. Indeed, this collaborative approach also mirrored the needs of the global software engineering (GSE) industry, where professionals often work in geographically distributed teams. It enabled students to simulate real-world collaborative environments, further aligning their skills with industry expectations.

Central to the success of the joint course was the implementation of technology-enhanced learning. Various digital platforms were employed to facilitate seamless communication, coordination, and learning. The use of video conferencing, shared platforms for course materials, and real-time project management tools enabled efficient collaboration despite geographical distances. This aligns with the Education 4.0 principle of promoting learning anywhere (P1), thus breaking down traditional barriers to education and fostering flexibility.

Principles P4 and P5: Project-Based and Hands-On Learning. A standout feature of the course was its focus on project-based learning (P4). Students engaged in real-world projects in collaboration with industry partners, which simulated the complexities of modern software development. This approach provided them with hands-on experience, allowing them to apply theoretical knowledge in practical settings (P5). Industry-driven challenges, provided by an external partner, further enhanced the learning experience, giving students insight into the expectations and realities of the industry. Project-based learning aligns with the Education 4.0 framework, which emphasizes the need for learners to develop skills through experiential learning. By working on real-world projects, students were able to develop critical thinking, creativity, and innovation, all of which are essential skills for the future workforce.

One of the most significant educational strategies employed in the course was the *complementary theory/practice* model. The course seamlessly integrated theoretical lectures with practical applications, particularly using SCRUM methodology in software development projects. This approach allowed students to see the direct application of theoretical concepts to real-world scenarios, which is a fundamental tenet of Education 4.0 (P5). The fusion of theory and practice not only deepened the students' understanding of complex software architecture concepts but also prepared them for the agile work environments that are prevalent in the industry. This experiential learning approach is essential for ensuring that students can transition smoothly from academic settings to professional roles.

In addition, over 90% of the students involved in the course reported a significant improvement in their soft skills. This is a critical achievement in the context of Education 4.0, which places a strong emphasis on the development of interpersonal skills alongside technical proficiency through hands-on learning (P5). The collaborative nature of the course fostered communication, teamwork, and leadership skills, as students worked in mixed teams, combining diverse skill sets from the two institutions. Finally, problem-solving was a core skill developed through this course. By addressing real-world challenges and working in interdisciplinary teams, students were able to enhance their ability to analyse complex issues, propose solutions, and adapt to changing project requirements – skills that are indispensable in the ever-evolving technology industry.

Principles P3, P7, and P8: Self-Paced Learning, Assessment and Curriculum Design. The flexible structure of the course gave students autonomy over how they wanted to engage with the material, giving them choices on pacing and learning pathways. This is a direct reflection of the principle that students should have control over how they learn, promoting more self-directed learning experiences (P3).

The use of self-assessment tools marks a shift from traditional assessment methods, providing a more dynamic way for students to evaluate their progress. This aligns with modern approaches where conventional assessment platforms become less relevant, and student feedback becomes key in shaping future learning experiences (P7).

Finally, the usage of tools for course design based on learning design and constructive alignment methods, the self-assessment tools and the pre- and post-surveys allowed the instructor to design and update the curriculum based on the student's knowledge and learning outcomes (P8).

4.2 Challenges

Despite the overall success of the course, several challenges and lessons were identified that highlight areas for improvement in future implementations of the joint course model.

Missed Deadlines and Project Overload. A significant challenge observed was that not all teams were able to meet the project deadlines. Specifically, two out of the six teams failed to deliver their project solutions by the end of the semester, necessitating continued work beyond the duration of the course. The complex and demanding nature of the assigned projects, which included intricate architectural designs, embedded systems, and extensive documentation requirements, may have contributed to this outcome. The scope and workload should be carefully managed

to ensure that all students can complete the tasks within the given timeframe.

Synchronization of Course Schedules. Another challenge faced was the difficulty in synchronizing the course schedules between the two participating universities. The universities had different start dates for the semester, which caused disruptions in the delivery of some joint lectures and delayed project kick-offs. This issue highlights the importance of international alignment of academic calendars or the incorporation of flexibility into the course structure to accommodate such discrepancies.

Effort to Bridge Intercultural Differences. Given the international nature of the course, students from diverse cultural backgrounds participated. Some intercultural communication challenges emerged, necessitating extra effort from both students and instructors to overcome differences in work habits, communication styles, and academic expectations. While this provided valuable learning experiences, it also required additional support mechanisms, such as increased mentoring and frequent meetings, to ensure smooth collaboration.

Coordination and Assessment Complexity. The joint course model involved two universities and an industry partner, which required clear coordination of roles in the assessment process. However, students from different courses possessed varying levels of technical expertise, leading to difficulties in establishing uniform assessment criteria. Balancing the evaluation of theoretical and practical skills across diverse student backgrounds remains a challenge that needs to be addressed more effectively in future courses.

5 Conclusion

The implementation of the *JCC* joint course represents a successful application of Education 4.0 principles in software engineering education. By fostering interdisciplinary, cross-university collaboration, and integrating technology-enhanced and project-based learning, the course provided students with a holistic and practical experience that mirrors the demands of the global software engineering industry.

An important aspect of the course's success was the careful use of *learning design* methodology and tools. By using structured learning design methodology, the course effectively integrated diverse content and learning objectives from two universities into a cohesive educational experience. This planning ensured that students engaged in a balanced mix of theory and practice, with clear learning outcomes aligned with industry needs. The thoughtful design of activities, assignments, and assessments played a crucial role in facilitating personalized learning experiences and optimizing student engagement across institutions.

Key lessons learned from the course highlight both the strengths and challenges of this innovative educational approach. The development of students' technical and soft skills, especially in real-world problem-solving, communication, and teamwork, was a notable success. Additionally, the use of advanced technologies facilitated flexible, cross-border learning, aligning with the core tenets of Education 4.0.

However, challenges such as schedule synchronization between universities, managing intercultural differences, and ensuring timely project completion underscore the need for careful planning and coordination in future implementations. Addressing these challenges will further improve the effectiveness of such courses and strengthen the alignment with Education 4.0 objectives.

In conclusion, the *JCC* joint course provides valuable insights into how software engineering education can evolve to meet the needs of the modern workforce. By embracing the principles of Education 4.0, it creates a model for future educational initiatives that can better prepare students for the complexities of the global and digital economy.

References

Angelov, S., & de Beer, P. (2017). Designing and applying an approach to software architecting in agile projects in education. *Journal of Systems and Software, 127,* 78–90.

Aničić, K. P., Divjak, B., & Arbanas, K. (2016). Preparing ICT graduates for real-world challenges: Results of a meta-analysis. *IEEE Transactions on Education, 60*(3), 191–197.

Beier, S., Bickel, M., Brockmann, P., & Choinzon, M. (2012). It takes a global village to teach global software engineering: A Mongolian-German team-teaching project. *In 2012 International Conference on E-Learning and E-Technologies in Education (ICEEE)* (pp. 152–157). IEEE.

Bonfield, C. A., Salter, M., Longmuir, A., Benson, M., & Adachi, C. (2020). Transformation or evolution? Education 4.0, teaching and learning in the digital age. *Higher Education Pedagogies, 5*(1), 223–246.

Bosnić, I., & Čavrak, I. (2019). Project work division in agile distributed student teams—Who develops what? In *2019 ACM/IEEE 14th International Conference on Global Software Engineering (ICGSE)* (pp. 162–171).

Bosnić, I., Čavrak, I., & Žagar, M. (2019). Assessing the impact of the distributed software development course on the careers of young software engineers. *ACM Transactions on Computing Education (TOCE), 19*(2), 1–27.

Czerska-Shaw, K., & Krzaklewska, E. (2022). Uneasy belonging in the mobility capsule: Erasmus Mundus

students in the European Higher Education Area. *Mobilities, 17*(3), 432–445.

Divjak, B., Grabar, D., Svetec, B., & Vondra, P. (2022). Balanced learning design planning: Concept and tool. *Journal of Information and Organizational Sciences, 46*(2), 361–375.

Fernández, O. B. (2009). A computer sciences teaching experience in the Erasmus Mundus Program. In *INTED2009 Proceedings* (pp. 2566–2570). IATED.

Fisk, P. (2017). Education 4.0 … the future of learning will be dramatically different, in school and throughout life. *Retrieved from https://www.peterfisk.com/2017/01/future-education-young-everyone-taught-together/*

Fu, Y., Reina, L. P., & Brockmann, P. (2018). Teaching global software engineering: Experience report comparing distributed, virtual collaborative courses at the bachelor's and master's degree levels. In *Proceedings of the 3rd European Conference of Software Engineering Education* (pp. 34–38). ACM.

Garousi, V., Giray, G., Tüzün, E., Catal, C., & Felderer, M. (2019). Aligning software engineering education with industrial needs: A meta-analysis. *Journal of Systems and Software, 156,* 65–83.

Guan, C., Mou, J., & Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies, 4*(4), 134–147.

Harkins, A. M. (2008). Leapfrog principles and practices: Core components of education 3.0 and 4.0. *Futures Research Quarterly, 24*(1), 19–31.

Herbsleb, J. D., & Moitra, D. (2001). Global software development. *IEEE Software, 18*(2), 16–20.

Hussin, A. A. (2018). Education 4.0 made simple: Ideas for teaching. *International Journal of Education and Literacy Studies, 6*(3), 92–98.

Iniesto, F., Sargent, J., Rienties, B., … Muccini, H. (2021). When industry meets Education 4.0: What do computer science companies need from higher education? In *Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21)* (pp. 367–372). ACM.

Jisc. (2019). Education 4.0 – transforming the future of education (through advanced technology). [Video]. YouTube. https://www.youtube.com/watch?v=aVWHp8FsV1w

Lago, P., Muccini, H., & Babar, M. A. (2008). Developing a course on designing software in globally distributed teams. In *2008 IEEE International Conference on Global Software Engineering* (pp. 249–253). IEEE.

Lago, P., Muccini, H., & Babar, M. A. (2012). An empirical study of learning by osmosis in global software engineering. *Journal of Software: Evolution and Process, 24*(6), 693–706.

Marutschke, D. M., Kryssanov, V. V., & Brockmann, P. (2020, January). Distributed virtual courses to teach global software engineering: Lessons learned and best practices. In *Proceedings of the 2020 11th International Conference on E-Education, E-Business, E-Management, and E-Learning* (pp. 256–260). ACM.

Medeiros, R. P., Ramalho, G. L., & Falcão, T. P. (2018). A systematic literature review on teaching and learning introductory programming in higher education. *IEEE Transactions on Education, 62*(2), 77–90.

Moraes, E. B., Kipper, L. M., Hackenhaar Kellermann, A. C., Austria, L., Leivas, P., Moraes, J. A. R., & Witczak, M. (2023). Integration of Industry 4.0 technologies with Education 4.0: Advantages for improvements in learning. *Interactive Technology and Smart Education, 20*(2), 271–287.

Mukul, E., & Büyüközkan, G. (2023). Digital transformation in education: A systematic review of education 4.0. *Technological Forecasting and Social Change, 194,* 122664.

Odom, A. L., & Kelly, P. V. (2001). Integrating concept mapping and the learning cycle to teach diffusion and osmosis concepts to high school biology students. *Science Education, 85*(6), 615–635.

Rienties, B., Balaban, I., Ferguson, R., Herodotou, C., Iniesto, F., Muccini, H., … Virkus, S. (2021). Education 4.0 and computer science: A European perspective. In *Proceedings of the Central European Conference on Information and Intelligent Systems* CECIIS (pp. 139–146)

Rienties, B., Ferguson, R., Gonda, D., Hajdin, G., Herodotou, C., Iniesto, F., … Isidori, M. V. (2023). Education 4.0 in higher education and computer science: A systematic review. *Computer Applications in Engineering Education, 31*(5), 1339–1357.

Salmon, G. (2019). May the Fourth be with you: Creating Education 4.0. *Journal of Learning for Development, 6*(2).

Scatalon, L. P., Garcia, R. E., & Barbosa, E. F. (2020). Teaching practices of software testing in programming education. In *2020 IEEE Frontiers in Education Conference (FIE)* (pp. 1–9). IEEE.

Salah, R. M., Cecil, J., & Atrushi, D. (2018, October). Collaborative remote laboratories for serving sciences and engineering education in Iraq: REXNET project. In *2018 International Conference on Advanced Science and Engineering (ICOASE)* (pp. 134–139). IEEE.

Publication Ethics And Malpractice Statement

Central European Conference on Information and Intelligent Systems – CECIIS is committed to upholding the highest standards of publication ethics and takes all possible measures against any publication malpractices for its double-blind peer-reviewed Conference Proceedings.

1. Duties of Editors and Track Directors

Editors are responsible for maintaining the integrity of the academic record, for having processes in place to assure the quality of the material they publish and for precluding business needs from compromising intellectual and ethical standards.

Track Directors are responsible that appropriate reviewers are selected for submissions (i.e. individuals who are able to judge the work and are free from disqualifying competing interests).

Publication decisions

Editors and track directors are responsible for deciding which manuscripts submitted to CECIIS will be accepted for presentation and for publication of the final paper in the Conference Proceedings. This decision is based on the recommendation of reviewers. Main selection criteria are the contribution's importance, originality and clarity, and the study's validity and its relevance. Decisions will not be influenced by the authors' race, gender, sexual orientation, religious belief, ethnic origin, citizenship, or political philosophy. Editors and track directors may be guided by the policies of the CECIIS Program Committee and constrained by such legal requirements as shall then be in force regarding libel, copyright infringement and plagiarism. Editors and track directors may confer with reviewers in making this decision.

Confidentiality

Editors and track directors must not disclose any information about a manuscript submitted to anyone other than the corresponding author, reviewers, potential reviewers, other editorial advisers, and the publisher, as appropriate.

Disclosure and conflicts of interest

Unpublished materials disclosed in a submitted manuscript will not be used in an editors´ or track directors' own research without the express written consent of the author.

2. Duties of Reviewers

Contribution to editorial decisions

Papers will be published in the Conference Proceedings after double-blind peer-reviewed process. The reviewers advise track directors and editors. Track directors communicate with authors, as required, and help them in improving the quality of their manuscript. The reviewers don't know the author's identity and reviewers' comments to track directors and editors are confidential and before passing on to the author will be made anonymous. The names of the reviewers remain strictly confidential; with their identities known only to editors and corresponding track directors.

Promptness

Any selected reviewer who feels unqualified to review the research reported in manuscript or knows that its prompt review will be impossible, should notify the track director and withdraw from the review process.

Confidentiality

Any manuscripts received for review must be treated as confidential documents. They must not be disclosed to or discussed with others except as authorized by editors.

Standards of objectivity

Reviews should be conducted objectively. Personal criticism of the author is inappropriate. Reviewers should express their views clearly with supporting arguments, if necessary with explanation.

Acknowledgement of sources

Reviewers should identify relevant published work that has not been cited by authors. They should point out whether observations or arguments derived from other publications are accompanied by the respective source. Reviewers will notify track directors or editors of any substantial similarity or overlap between the manuscript under consideration and any other published paper of which they have personal knowledge.

Disclosure and conflict of interest

Privileged information or ideas obtained through the review process must be kept confidential and not used for personal advantage. Reviewers should not consider manuscripts in which they have conflicts of interest resulting from competitive, collaborative, or other relationships or connections with any of the authors, companies, or institutions associated with the papers.

3. Duties of Authors

Reporting standards

Authors of manuscripts should present an accurate account of the work performed as well as an objective discussion of its significance. Underlying data should be represented accurately in the manuscript. A manuscript should contain sufficient detail and references to permit others to replicate the work. Fraudulent or knowingly inaccurate statements constitute unethical behavior and are unacceptable.

Data access and retention

Authors could be asked to provide the raw data in connection with a manuscript for review and should be prepared to provide public access to such data, if practicable, and should in any event be prepared to retain such data for a reasonable time after publication.

Originality and plagiarism

Authors should ensure that they have written entirely original works, and if authors have used the work and/or words of others that this has been appropriately cited or quoted. Plagiarism in any form, including the touting of material contained in another paper (of the same authors or some other author) with cosmetic changes as a new paper; copying or paraphrasing substantial parts of another’s paper (without attribution), and claiming results from research conducted by others are among the numerous forms of plagiarism. In all its forms plagiarism constitutes unethical publishing behavior and is unacceptable.

Acknowledgement of sources

Proper acknowledgment of the work of others must always be given. Authors should cite publications that have been influential in determining the nature of the reported work.

Multiple, redundant or concurrent publication

An author should not in general publish manuscripts describing essentially the same research in more than one journal or primary publication. Submitting the same manuscript to more than one journal or conference concurrently constitutes unethical publishing behaviour and is unacceptable.

Authorship of the paper

Authorship should be limited to those who have made a significant contribution to the conception, design, execution, or interpretation of the reported study. All those who have made significant contributions should be listed as co-authors. Where there are others who have participated in certain substantive aspects of the manuscript, they should be acknowledged or listed as contributors.

The corresponding author should ensure that all appropriate co-authors and no inappropriate co-authors are included on the manuscript, and that all co-authors have seen and approved the final version of the manuscript and have agreed to its submission for publication.

Fundamental errors in published work

When an author discovers a significant error or inaccuracy in his/her own published work, it is the author’s obligation to promptly notify editors or publishers and cooperate with editors to retract or correct the paper.

If editors or the publisher learns from a third party that a published work contains a significant error, it is the obligation of the author to promptly retract or correct the paper or provide evidence to editors of the correctness of the original paper.

Disclosure and conflicts of interest

All authors should disclose in their manuscript any financial or other substantive conflict of interest that might be construed to influence the results or interpretation of their manuscript. All sources of financial support for the paper should be disclosed.

Publisher's confirmation

In cases of alleged or proven scientific misconduct, fraudulent publication or plagiarism the publisher, in close collaboration with the editors, will take all appropriate measures to clarify the situation and to amend the paper in question. This includes the prompt publication of an erratum or, in the most severe cases, the complete retraction of the affected work.

Author Guidlines

Microsoft Word Template for the Central European Conference on Information and Intelligent Systems (CECIIS)

Author 1, Author 2
Author's Common Affiliation
Department/Institute
Full Address
{author1, author2}@domain.com

Author(s) Name(s)
Author Affiliation(s)
Department/Institute
Full Address(es)
E-mail(s)

Abstract. *The abstract is to be in fully-justified italicized text, at the top of the left-hand column as presented here, below the author information. Use the word "Abstract." in 10-point Times, boldface type, left positioned, initially capitalized, followed by the abstract in 10-point, single-spaced type, up to 100 words long.*
Leave one blank line after the abstract, and then begin the keywords. Use the word "Keywords." in 10-point Times, boldface type, left positioned, initially capitalized, followed by up to ten keywords in 10-point, separated by comma, as below.
Leave two blank lines after the keywords, and then begin the main text.

Keywords. CECIIS, conference paper, template

1 Introduction

This is an example of a Central European Conference on Information and Intelligent Systems (CECIIS) submission. These guidelines include complete description of the paper style including formatting, fonts, spacing, and related information for producing your proceedings manuscripts.

This template can be used to create your own paper but if you have any troubles or if in doubt do not hesitate to contact us via e-mail: ceciisoo@foi.hr or via the conference web site at http://www.ceciis.foi.hr.

2 Structure and Requirements for Research Paper (Examples)

Central European Conference on Information and Intelligent Systems (CECIIS) welcomes research papers on all topics of interest (Computer Games, Data and Knowledge Bases, Digital transformation and corporate culture, Innovation, entrepreneurship and economic growth, Intelligent Energy, Intelligent Systems, Learning analytics, AI and technologies in education, Strategic planning and decision making, Systems for Smart Mobility, Trends in Software Engineering) as well as on the Conference theme/special emphasis announced on yearly base, addressing both, research theory and practice.

The following paper structure examples regarding the research type could be used for preparing a CECIIS paper:

1. Quantitative questionnaire-based: background research (state of the art); research problem; hypothesis or research question; description of instrument used; sampling; statistical methods used for analysis; discussion of results with limitations of research.
2. Qualitative research: background research (state of the art); research problem; research question; qualitative method used (case study research, interviews, content analysis etc.); discussion of results with limitations of research.
3. Algorithms and theoretical research: theory description with references; original theoretical contribution (theorem, algorithm etc.) with proofs or confirmations; future research and conclusion
4. Systematic literature review with meta-analysis: scope of research, research questions or research purpose; sources of literature with period covered (data bases, journals etc.), methods of analysis, conclusions with future research

Research paper should be theoretically and evidence based, clearly and consistently as well as ethically prepared and written. Central European Conference on Information and Intelligent Systems (CECIIS) has zero-tolerance policy on any kind of plagiarism. Paper containing text copied and used from another author(s) work without permission or from authors previously published own work, without referencing, will be retracted.

3 Paper Organization and Formatting

All papers should be written in English, up to 8 pages long (approximately 20 000 characters), and arranged in the following order:

- Main title
- Author(s), affiliation(s), full and e-mail addresses
- Abstract
- Keywords
- Body text (Main text)
- Footnotes
- Acknowledgements
- References

All printed material, including text and figures, must be kept within a print area of 16 cm (6.3") wide by 24.7 (9.7") high. Do not write or print anything outside the print area. Paper size is A4: 21x29.7 cm (8.3x11.7"). Margins (top, bottom, left, right) are 2.5 cm (1"). All *text* must be in a two-column format. Columns are to be 7.7 cm (3") wide, with a 0.6 cm (0.24") space between them. Text must be fully justified. Indent each paragraph by 0.5 cm (0.2").

The final submission has to be submitted in a single PDF file with all fonts that are used embedded, and also the source file (Microsoft Word, LaTeX - zip original files and images).

4 Paper Submission in Croatian Language

Authors who also want to publish the translation of their paper in Croatian, need to upload both the English and Croatian version of the paper (in the same document) for review. English and Croatian versions need to be in parallel columns (side by side). Left column is the paper in English and right column is the paper in Croatian. A camera-ready full paper should be prepared according to the Instructions for Authors and submitted following all the stages of the Paper Submission Process.

5 Main Title

The main title (on the first page) should begin 3,5 cm from the top edge of the page, centred, and in Times 18-point, bold face. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave a blank line after the title.

6 Author Name(s) and Affiliation(s)

Author names and affiliations are to be centred beneath the title and printed in Times 10-point type. Author names should be in bold face. Multiple authors shall be grouped by affiliation as shown in the title above. Include also e-mail addresses in 9-point `Courier/Typewriter` face. Follow the author information by two blank lines before main text.

7 Type-style and Fonts

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times that you have access to. Please avoid using bit-mapped fonts if possible. True-Type 1 fonts are preferred.

Also make sure that all fonts in the final PDF file are embedded. Thus we strongly encourage the use of Latex especially if you use non-standard fonts for formulas for your examples. You can make sure that your fonts are embedded by opening the PDF file in some PDF viewing program and taking a look at the properties of the document.

8 First-order Headings (May Extend to the Next Line)

For example, “1 Introduction”, should be Times 14-point boldface, initially capitalized, flush left, with two blank lines before, and one blank line after.

Don't use periods (“.”) after the heading number, use them only in lower order headings to separate them from higher order headings. For *long* headings use *a hanging indent* aligning the text to the right of the heading number as shown above.

All paragraphs in the text should be indented except for the first paragraph in a section.

8.1 Second-order Headings (May Extend to the Next Line)

As in this heading, they should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one after. Use a hanging indent for long headings.

8.1.1 Third-order Headings (May Extend to the Next Line)

Third-order headings, as in this paragraph, are discouraged. However, if you must use them, use 10-point Times, boldface, initially capitalized, flush left, with one blank line before, and one after. Use a hanging indent for long headings.

9 Main Text

Type your main text in 10-point Times, single-spaced. Do not use double-spacing. All paragraphs should be indented 0.5 cm (0.2") except for the first paragraph in a section.

Be sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

9.1 Figures

All figures are to be included within the text. **Figure captions** are to be *below* the figures, in 10-point Times (or a similar serif font), normal face. Initially capitalize only the first word of each figure caption.

Figures are to be numbered consecutively with Arabic numerals throughout the paper, for example: "Figure 1. Database contexts", and are referred to in the text as Fig. 1, Fig. 2, etc.

9.1.1 Illustrations, Photographs and Graphs

Illustrations, photographs and graphs are considered as figures. All graphics should be centred. Your artwork must be in place in the article (preferably printed as part of the text rather than pasted up). If you are using photographs and are able to have halftones made at a print shop, use a 100- or 110-line screen. Supply the best quality photographs and illustrations possible.

Pencilled lines and very fine lines do not reproduce well. Remember, the quality of the book cannot be better than the originals provided. The conference proceedings will be printed in grey scale, but you can provide images in colour since the papers will also be published on the web site of the conference.

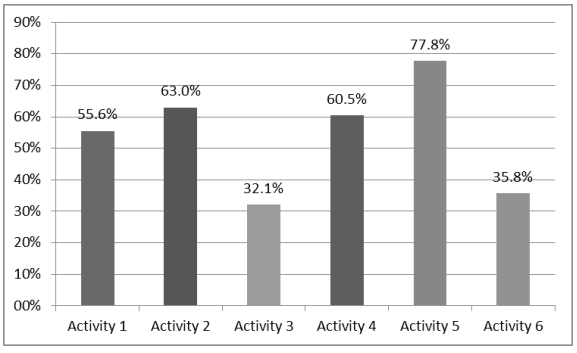


Figure 1. Times, 10 points, normal face

9.2 Tables

All tables are to be included within the text. **Table titles** are to be *above* the tables, in 10-point Times (or a similar serif font), normal face. Initially capitalize only the first word of each table title. Tables are to be numbered consecutively with Arabic numerals throughout the paper, for example: "Table 1. Input data", and are referred to in the text as Table 1, Table 2, etc. Column headings should be as brief as possible.

Table 1. Times, 10 points, normal face

	Column 1	Column 2	Column 3
Row 1	1	2	3
Row 2	4	5	6
Row 3	3	4	7
Row 4	2	6	4
Row 5	2	2	4

9.3 Equations

Equations should be numbered serially on the right-hand side by Arabic numerals in parentheses, and referred to in the text by eq. 1, eq. 2 etc.

$$3x^2 + 3x - 23 = 0 \quad (1)$$

9.4 Footnotes

Use footnotes sparingly (or not at all!) and place them at the bottom of the column on the page on which they are referenced.[1]

10 In-text Citations and Referencing

In-text citation must be used to denote all text used from another author(s) work or from authors previously published own work. When referenced in the text, enclose the author surname(s) and the publication year in brackets, for example (Surname, year) or (Surname, year, p. x) where x is the page number of the cited reference. When referenced paper has more than three authors, cite the paper as (Surname et al., year).

List and number all bibliographical references, alphabetically sorted, in 10-point Times, single-spaced, with a hanging indent, at the end of your paper. Use 6-point paragraph spacing after each reference. References should be distributed evenly in both columns as much as possible, so use a column break where appropriate.

In the References section below you will find examples of common types of bibliographical items: book (Mayer, 2009, p. 52), journal paper (Bule & Peer, 2014), journal paper with more than three authors (Shukor et al., 2014), conference proceedings paper (Steingartner & Novitzká, 2015), report or standard

[1] Use Times 8-point type, single-spaced. Avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

(*ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models*, 2011) and web page ("Web Content Accessibility Guidelines (WCAG) 2.0," 2008). For other types of bibliographical items refer the **APA style guidelines**. It should be ensured that every reference cited in the text is also listed in the References section (and vice versa).

Acknowledgments

Acknowledgments, if necessary, should appear in a separate paragraph preceding the references.

References

Bule, J., & Peer, P. (2014). Technical, Legal, Economic and Social Aspects of Biometrics for Cloud Computing. *Journal of Information and Organizational Sciences*, 38(2), 83–95.

ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. (2011).

Mayer, R. E. (2009). *Multimedia Learning* (Second Edi.). New York: Cambridge University Press.

Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). A Predictive Model to Evaluate Students' Cognitive Engagement in Online Learning. *Procedia - Social and Behavioral Sciences*, 116, 4844–4853. doi:10.1016/j.sbspro.2014.01.1036

Steingartner, W., & Novitzká, V. (2015). A new approach to operational semantics by categories. In T. Hunjak, V. Kirinić, & M. Konecki (Eds.), *Proceedings of the 26th Central European Conference on Information and Intelligent System (CECIIS 2015)* (pp. 247–254). University of Zagreb, Faculty of Organization and Informatics Varaždin.

Web Content Accessibility Guidelines (WCAG) 2.0. (2008). Retrieved from https://www.w3.org/TR/WCAG20/

Microsoft Word Template for papers in Croatian Language for the Central European Conference on Information and Intelligent Systems

Microsoft Word predložak za radove na hrvatskome jeziku za Central European Conference on Information and Intelligent Systems

Author 1, Author 2
Author's Common Affiliation
Naziv ustanove
Department/Institute
Naziv odjela/instituta
Full Address / Adresa
{author1, author2}@domain.com

Author(s) Name(s)
Author Affiliation(s)
Naziv ustanove
Department/Institute
Naziv odjela/instituta
Full Address / Adresa
E-mail(s)

Abstract*. The abstract is to be in fully-justified italicized text, at the top of the column as presented here, below the author information. Use the word "Abstract." in 10-point Times, boldface type, left positioned, initially capitalized, followed by the abstract in 10-point, single-spaced type, up to 100 words long.*
Leave one blank line after the abstract, and then begin the keywords. Use the word "Keywords." in 10-point Times, boldface type, left positioned, initially capitalized, followed by up to ten keywords in 10-point, separated by comma, as below.
Leave two blank lines after the keywords, and then begin the main text.

Keywords. CECIIS, conference paper

Sažetak*. Tekst sažetka treba biti obostrano poravnat i pisan kurzivom, na vrhu stupca kako je prikazano ovdje ispod podataka o autoru. Upotrijebite riječ "Sažetak." u fontu Times veličine 10, podebljano, poravnato ulijevo, velikim početnim slovom. Nakon riječi "Sažetak." slijedi sažetak rada u fontu veličine 10, jednostruki prored, dužine do 100 riječi.*
Nakon sažetka jedan redak ostaje prazan, a zatim slijede ključne riječi. Upotrijebite riječi "Ključne riječi." u fontu Times veličine 10, podebljano, lijevo poravnanje, s velikim početnim slovom, nakon čega slijedi do deset ključnih riječi u fontu veličine 10, odvojenih zarezom, kao u nastavku.
Ostavite dva prazna retka nakon ključnih riječi, a zatim započnite glavni tekst.

Ključne riječi. CECIIS, konferencijski rad

1 Introduction

This is an example of a Central European Conference on Information and Intelligent Systems (CECIIS) submission. These guidelines include complete description of the paper style including formatting, fonts, spacing, and related information for producing your proceedings manuscripts.

This template can be used to create your own paper but if you have any troubles or if in doubt do

1 Uvod

Ovo je predložak rada za Central European Conference on Information and Intelligent Systems (CECIIS). Smjernice uključuju cjelovit opis stila rada, uključujući oblikovanje, fontove, razmake i srodne informacije za objavu članaka u zborniku.

Predložak možete koristiti za izradu vlastitog rada, a ako imate bilo kakvih problema ili pitanja, ne

not hesitate to contact us via e-mail: ceciis@foi.unizg.hr.

2 Structure and Requirements for Research Paper (Examples)

Central European Conference on Information and Intelligent Systems (CECIIS) welcomes research papers on all topics of interest (Computer Games, Data and Knowledge Bases, Digital transformation and corporate culture, Innovation, entrepreneurship and economic growth, Intelligent Energy, Intelligent Systems, Learning analytics, AI and technologies in education, Strategic planning and decision making, Systems for Smart Mobility, Trends in Software Engineering) as well as on the Conference theme/special emphasis announced on yearly base, addressing both, research theory and practice.

The following paper structure examples could be used for preparing a CECIIS paper:

1. Quantitative questionnaire-based: background research (state of the art); research problem; hypothesis or research question; description of instrument used; sampling; statistical methods used for analysis; discussion of results with limitations of research.
2. Qualitative research: background research (state of the art); research problem; research question; qualitative method used (case study research, interviews, content analysis etc.); discussion of results with limitations of research.
3. Algorithms and theoretical research: theory description with references; original theoretical contribution (theorem, algorithm etc.) with proofs or confirmations; future research and conclusion
4. Systematic literature review with meta-analysis: scope of research, research questions or research purpose; sources of literature with period covered (data bases, journals etc.), methods of analysis, conclusions with future research

Research paper should be theoretically and evidence based, clearly and consistently as well as ethically prepared and written. Central European Conference on Information and Intelligent Systems (CECIIS) has zero-tolerance policy on any kind of plagiarism. Paper containing text copied and used from another author(s) work without permission or from authors previously published own work, without referencing, will be retracted.

3 Paper Organization and Formatting

ustručavajte se kontaktirati nas putem e-pošte: ceciis@foi.unizg.hr.

2 Struktura i zahtjevi za istraživački rad (primjeri)

Central European Conference on Information and Intelligent Systems (CECIIS) očekuje istraživačke radove na teme od interesa (Računalne igre, Baze podataka i baze znanja, Digitalne transformacije i korporativna kultura, Inovacije, poduzetništvo i ekonomski rast, Pametna energija, Inteligentni sustavi, Analitike učenja, umjetna inteligencija i tehnologije u obrazovanju, Strateško planiranje i donošenje odluka, Sustavi za pametnu mobilnost, Trendovi u softverskom inženjerstvu), kao i na godišnju temu konferencije, a bave se kako teorijskim, tako i primijenjivim istraživanjem.

Sljedeći primjeri strukture rada mogli bi se koristiti za pripremu rada za CECIIS:

1. Kvantitativni upitnik: pregled dosadašnjih istraživanja; problem istraživanja; hipoteza ili istraživačko pitanje; opis korištenog instrumenta; uzorkovanje; statističke metode korištene za analizu; rasprava o rezultatima s ograničenjima istraživanja.
2. Kvalitativno istraživanje: pregled dosadašnjih istraživanja; problem istraživanja; istraživačko pitanje; korištena kvalitativna metoda (istraživanje slučaja, intervjui, analiza sadržaja itd.); rasprava o rezultatima s ograničenjima istraživanja.
3. Algoritmi i teorijsko istraživanje: opis teorije s referencama; izvorni teorijski doprinos (teorem, algoritam, itd.) s dokazima ili potvrdama; buduće istraživanje i zaključak
4. Sustavni pregled literature s metaanalizom: opseg istraživanja, istraživačka pitanja ili svrha istraživanja; izvori literature s obuhvaćenim razdobljem (baze podataka, časopisi itd.), metode analize, zaključci o budućim istraživanjima

Istraživački rad trebao bi biti teorijski i zasnovan na dokazima, jasno i dosljedno te etički pripremljen i napisan. Central European Conference on Information and Intelligent Systems (CECIIS) ima politiku nulte tolerancije prema bilo kojoj vrsti plagijarizma. Rad koji sadrži tekst kopiran i korišten iz djela drugog autora bez odobrenja ili iz prethodno objavljenog vlastitog djela, bez referenciranja, bit će odbačen.

3 Organizacija i formatiranje rada

Svi radovi trebaju biti predani na engleskom jeziku, dužine do 8 stranica (oko 20 000 znakova) i organizirani prema sljedećem redoslijedu:

All papers should be written in English, up to 8 pages long (approximately 20 000 characters), and arranged in the following order:

- Main title
- Author(s), affiliation(s), full and e-mail addresses
- Abstract
- Keywords
- Body text (Main text)
- Footnotes
- Acknowledgements
- References

All printed material, including text and figures, must be kept within a print area of 16 cm (6.3") wide by 24.7 (9.7") high. Do not write or print anything outside the print area. Paper size is A4: 21x29.7 cm (8.3x11.7"). Margins (top, bottom, left, right) are 2.5 cm (1"). All *text* must be in a two-column format. Columns are to be 7.7 cm (3") wide, with a 0.6 cm (0.24") space between them. Text must be fully justified. Indent each paragraph by 0.5 cm (0.2").

The final submission has to be submitted in a single PDF file with all fonts that are used embedded, and also the source file (Microsoft Word, LaTeX).

4 Paper Submission in Croatian Language

Authors who also want to publish the translation of their paper in Croatian, need to upload both the English and Croatian version of the paper (in the same document) for review. English and Croatian versions need to be in parallel columns (side by side). Left column is the paper in English and right column is the paper in Croatian. A camera-ready full paper should be prepared according to the Instructions for Authors and submitted following all the stages of the Paper Submission Process.

5 Main Title

The main title should begin 3,5 cm from the top edge of the page, centred, and in Times 18-point, bold face. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave a blank line after the title.

6 Author Name(s) and Affiliation(s)

- Naslov rada
- Autor(i), pripadnost(i), adrese i e-mail adrese
- Sažetak
- Ključne riječi
- Tekst rada
- Fusnote
- Zahvale
- Reference

Sav materijal, uključujući tekst i slike, mora biti u području ispisa od 16 cm (6,3") širine i 24,7 (9,7") visine. Ne pišite i ne ispisujte ništa izvan područja ispisa. Veličina papira je A4: 21x29,7 cm (8,3x11,7"). Margine (gore, dolje, lijevo, desno) su 2,5 cm (1"). Sav tekst mora biti u formatu s dva stupca. Stupci trebaju biti široki 7,7 cm (3"), a između njih moraju biti razmaci od 0,6 cm (0,24"). Tekst mora biti obostrano poravnat. Uvucite svaki odlomak za 0,5 cm (0,2"), osim prvog odlomka u odjeljku.

Konačni rad mora se poslati u jednoj PDF datoteci s ugrađenim svim fontovima koji se koriste, kao i izvornoj datoteci (Microsoft Word, LaTeX).

4 Predaja rada na hrvatskome jeziku

Autori koji žele objaviti i prijevod svog rada na hrvatskome jeziku, trebaju predati i englesku i hrvatsku verziju rada (u istom dokumentu) na recenziju. Engleska i hrvatska verzija trebaju biti u paralelnim stupcima (jedan pored drugog). Lijevi stupac je rad na engleskom jeziku, a desni stupac je rad na hrvatskome jeziku. Finalnu verziju cjelovitog rada treba pripremiti u skladu s Uputama za autore i predati slijedeći sve faze procesa predaje radova.

5 Naslov rada

Naslov rada (na prvoj stranici) trebao bi započeti 3,5 cm od gornjeg ruba stranice, centrirano, u fontu Times veličine 18, podebljano. Naslov započinje velikim početnim slovom. Ostavite jedan prazan redak nakon naslova.

6 Imena autora i pripadnost(i)

Imena autora i pripadnost trebaju biti centrirani ispod naslova i pisani fontom Times veličine 10. Imena autora trebaju biti podebljana. Više autora bit će grupirano prema pripadnosti kao što je prikazano u ovom predlošku. Uključite i adrese e-pošte u fontu `Courier/Typewriter` veličine 9. Nakon podataka

Author names and affiliations are to be centred beneath the title and printed in Times 10-point type. Author names should be in bold. Multiple authors shall be grouped by affiliation as shown in the title above. Include also e-mail addresses in 9-point `Courier/Typewriter` face. Follow the author information by two blank lines before main text.

7 Type-style and Fonts

Wherever Times is specified, Times Roman or Times New Roman may be used. Please avoid using bit-mapped fonts if possible. True-Type 1 fonts are preferred.

Also make sure that all fonts in the final PDF file are embedded. We strongly encourage the use of LaTeX, especially if you use non-standard fonts for formulas for your examples.

8 First-order Headings (May Extend to the Next Line)

For example, “1 Introduction”, should be Times 14-point boldface, initially capitalized, flush left, with two blank lines before, and one blank line after.

Don't use periods (“.”) after the heading number, use them only in lower order headings to separate them from higher order headings. For *long* headings use *a hanging indent* aligning the text to the right of the heading number as shown above.

All paragraphs in the text should be indented except for the first paragraph in a section.

8.1 Second-order Headings (May Extend to the Next Line)

As in this heading, they should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one after. Use a hanging indent for long headings.

8.1.1 Third-order Headings (May Extend to the Next Line)

Third-order headings, as in this paragraph, are discouraged. However, if you must use them, use 10-point Times, boldface, initially capitalized, flush left, with one blank line before, and one after. Use a hanging indent for long headings.

9 Main Text

Type your main text in 10-point Times, single-spaced. Do not use double-spacing. All paragraphs should be indented 0.5 cm (0.2") except for the first paragraph in a section.

o autoru slijede dva prazna retka prije nego počne glavni tekst.

7 Stil i fontovi

Gdje god je naveden font Times, može se koristiti Times Roman ili Times New Roman. Molimo izbjegavajte upotrebu bitmapiranih fontova ako je moguće. Poželjni su fontovi True-Type 1.

Također provjerite jesu li ugrađeni svi fontovi u konačnu PDF datoteku. Pritom, snažno preporučamo upotrebu LaTeX-a, posebno ako za svoje primjere koristite nestandardne fontove za formule.

8 Naslovi prvog reda (mogu se nastaviti u sljedeći redak)

Na primjer, "1 Uvod" treba biti font Times veličine 14, podebljano, s velikim početnim slovom, poravnato ulijevo, s dva prazna retka prije i jednim praznim redom nakon naslova.
Ne upotrebljavajte točke (".") iza broja naslova, koristite ih samo u naslovima nižeg reda kako biste ih odvojili od naslova višeg reda. Za duge naslove upotrijebite “hanging indent” poravnavajući tekst udesno od broja naslova, kao što je prikazano gore.

Svi odlomci u tekstu trebaju biti uvučeni, osim prvog odlomka u odjeljku.

8.1 Naslovi drugog reda (mogu se nastaviti u sljedeći redak)

Kao i u ovom naslovu, trebali bi biti font Times veličine 12, s velikim početnim slovom, poravnati lijevo, s jednim praznim retkom prije i jednim nakon. Koristite “hanging indent” za duge naslove.

8.1.1 Naslovi trećeg reda (mogu se nastaviti u sljedeći redak)

Naslovi trećeg reda, kao u ovom paragrafu, ne preporučuju se. Međutim, ako ih morate koristiti, upotrijebite font Times veličine 10, podebljano, s velikim početnim slovom, poravnati ulijevo, s jednim praznim retkom prije i jednim nakon. Koristite “hanging indent” za duge naslove.

9 Tekst rada

Tekst rada pišite fontom Times veličine 10, s jednostrukim proredom. Ne koristite dvostruki prored. Svi odlomci trebaju biti uvučeni 0,5 cm (0,2"), osim prvog odlomka u odjeljku.

Be sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

9.1 Figures

All figures are to be included within the text. **Figure captions** are to be *below* the figures, in 10-point Times (or a similar serif font), normal face. Initially capitalize only the first word of each figure caption.

Figures are to be numbered consecutively with Arabic numerals throughout the paper, for example: “Figure 1. Database contexts”, and are referred to in the text as Fig. 1, Fig. 2, etc.

9.1.1 Illustrations, Photographs and Graphs

Illustrations, photographs and graphs are considered as figures. All graphics should be centred. Your artwork must be in place in the article (preferably printed as part of the text rather than pasted up). Supply the best quality photographs and illustrations possible.

Very fine lines do not reproduce well. The conference proceedings will be printed in grey scale, but you can provide images in colour since the papers will also be published on the conference web site.

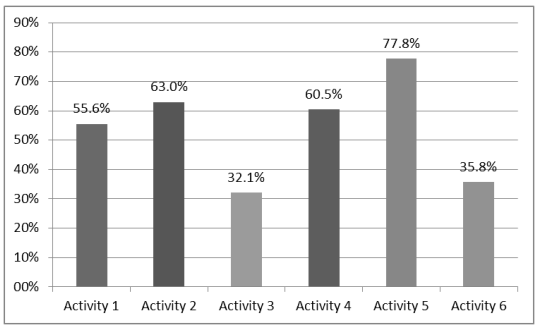


Figure 1. Times, 10 points, normal face

9.2 Tables

All tables are to be included within the text. Table titles are to be *above* the tables, in 10-point Times, normal face. Initially capitalize only the first word of each table title. Tables are to be numbered consecutively with Arabic numerals throughout the paper, for example: “Table 1. Input data”, and is referred to in the text as Table 1.

Table 1. Times, 10 points, normal face

	Column 1	Column 2	Column 3
Row 1	1	2	3
Row 2	4	5	6
Row 3	3	4	7
Row 4	2	6	4
Row 5	2	2	4

Provjerite je li vaš tekst poravnat obostrano - to jest, i lijevo i desno. Molimo ne stavljajte dodatne prazne retke između odlomaka.

9.1 Slike

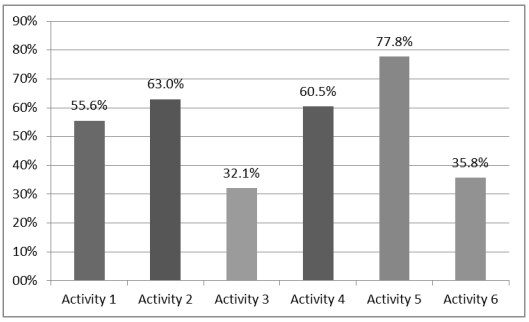
Sve slike trebaju biti uključene u tekst. Nazivi slika trebaju biti ispod slika, u fontu Times veličine 10 (ili sličnom fontu). Samo prvu riječ svakog naziva slike pišite velikim početnim slovom.

Slike se numerira uzastopno arapskim brojevima u cijelom radu, na primjer: "Slika 1. Konteksti baze podataka", a u tekstu se navodi kao Slika 1.

9.1.1 Ilustracije, fotografije i grafovi

Ilustracije, fotografije i grafikoni smatraju se slikama. Sva grafika treba biti centrirana. Svaka slika ima svoje mjesto u članku. Osigurajte najbolju moguću kvalitetu fotografija i ilustracija.

Vrlo tanke linije ne reproduciraju se dobro. Zbornik radova s konferencije bit će tiskan u nijansama sive boje, ali možete osigurati slike u boji jer će radovi biti objavljeni i na mrežnoj stranici konferencije.



Slika 1. Times, veličina 10

9.2 Tablice

Sve tablice trebaju biti uključene u tekst. Naslovi tablica trebaju biti iznad tablica, pisane fontom Times veličine 10. Velikim početnim slovom pišite samo prvu riječ svakog naslova tablice. Tablice se u radu redom numeriraju arapskim brojevima, na primjer: "Tablica 1. Ulazni podaci", a u tekstu se naziva Tablica 1.

Tablica 1. Times, 10 points, normal face

	Column 1	Column 2	Column 3
Row 1	1	2	3
Row 2	4	5	6
Row 3	3	4	7
Row 4	2	6	4
Row 5	2	2	4

9.3 Equations

Equations should be numbered serially on the right-hand side by Arabic numerals in parentheses, and referred to in the text by eq. 1, eq. 2 etc.

$$3x^2 + 3x - 23 = 0 \quad (1)$$

9.4 Footnotes

Use footnotes sparingly (or not at all!) and place them at the bottom of the column on the page on which they are referenced.[1]

10 In-text Citations and Referencing

In-text citation must be used to denote all text used from another author(s) work or from authors previously published own work. When referenced in the text, enclose the author surname(s) and the publication year in brackets, for example (Surname, year) or (Surname, year, p. x) where x is the page number of the cited reference. When referenced paper has more than three authors, cite the paper as (Surname et al., year).

List and number all bibliographical references, alphabetically sorted, in 10-point Times, single-spaced, with a hanging indent, at the end of your paper. Use 6-point paragraph spacing after each reference. References should be distributed evenly in both columns as much as possible, so use a column break where appropriate.

In the References section below you will find examples of common types of bibliographical items: book (Mayer, 2009, p. 52), journal paper (Bule & Peer, 2014), journal paper with more than three authors (Shukor et al., 2014), conference proceedings paper (Steingartner & Novitzká, 2015), report or standard (*ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models*, 2011) and web page ("Web Content Accessibility Guidelines (WCAG) 2.0," 2008). For other types of bibliographical items refer the **APA style guidelines**. It should be ensured that every reference cited in the text is also listed in the References section (and vice versa).

Acknowledgments

9.3 Jednadžbe

Jednadžbe se numeriraju slijedno s desne strane arapskim brojevima u zagradama, a u tekstu se na njih poziva tekstom Jednadžba 1, Jednadžba 2 itd.

$$3x^2 + 3x - 23 = 0 \quad (1)$$

9.4 Fusnote

Koristite fusnote štedljivo (ili ih uopće ne koristite!) i stavite ih na dno stupca na stranici na kojoj su navedeni.[2]

10 Navodi u tekstu i reference

Citiranje u tekstu mora se koristiti za označavanje cijelog teksta korištenog iz djela drugog autora ili prethodno objavljenih vlastitih djela. Kad se citira u tekstu, prezimena autora i godinu izdanja priložite u zagrade, na primjer (Prezime, godina) ili (Prezime, godina, str. X) gdje je X broj stranice citirane reference. Kada referentni rad ima više od tri autora, navedite ga kao (Prezime i sur., Godina).

Navedite sve bibliografske reference, poredane po abecedi, u fontu Times veličine 10, s jednostrukim proredom, s "hanging indent", na kraju vašeg rada. Upotrijebite razmak paragrafa veličine 6 nakon svake reference. Reference bi trebale biti raspoređene u oba stupca što ravnomjernije, stoga upotrijebite prijelom stupca tamo gdje je to prikladno.

U odjeljku Reference u nastavku naći ćete primjere uobičajenih vrsta bibliografskih jedinica: knjiga (Mayer, 2009., str. 52.), rad u časopisu (Bule & Peer, 2014.), članak s više od tri autora (Shukor i sur., 2014), članak s konferencije (Steingartner i Novitzká, 2015), izvještaj ili standard (*ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models, 2011*) i mrežna stranica ("Web Content Accessibility Guidelines (WCAG) 2.0," 2008). Za ostale vrste bibliografskih jedinica pogledajte **smjernice za stil APA**. Potrebno je osigurati da svaka referenca navedena u tekstu bude navedena i u odjeljku Reference (i obrnuto).

Zahvale

Zahvale, ako ih ima trebaju biti u zasebnom odlomku koji prethodi referencama.

[1] Use Times 8-point type, single-spaced. Avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

[2] Koristite font Times veličine 8, s jednostrukim proredom. Izbjegavajte potpuno koristiti fusnote i u tekst uvrstite potrebna periferna zapažanja (u zagrade, ako želite, kao u ovoj rečenici).

Acknowledgments, if necessary, should appear in a separate paragraph preceding the references.

References

Bule, J., & Peer, P. (2014). Technical, Legal, Economic and Social Aspects of Biometrics for Cloud Computing. *Journal of Information and Organizational Sciences*, 38(2), 83–95.

ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. (2011).

Mayer, R. E. (2009). *Multimedia Learning* (Second Edi.). New York: Cambridge University Press.

Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). A Predictive Model to Evaluate Students' Cognitive Engagement in Online Learning. *Procedia - Social and Behavioral Sciences*, 116, 4844–4853. doi:10.1016/j.sbspro.2014.01.1036

Steingartner, W., & Novitzká, V. (2015). A new approach to operational semantics by categories. In T. Hunjak, V. Kirinić, & M. Konecki (Eds.), *Proceedings of the 26th Central European Conference on Information and Intelligent System (CECIIS 2015)* (pp. 247–254). University of Zagreb, Faculty of Organization and Informatics Varaždin.

Web Content Accessibility Guidelines (WCAG) 2.0. (2008). Retrieved from https://www.w3.org/TR/WCAG20/

Reference

Bule, J., & Peer, P. (2014). Technical, Legal, Economic and Social Aspects of Biometrics for Cloud Computing. *Journal of Information and Organizational Sciences*, 38(2), 83–95.

ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. (2011).

Mayer, R. E. (2009). *Multimedia Learning* (Second Edi.). New York: Cambridge University Press.

Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). A Predictive Model to Evaluate Students' Cognitive Engagement in Online Learning. *Procedia - Social and Behavioral Sciences*, 116, 4844–4853. doi:10.1016/j.sbspro.2014.01.1036

Steingartner, W., & Novitzká, V. (2015). A new approach to operational semantics by categories. In T. Hunjak, V. Kirinić, & M. Konecki (Eds.), *Proceedings of the 26th Central European Conference on Information and Intelligent System (CECIIS 2015)* (pp. 247–254). University of Zagreb, Faculty of Organization and Informatics Varaždin.

Web Content Accessibility Guidelines (WCAG) 2.0. (2008). Retrieved from https://www.w3.org/TR/WCAG20/

CECIIS 2025 Sponsors

CECIIS 2025 GOLD SPONSOR

INFODOM

CECIIS 2025 GOLD SPONSOR



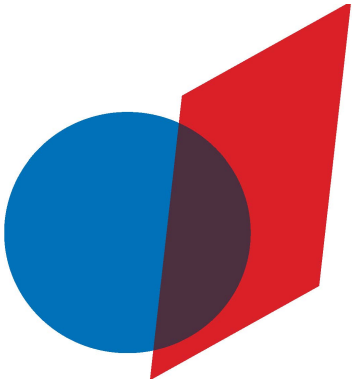
CECIIS 2025 GOLD SPONSOR



CECIIS 2025 SILVER SPONSORS

HBOR

HRVATSKA BANKA
ZA OBNOVU I RAZVITAK



JANAF

CECIIS 2025 SILVER SPONSORS



CECIIS 2025 BRONZE SPONSORS

VARAŽDINSKA ŽUPANIJA



Varaždin County

VARAŽDIN

City of Varaždin

SICK

Sensor Intelligence.

CECIIS 2025 Program Committee

Nina Begičević Ređep, Program Committee Chair, University of Zagreb, Croatia

Victoria Appatova, University of Cincinnati, USA
Igor Balaban, University of Zagreb, Croatia
Marko Bohanec, Jozef Stefan Institute, Slovenia
Leo Budin, University of Zagreb, Croatia
Aruna Chandra, Indiana State University, USA
Violeta Cvetkoska, University in Skopje, Macedonia
Luis de Marcos Ortega, University of Alcalá, Spain
Blaženka Divjak, University of Zagreb, Croatia
Jasminka Dobša, University of Zagreb, Croatia
Juraj Dud'ak, Slovak University of Technology in Bratislava, Slovakia
Matjaž Garms, Jozef Stefan Institute, Slovenia
Gabriel Gaspar, Slovak University of Technology in Bratislava, Slovakia
Gordan Gledec, University of Zagreb, Croatia
Andrina Granić, University of Split, Croatia
Marjan Heričko, University of Maribor, Slovenia
Dieter Hertweck, Herman Hollerith Research Centre, Germany
Marina Klačmer Čalopa, University of Zagreb, Croatia
Michal Kohani, University of Žilina, Slovakia
Mario Konecki, University of Zagreb, Croatia
Vladimir Kovšca, University of Zagreb, Croatia
Marjan Krašna, University of Maribor, Slovenia
Lode Lauwaert, KU Leuven, Belgium
Sandra Lovrenčić, University of Zagreb, Croatia
Ivan Luković, University of Belgrade, Serbia
Ivan Magdalenić, University of Zagreb, Croatia
Marcelo Fabian Maina, Open University of Catalonia, Spain
Aleksandar Marković, University of Belgrade, Serbia
Peter Marton, University of Žilina, Slovakia
Marta Mas Machuca, UIC Barcelona, Spain
Dunja Mladenič, Jožef Stefan Institute, Slovenia
Oliver Moravčik, Slovak University of Technology, Slovakia
Enrique Mu, Carlow University, USA
Karlo Nyarko, University of Osijek, Croatia
Jan Paralič, Technical University of Košice, Slovakia
Elisabeth Pergler, CAMPUS 02 - University of Applied Sciences, Austria
Krzysztof Podlaski, University of Lodz, Poland
Patrizia Poščić, University of Rijeka, Croatia
Sonja Ristić, University of Novi Sad, Serbia
Markus Schatten, University of Zagreb, Croatia
Zlatko Stapić, University of Zagreb, Croatia
William Steingartner, Technical University of Kosice, Slovakia
Vjeran Strahonja, University of Zagreb, Croatia
Marko Tadić, University of Zagreb, Croatia
Martina Tomičić Furjan, University of Zagreb, Croatia
Katarina Tomičić Pupek, University of Zagreb, Croatia
Jan Tretmans, Radboud University Nijmegen, Netherlands
Wim Van Petegem, KU Leuven, Belgium
Violeta Vidaček Hainš, University of Zagreb, Croatia
Sirje Virkus, Tallinn University, Estonia
Neven Vrček, University of Zagreb, Croatia
Mladen Vouk, North Carolina State University, USA
Diana Šimić, University of Zagreb, Croatia

Honorary Members - Former Program Committee Chairs

Tihomir Hunjak, University of Zagreb, Croatia
Mirko Maleković, University of Zagreb, Croatia
Boris Aurer, University of Zagreb, Croatia
Prof. Wolf Rauch, University of Graz, Austria
Prof. Günter Müller, University of Freiburg, Germany

CECIIS 2025 Organizing Committee

Marijana Bubanić, Organizing Committee Chair, University of Zagreb, Croatia

Dina Korent, University of Zagreb, Croatia
Marko Matus, University of Zagreb, Croatia
Lea Masnec, University of Zagreb, Croatia
Antonela Devčić, University of Zagreb, Croatia
Jelena Gusić Munđar, University of Zagreb, Croatia
Bernarda Kos, University of Zagreb, Croatia
Karlo Nadoveza, University of Zagreb, Croatia
Martina Đuras Sekovanić, University of Zagreb, Croatia
Karlo Pepelnik, University of Zagreb, Croatia
Marija Kuštelega, University of Zagreb, Croatia

CECIIS 2025 Research Tracks' Chairs

Computer Games: Markus Schatten, University of Zagreb, Croatia; Jaume Jordán Prunera, University Politècnica de València, Spain; Bogdan Okreša Đurić, University of Zagreb, Croatia

Cyber-Physical Systems and Intelligent Energy: Ivan Magdalenić, University of Zagreb, Croatia; Marinko Barukčić, University of Osijek, Croatia; Danijel Topić, University of Osijek, Croatia

Data and Knowledge Bases: Sandra Lovrenčić, University of Zagreb, Croatia; Patrizia Poščić, University of Rijeka, Croatia

Designing Innovation through Productive Disruption: Neven Vrček, University of Zagreb, Croatia; Katarina Tomičić-Pupek, University of Zagreb, Croatia

Digital Transformation and Corporate Culture: Katarina Tomičić-Pupek, University of Zagreb, Croatia; Gabriel Gaspar, Slovak University of Technology in Bratislava, Slovakia; Peter Marton, University of Žilina, Slovakia

Innovation, Entrepreneurship and Economic Growth: Vladimir Kovšca**,** University of Zagreb, Croatia; Martina Tomičić Furjan, University of Zagreb, Croatia

Intelligent Systems: Jasminka Dobša, University of Zagreb, Croatia; Dunja Mladenić, Jožef Stefan Institute, Slovenia

Learning Analytics, AI and Technologies in Education: Blaženka Divjak, University of Zagreb, Croatia; Igor Balaban, University of Zagreb, Croatia; Marjan Krašna, University of Maribor, Slovenia

Strategic Planning and Decision Making: Nina Begičević Ređep, University of Zagreb, Croatia; Nikola Kadoić, University of Zagreb, Croatia

Systems for Smart Mobility: Ivan Magdalenić, University of Zagreb, Croatia; Eduard Ivanjko, University of Zagreb, Croatia

Trends in Software Engineering: Marko Mijač, University of Zagreb, Croatia; William Steingartner, Technical University of Kosice, Slovakia; Luis de Marcos Ortega, University of Alcalá, Spain

CECIIS 2025 Poster Section Chairs

Nina Begičević Ređep, University of Zagreb, Croatia

Dina Korent, University of Zagreb, Croatia

Marijana Bubanić, University of Zagreb, Croatia

CECIIS 2025 Student Poster and Presentation Section Chairs

Violeta Vidaček Hainš, University of Zagreb, Croatia

Victoria Appatova, University of Cincinnati, USA



ISSN 1847-2001 (Print)
ISSN 1848-2295 (Online)