14th Higher Education Institutions Conference

03 – 04 September, 2026

THE FUTURE OF HIGHER EDUCATION:
THE ROLE OF ARTIFICIAL INTELLIGENCE IN LEARNING, TEACHING AND RESEARCH

PROCEEDINGS

Double-Blind Peer Reviewed

Edited by: Karmela Aleksić-Maslać and Mateja Kovačić



ZSEM Zagreb School of Economics and Management

EUonAIR European University Alliance

LSB Luxembourg School of Business

14th Higher Education Institutions Conference

03 – 04 September, 2026

THE FUTURE OF HIGHER EDUCATION:
THE ROLE OF ARTIFICIAL INTELLIGENCE IN LEARNING, TEACHING AND RESEARCH

PROCEEDINGS

Double Blind Peer Reviewed

Copyright © 2026 Mate Ltd, Zagreb, Croatia

All rights reserved. No part of this book may be reproduced in any form by any means without permission in writing from the publisher.

Publisher	MATE Ltd., Zagreb
For Publisher	Vesna Njavro
Chief Editor	PhD Mato Njavro
Editors	Karmela Aleksić-Maslać and Mateja Kovačić

Contents

Welcome Note V
Organizing Committee VI
Program Committee VII
Keynote speakers 2026 VIII
Panelists 2026 4

AI, HIGHER EDUCATION AND INSTITUTIONAL TRANSFORMATION 18

Directing Autonomous AgentsNew core competency for the MBA graduate 19
Mikael Alemu Gorsky and Ilya Levin

Integrating AI Avatars into Higher Ed: Balancing Technology and Academic Quality 32
Andrzej Rudko

Accreditations as Drivers of Organizational Learning and Strategic Development in Higher Education 33
Edda Apfenthaler, Maja Martinović and Zoran Barac

The Impact of Artificial Intelligence on the Economic Professions 50
But Dedaj

Beyond Screen Time: Smartphone Salience as a Cognitive Predictor of Problematic Smartphone Use 51
Matej Nakić, Ivan David Dogan and Goran Oblaković

AI, FUTURE SKILLS AND BUSINESS EDUCATION 62

What Knowledge-IntensiveEntrepreneurs Still Need to Learn: Skill Gaps in the Age of AI 63
Vitaliy Roud

Artificial Intelligence in Policing Technologies, Ethical Challenges, and Implications for Democratic Governance 74
Krunoslav Antoliš

From Knowing to Doing: Redesigning Business Education for AI 90
Monika Sońta, Lyubomyr Matsekh-Ukrayinskyy and Maciej Mąkosa

From Entrepreneurial Practice to Entrepreneurship Education: A Situated Experiential Framework for Critical Digital Literacy Development 105
Thao Laavunen

AI, LEARNING AND INNOVATION IN HIGHER EDUCATION 119

Cognitive Offloading in the Classroom: Evidence from Written and Oral Statistics Exams Before and After ChatGPT 120
Marija Ivanković

AI-Driven Learning Analytics Capabilities and Learning Data Quality for Intelligent Academic Decision-Making in Jordanian Higher Education: A Pilot Study 135

Maha Alkhaffaf

Implementation of the GAMBL-program – A Psychoeducational Intervention with a Gender Perspective for Gambling Behavior Prevention in Higher Education 149

Marina Ledesma-Casas, Loreto Camarero-Grados and Yolanda López-del-Hoyo

Exploring the Role of Artificial Intelligence in Accounting, Audit, and Tax Education: Insights from Croatian Academia and Practice 157

Branka Remenarić, Ivan Čevizović and Tonka Čevizović

AI-ENHANCED LEARNING, TEACHING AND PROFESSIONAL PRACTICE 172

Teaching Digital Transformation Through AI-Simulated Organisations: A Design Blueprint for Interactive Teaching Cases 173

Sven Dittes, Isabell Steidel, Ralf Dillerup and Brian Boyd

The Use of Artificial Intelligence in Business Communication within Croatian Corporate Settings 189

Zdravka Biočina and Sunčica Žnidar

Orchestrating Generative AI in an Immersive Learning Environment: A Participatory Design Study in Higher Education 200

Riem Yasin and Carsten Lanquillon

From Task Execution to Task Stewardship: AI-Induced Transformation of Task Profiles in Architecture and Engineering and Its Consequences for Higher Education 214

Enes Rischert, Viktor Grinewitschus, Niels Bartels, Uta Pottgiesser and Felix Meckmann

AI, DIGITAL TRANSFORMATION AND EMERGING CHALLENGES 226

Prompting as a Professional Competence: Analysing Pre-service Teachers' Prompts and Critical Appraisal of AI-Generated Lesson Plans 227

Maria Rosaria Re, Stefania Capogna and Alessandra Natalini

From Usage Intensity to Relationship Quality: A Typology of Student-AI Use and Design Propositions for Generative AI in Higher Education 242

Riem Yasin, Samir Yasin and Carsten Lanquillon

Standards for Machine-Readable Course Descriptions 255

Yannik Langer, Carsten Lanquillon, Seline Löwe, Tom Ochs, Elias Ottenbacher and Joshua Zürn

Input-Output Shock Propagation and Spectral Amplification 270

Matej Ivaniček and Yaraslau Svintsitski

Geopolitical Risk, Oil Supply Shocks, and the Macroeconomic Transmission of the 2026 Iran-Israel-U.S. War 282

Josip Tomašković

Index 299



Welcome Note

Dear guests and friends,

It is with great pleasure that I welcome you to the 14th Higher Education Institutions Conference (HEIC 2026), hosted by the Zagreb School of Economics and Management (ZSEM). Over the years, HEIC has grown into an international forum bringing together academics, higher education leaders, researchers, and practitioners to exchange ideas and reflect on the challenges and opportunities shaping the future of higher education.

This year's conference theme, "The Future of Higher Education: The Role of AI in Learning, Teaching and Research," reflects a transformation that is already profoundly changing universities. Artificial intelligence is influencing not only how students learn and faculty teach, but also how research is conducted, how institutions make decisions, and how we understand the roles and responsibilities of universities in society.

The HEIC 2026 programme approaches these developments from several complementary perspectives. Our keynote lectures, panel discussions, workshops, and research presentations explore questions of leadership and institutional purpose, the ethical and responsible use of AI in education, emerging models of human–AI collaboration in research, new competencies required by students and professionals, and innovative approaches to learning and teaching.

The abstracts collected in this volume reflect the international and interdisciplinary character of these discussions. They bring together research and experiences from different countries, institutions, and academic fields, illustrating both the remarkable opportunities created by artificial intelligence and the questions that still require careful research, critical reflection, and open dialogue.

HEIC has always been more than a conference programme. Its value comes from the exchange of perspectives, the conversations that continue beyond individual sessions, and the connections established among colleagues from different institutions and countries. I hope that HEIC 2026 will once again provide an inspiring environment for such exchange and create opportunities for new ideas and future collaboration.

On behalf of the Zagreb School of Economics and Management, I would like to thank our keynote speakers, panelists, authors, reviewers, partners, and all participants whose contributions have made HEIC 2026 possible.

I wish you an inspiring and enjoyable conference and a memorable stay in Dubrovnik.

Best regards,

Mato Njavro, PhD
Dean, Zagreb School of Economics and Management

Organizing Committee

dr. sc. Mato Njavro
Conference chair, Dean, Zagreb School of Economics and Management

mr. sc. Karmela Aleksić-Maslać
Associate Dean, Zagreb School of Economics and Management

dr. sc. Zoran Barac
Managing Director, Zagreb School of Economics and Management

Prof. dr. Patrick Vanhoudt
Dean, Luxembourg School of Business

dr. sc. Maja Martinović
Associate Dean, Graduate MBA Program, Zagreb School of Economics and Management

dr. sc. Goran Oblaković
Associate Dean, Zagreb School of Economics and Management

dr. sc. Mirna Koričan Lajtman
Academic Director,
ZSEM – Business Academy

Ante Kaleb
HEIC Event Coordinator,
ZSEM – Business Academy

Program Committee

dr. sc. Abdelkrim Haqiq
Hassan 1st University, Settat, Morocco

dr. sc. Alimshan Faizulayev
KIMEP University, Almaty, Kazakhstan

dr. sc. Boris Podobnik
Boston University and Zagreb School of Economics and Management

dr. sc. Patrick Vanhoudt
Dean, Luxembourg School of Business

dr. sc. Đuro Njavro
Zagreb School of Economics and Management

dr. sc. Emmanuel Glakpe
Howard University, Washington DC, USA

dr. sc. Ferdinand Daňo
Rector of the University of Economics in Bratislava, Slovakia

dr. sc. James Uhomoibhi
Ulster University, Northern Ireland, United Kingdom

dr. sc. Julie Felker
University of Michigan-Dearborn, USA

mr. sc. Karmela Aleksić Maslać
Zagreb School of Economics and Management

dr. sc. Maja Martinović
Zagreb School of Economics and Management

dr. sc. Mato Njavro
Zagreb School of Economics and Management

dr. sc. Metka Tekavčić
University of Ljubljana, Slovenia

dr. sc. Ralf Dillerup
Heilbronn University, Germany

dr. sc. Shab Hundal
Jyväskylä University of Applied Sciences, Finland

dr. sc. Siniša Krajnović
CEO and President of the Management Board at ENT – Ericsson Nikola Tesla

KEYNOTE SPEAKERS / 2026



Peter van der Hijden

Peter van der Hijden is an independent higher education strategy adviser and policy influencer based in Brussels. He worked for the European Commission for 23 years in the fields of education, transport, research and innovation. At the Directorate-General for Education and Culture, he served as Head of Sector for Higher Education Policy. At the Directorate-General for Research and Innovation, he contributed to the Human Resources Strategy for the European Research Area.

Peter played a central role in the development of the Erasmus programme, the Bologna Process and the EU Modernisation Agenda for universities. He studied law at Maastricht University in the Netherlands, where he began his career as Vice-Registrar of the University Council and Policy Officer for Education and Research.

In 2025, University World News and six other journals published his article on the future of Erasmus+, entitled “Could Europe Achieve Higher Impact with a Lighter Touch?”



Johan Roos, PhD

Johan Roos is an internationally recognized strategist, author, educator, and advisor specializing in leadership, strategy, and the impact of artificial intelligence on organizations.

A co-inventor of LEGO® Serious Play®, he has held senior leadership positions at leading international business schools, including Hult International Business School, Copenhagen Business School, and the Stockholm School of Economics. Today, he advises executives and boards around the world on navigating AI-driven transformation.

His latest book, Human Magic: Leading with Wisdom in an Age of Algorithms, explores how leaders can combine human judgment, creativity, and wisdom with the power of AI to build resilient, future-ready organizations.



Chief Thought Leadership Officer & EMEA, AACSB International

Prof. Eileen McAuliffe, PhD

Eileen McAuliffe serves as Executive Vice President, Chief Thought Leadership Officer, and Managing Director, EMEA at AACSB. As a key member of the executive team, she leads global thought leadership initiatives, positioning AACSB as the foremost voice in business education. In her role, she champions the quality, relevance, and societal impact of business schools worldwide, while also overseeing strategic direction and operations across the EMEA region to ensure strong engagement and support for AACSB members.

Prior to joining AACSB, McAuliffe was Pro Vice-Chancellor (Global) and Executive Dean of the Faculty of Business, Law, and Social Sciences at Birmingham City University. There, she played a central role in advancing the university's 2030 and Beyond Strategy, aligning international initiatives with institutional goals. As Executive Dean, she led a faculty of 400 and a student body of 12,000, managing an annual budget of approximately £100 million.

McAuliffe also brings a wealth of industry experience, having held executive tax advisory roles at ConocoPhillips in the oil sector. She earned her PhD in 2017 and was appointed Professor of International Taxation in 2020. In addition to her academic and professional leadership, she actively contributes to global policy as a member of the United Nations Platform for the Collaboration on Tax, and is a former member of the AACSB Board of Directors.

PANELISTS / 2026

PANEL 1

Strategic Futures of Higher Education in the Age of Artificial Inteligence

Moderator: Mato Njavro, PhD (ZSEM)



Zagreb School of Economics and Management

https://zsem.hr/en/dr-sc-mato-njavro/

Dr. Mato Njavro serves as the Dean of Zagreb School of Economics and Management (ZSEM) in Croatia. He also teaches at the Luxembourg School of Business and the University of St. Gallen in Switzerland. With a diverse academic and professional background, Dr. Njavro has worked at the St. Gallen Institute of Management in Asia, in Singapore, where he served as a senior research fellow. He was also a lecturer at the Singapore Management University, where he taught a course on China's economic development.

Prior to joining the University of St. Gallen and Singapore Management University, Dr. Njavro was a visiting research fellow at Harvard University's Institute for Quantitative Social Sciences (IQSS). Dr. Njavro authored and co-authored several papers and case studies published by the Harvard Business School publishing. Dr. Njavro has earned his bachelor's and master's degree in economics and finance from Bocconi University in Milan, Italy. He earned his PhD in finance from the University of St. Gallen in Switzerland. His professional experience includes working in the investment banking divisions of Lehman Brothers and Nomura in London.

Panelists: Peter van der Hijden



Peter van der Hijden is an independent higher education strategy adviser and policy influencer based in Brussels. He worked for the European Commission for 23 years in the fields of education, transport, research and innovation. At the Directorate-General for Education and Culture, he served as Head of Sector for Higher Education Policy. At the Directorate-General for Research and Innovation, he contributed to the Human Resources Strategy for the European Research Area.

Peter played a central role in the development of the Erasmus programme, the Bologna Process and the EU Modernisation Agenda for universities. He studied law at Maastricht University in the Netherlands, where he began his career as Vice-Registrar of the University Council and Policy Officer for Education and Research.

In 2025, University World News and six other journals published his article on the future of Erasmus+, entitled "Could Europe Achieve Higher Impact with a Lighter Touch?"

Panelists: Johan Roos, PhD



Johan Roos is an internationally recognized strategist, author, educator, and advisor specializing in leadership, strategy, and the impact of artificial intelligence on organizations.

A co-inventor of LEGO® Serious Play®, he has held senior leadership positions at leading international business schools, including Hult International Business School, Copenhagen Business School, and the Stockholm School of Economics. Today, he advises executives and boards around the world on navigating AI-driven transformation.

His latest book, Human Magic: Leading with Wisdom in an Age of Algorithms, explores how leaders can combine human judgment, creativity, and wisdom with the power of AI to build resilient, future-ready organizations.

Panelists: Prof. Eileen McAuliffe, PhD



Chief Thought Leadership Officer & EMEA, AACSB International

Eileen McAuliffe serves as Executive Vice President, Chief Thought Leadership Officer, and Managing Director, EMEA at AACSB. As a key member of the executive team, she leads global thought leadership initiatives, positioning AACSB as the foremost voice in business education. In her role, she champions the quality, relevance, and societal impact of business schools worldwide, while also overseeing strategic direction and operations across the EMEA region to ensure strong engagement and support for AACSB members.

Prior to joining AACSB, McAuliffe was Pro Vice-Chancellor (Global) and Executive Dean of the Faculty of Business, Law, and Social Sciences at Birmingham City University. There, she played a central role in advancing the university's 2030 and Beyond Strategy, aligning international initiatives with institutional goals. As Executive Dean, she led a faculty of 400 and a student body of 12,000, managing an annual budget of approximately £100 million.

McAuliffe also brings a wealth of industry experience, having held executive tax advisory roles at ConocoPhillips in the oil sector. She earned her PhD in 2017 and was appointed Professor of International Taxation in 2020. In addition to her academic and professional leadership, she actively contributes to global policy as a member of the United Nations Platform for the Collaboration on Tax, and is a former member of the AACSB Board of Directors.

Panelists: Dr. Carlo Maria Gallucci Calabrese



Esade-URL (Barcelona, Spain)

Born in Spezzano Albanese, Italy (1956). He holds a Bachelor degree in Economics and Commerce by Messina University (Italy), an MBA by Esade-URL and a PhD in Economics and Business Sciences by Comillas Pontificia University, ICADE-ICAI (Madrid). He also followed the Executive Program in Management at University of California at Los Angeles UCLA (USA) and the Strategic Marketing Management Program at Harvard University, Boston (USA).

He is currently Full Professor of Marketing at Esade-URL (Barcelona, Spain), where he works since 1986 and where he has held various top positions, particularly in the area of international affairs. He is member of the Esade International Advisory Board. Since January 2013 he is part of the URL Rector's Team as Vice-Rector for International Relations and Students.

Among other recognitions, in 2013 he received by the President of the Italian Republic the distinction of "Ufficiale dell'Ordine della Stella d'Italia", and in 2019 by the French Government the distinction of "Chevalier dans l'Ordre des Palmes académiques". In 2018 he received the Doctorate Honoris Causa of the Córdoba Catholic University, Argentina. He is currently member of the board of the Federation of European Catholic Universities (FUCE) and the International Federation of Catholic Universities (FIUC).

Professor Carlo Maria Gallucci Calabrese has been appointed as the new Director General of Università Cattolica del Sacro Cuore, effective 1 September 2026.

PANEL 2

Ethics of AI in Education: Key Challenges and Practical Solutions

Moderator: Kristijan Krkač, PhD



Zagreb School of Economics and Management

Assoc. Prof. Kristijan Krkač, PhD (b. 1970) has been a tenured professor at the Zagreb School of Economics and Management (ZSEM) since 2003. Before joining ZSEM, he was an associate professor at the Faculty of Philosophy and Religious Studies, University of Zagreb (1996-2017). He was a visiting professor in France 2010-2014 (Science Po Lille), in Latvia (RISEBA University) and in Italy (H-FARM College).

His research interests are mainly in the areas of business ethics, CSR and sustainability, and in the philosophy of Ludwig Wittgenstein. To date, he has published 17 authored and 11 edited books (textbooks, conference and symposium proceedings, journal issues and two encyclopedias as a section editor). He is the author and co-author of more than 150 original scientific, professional and review articles, book chapters and encyclopedia articles. He has published with Palgrave MacMillan, De Gruyter, Springer, Routledge, Ashgate, Emerald, Rowman & Littlefield and the Austrian Society of Ludwig Wittgenstein. His books in English include: A Custodian of Grammar, Essays on Wittgenstein's Philosophical Morphology (Maryland: Rowman & Littlefield, 2012), Conversations on Ethics and Business. A Guide to Thinking about Workplace Ethics (with R. Richards, and B. Jalšenjak, Cambridge: Ethics International Press, 2023) and Alone amidst the ruins of the future – Philosophy and unsustainability of multiple simultaneous disasters (Zagreb: MATE, 2025).

Panelists: Dr. Markus Raueiser



CBS International Business School in Cologne, Germany

Dr. Markus Raueiser is Director of Teaching and Learning and Professor of International Business at CBS International Business School in Cologne, Germany. Over the past two decades, he has held several academic and management positions at CBS, including Dean and Vice President for International Relations.

He teaches Strategic Management at Bachelor's, Master's, and MBA levels, with particular expertise in interactive and action-learning formats. He also serves as Visiting Professor at Antwerp Management School and has professional experience with Deutsche Bank and KPMG.

His research focuses on higher education, university management, strategy, and organisational design. Dr. Raueiser also has extensive experience in accreditation and quality assurance, including institutional accreditation processes such as IACBE and AACSB, as well as serving as a reviewer and assessor in accreditation procedures.

Panelists: Prof. Maha Mahdi Al-Khaffaf



Applied Science Private University (ASU), Jordan

Prof. Maha Mahdi Al-Khaffaf is Professor of Management Information Systems and Director of International Relations at Applied Science Private University (ASU), Jordan. She has extensive academic and leadership experience in Jordanian higher education and served as Head of the Management Information Systems Department for more than five years.

Her research interests focus on artificial intelligence in higher education, digital transformation, intelligent decision support, and the responsible and ethical use of emerging technologies, particularly in teaching, research, and academic decision-making.

Throughout her academic career, she has held several leadership positions, including Director of the Planning and Development Center and Director of the Consultation and Training Center, reflecting her experience in institutional development, strategic planning, and professional training.

Panelists: Dr. Monika Sońta



Kozminski University

Dr. Monika Sońta is an expert in communication, creativity, and artificial intelligence in education. A technology sociologist specializing in organizational culture and internal communication. She is a certified facilitator of FORTH, LEGO® SERIOUS PLAY®, PLAYMOBILpro, Design Sprint, and PROSCI methodologies.

She has over 20 years of international experience in HR and communication. She is an Assistant Professor at Kozminski University and the local coordinator of the 4InnoPipe2 project. Her research focuses on creativity in organizations in the age of AI. She is also a Visiting Scholar at Riga Business School.

PANEL 3

The Co-Scientist Era:
Self-Driving Labs, Multimodal Discovery & the Postplagiarism University

Moderator: Andrej Novak, PhD



University of Vienna,
University of Zagreb

Dr. Andrej Novak is a Research Associate at the Faculty of Mathematics and Economics at the University of Vienna and a Professor at the Faculty of Science, University of Zagreb. He holds both a Master's and a PhD in Applied Mathematics, as well as a medical degree from the University of Zagreb School of Medicine, where he graduated with a thesis on cellular models in neuroscience.

His research focuses on applied mathematics and computer science, with particular interests in partial differential equations, optimization theory, medical image processing, and the use of interpretable predictive analytics in managing cardiac pathologies.

Throughout his academic career, Dr. Novak has led and coordinated more than ten university courses in mathematics, quantitative methods, and computer science at both undergraduate and graduate levels. He has also contributed to curriculum development for several computer science programs and has taught quantitative methods across multiple institutions.

He has been actively involved in numerous national and international projects, serving as both a researcher and a member of advisory boards. As a principal investigator, he has led scientific projects and collaborated with international companies specializing in algorithmic solutions for image processing and computationally intensive problems.

Panelists: Dr. Swen Seebach



Universitat Abat Oliba CEU

Dr. Swen Seebach has served as Vice-Rector for Quality, University Strategy and Internationalisation at Universitat Abat Oliba CEU since 2025, following his tenure as Vice-Rector for Academic Affairs and Quality between 2021 and 2025. He is currently Associate Professor in the Department of Communication, where he teaches sociology, political sociology, ethics, and communication theory.

His academic career is characterised by a strongly interdisciplinary approach that integrates political science, sociology, communication theory, and Science and Technology Studies (STS). His research focuses on the analysis of emotions and the social and cultural imaginaries that shape the adoption, interpretation, and incorporation of emerging technologies into society. His work has contributed to understanding how collective expectations, emotions, narratives, and visions of the future influence innovation processes, professional practices, and social transformations associated with digitalisation and artificial intelligence.

He published a monograph Love and Society (Routledge, 2017), which has gained international recognition within the fields of social relationships, emotions, and modernity studies and he is a recognised scholar in the study of technological imaginaries and the social dimensions of artificial intelligence, digital communication, and technological transformation.

He is Institutional leader of Universitat Abat Oliba CEU within the European University Alliance EUonAIR, and director of the consolidated research group COMEMFUTURES, recognised by the Agency for Management of University and Research Grants (AGAUR).

Panelists: Prof. Vilte Auruskeviciene, PhD



Social Sciences at ISM University of Management and Economics (Lithuania)

Prof. Vilte Auruskeviciene, PhD is a Professor of Social Sciences at ISM University of Management and Economics (Lithuania). She is a marketing scholar with extensive experience in academic leadership and international higher education development.

Prof. Auruskeviciene joined ISM in 2004 and has since held a number of senior academic and executive positions, including Dean, Vice-President for Studies, Vice-President for Studies and Research and Interim President. Since 2025, she has been leading ISM activities within the European University Alliance EUonAIR.

She is a Board Member and International Director of the Marketing Educators' Association (MEA, USA), a member of the American Marketing Association (AMA), Academy of Marketing Science (AMS) and a member of the Editorial Board of the Baltic Journal of Management. She also serves as an expert for the Research Council of Lithuania and the Centre for Quality Assessment in Higher Education (Lithuania).

Her research focuses on digital marketing, consumer behaviour, and the application of artificial intelligence in marketing and consumption contexts. She has published extensively in high-impact international journals, including the Journal of Public Management, Journal of Business Research, Electronic Markets, and the Journal of Product and Brand Management.

Her research excellence has been recognized internationally. Most recently, her article (co-authored) "AI-based Chatbots in Conversational Commerce and Their Effects on Product and Price Perceptions," published in Electronic Markets, was recognized as the Paper of the Year 2024.

Panelists: Marin Njavro



Marin Njavro is the co-founder and managing director of Luxembourg School of Business. He is responsible for the School's overall strategy, operations and relations with the School's main stakeholders.

Marin has graduated from Sciences Po Paris with a Bachelor degree in political science and a Master degree in Business Law. He has previously worked for Barclays Capital in London and the corporate law department of Loyens & Loeff in Luxembourg.

Luxembourg School of Business

WORKSHOP

Directing autonomous agents for management tasks: practical workshop

Panelists: Mikael Gorsky, PhD



Holon Institute of Technology

Mikael Gorsky is a senior AI researcher at the Holon Institute of Technology (HIT), where he focuses on researching the transformative impact of generative AI on education and knowledge work. His work centers on AI-first solutions for transformation of knowledge work, drawing from decades of experience in technology and business strategy. He has been engaged with knowledge work for many years, starting from software development for 8-bit processors. He co-founded one of the first management consulting firms in new, 'capitalist' Russia and started one of the first economic development think tanks. His work focused on strategies and performance management for the financial sector, retail, real estate and IT in Kazakhstan, Ukraine and Russia. He was also implementing various ERP solutions across countries of the former USSR.

His deepest interest in knowledge work led to getting HBS education on leading consulting and advisory organizations, and he remains fascinated by intricacies and complexities of those matters. Currently at HIT he does applied research on AI in knowledge work across fields of computer science, education and management while building international partnership networks and creating training programs for students and global audiences. His work focuses on practical applications that solve real knowledge work problems, capitalizing on the revolutionary advances in AI that are reshaping how people think, analyze, create and make decisions. Knowledge work is being fundamentally transformed by AI, creating both unprecedented opportunities and serious challenges.

Traditional roles in research, analysis, writing, and decision-making are being reshaped as AI handles routine cognitive tasks with increasing sophistication. This creates inevitable changes in professions, jobs, and social fabric – some knowledge workers will find their roles eliminated, others will see dramatic productivity gains, and entirely new types of work will emerge. Generative AI may be the most important technology discovered by humans. Nations, firms and individuals have to rethink how they learn, decide and create, and this revolution calls for hard intellectual work to reflect on the new reality, define and implement structures, policies and guidelines for the era of AI.

AI, Higher Education and Institutional Transformation

SESSION CHAIR:

Karmela Aleksić-Maslać

Directing Autonomous Agents New core competency for the MBA graduate

Mikael Alemu Gorsky[1] and Ilya Levin[2]

[1]Holon Institute of Technology, Holon, Israel, mikaelg@hit.ac.il
[2]Holon School of Computer Science, Holon Institute of Technology, Holon, Israel, levini@hit.ac.il

Abstract

Business schools teach students to converse with generative AI. The work those graduates will do is being reshaped by intentional autonomous agents that execute multi-step tasks across files and applications and return finished deliverables. This paper argues that the gap is wide enough to require a change at the level of the curriculum, and that fluency in directing such agents should become a graded competency in the MBA core. The argument draws on the authors' experience teaching the adjacent discipline to software engineers, and it adapts a discipline-based curriculum for agentic software engineering to the deliverable-centred work of the business graduate. The competency is a single analytical judgment, the direction of an agent, with oversight, verification, and the safe handling of data as dimensions of doing that one thing well. Empirical work on knowledge-worker productivity, on the failure modes of agentic systems, and on how students actually use these tools supports the diagnosis and the design. The paper offers a teaching pattern built around a planted error: the agent produces a result that is plausible and wrong, and the student has to catch it. The value a graduate adds lives in the check. This is a conceptual and pedagogical framework grounded in teaching practice, and the paper sets out the classroom it comes from, the limits of what that classroom supports, and the case for carrying a discipline developed with engineering students into the business school. It closes with two models for placing the competency in a programme, the obstacle of faculty readiness, and an admission that the framework awaits empirical validation.

Keywords: *agentic AI, MBA curriculum, autonomous agents, assessment, verification, faculty development.*

1. The training and the workplace have come apart

A business school today can teach a student to use AI well and still send that student into a job for which the training does not prepare them. The lesson and the work have drifted apart, and the drift is recent enough that few faculty have noticed it. Most AI instruction in MBA programmes treat's the technology as something a person talks to. The student learns to write a good prompt, to ask a chat assistant for a summary or a first draft, to push back on a weak answer and ask again. These are real skills. A year ago they were close to the whole of what the workplace asked for.

The workplace has since moved. Tools that now enter business roles do not wait for a prompt and return a paragraph. They accept a goal stated as an outcome, plan the steps that reach it, work across a person's files and applications, and hand back a finished deliverable for the person to check. A junior analyst no longer asks the model how to structure a valuation and then builds it by hand. The analyst gives the model the folder of statements and the instruction to build it, then judges what comes back. What matters has shifted from holding a good conversation to directing a capable, autonomous co-worker. These are different skills.

This paper proposes a way to close that gap. It sets out a framework for teaching agentic AI as a graded competency in the MBA core, and uses Claude Cowork as the concrete instance of the tools graduates will meet, a system whose name signals what it is. An argument about a category is easier to follow with one real example in view. The framework itself depends on no particular tool. It rests on a claim about what a person must know to stand behind work an agent produced, and that claim is the subject of everything that follows. Section 4 sets out the classroom the framework comes from and the limits of what that classroom supports. The framework awaits empirical validation, and cannot receive it until enough programmes adopt the position argued here to produce cohorts worth measuring. A pedagogy has to be proposed and taught before it can be tested. The proposal should not wait for evidence that only adoption can generate.

2. What changed, and what the change is called

The shift has a clean dividing line, and the whole argument for changing the curriculum turns on that line being real, not a matter of degree.

On one side sits the generative model used as a conversational tool. The person supplies a prompt, the model returns text or code or an image, and the person does everything else. Answering is all it does. Orchestration, file handling, and the decision about what to do with the answer all remain with the human. A student who has only ever used AI this way has used a fast source of drafts and has assembled all the work themselves.

On the other side sits the agent, which the graduate will meet less as a tool than as a new kind of co-worker. The person supplies an outcome rather than a step, the autonomous system plans

the path to that outcome, and it carries the plan out across real files and applications before returning a deliverable. A compact definition has taken hold: language models that use tools in a loop, with autonomy rising as the underlying models improve [1], [2]. This agent is no better as a conversationalist. It does the work, inside a bounded environment the person controls, and reports back. For a knowledge worker this is the decisive change. It moves the human from doing the task to supervising it, and supervising an autonomous worker is a different job from doing the task by hand.

Claude Cowork shows what the second category looks like in practice, and its name is descriptive: it is built to be a co-worker. It is an agentic desktop system, intentional and autonomous, that carries out multi-step knowledge work on a person's behalf. It operates on local files and ordinary applications and returns completed artefacts: a report, a structured spreadsheet, a folder sorted and renamed [3]. By its maker's own account it is built around the outcome, not the prompt, and built for the people who do analytical work, not for people who write software, which is to say for analysts, consultants, operations managers, and finance professionals. It runs from the desktop application without anyone writing a line of code [3]. That last point makes it the right example for this audience. The agent the MBA graduate meets first will not be a developer's instrument. It will be outcome-oriented, centred on files, and usable by someone whose training is in business, and a graduate who has never learned to direct such a thing will be unprepared for the first week of the job.

3. The gap, and why it belongs in the graded core

Most business schools do not teach what the agentic workplace demands, and the shortfall has a specific shape worth stating before turning to what should replace it.

Where AI appears in the MBA curriculum today, it tends to appear as an elective, and the elective tends to teach conversation. A course called something like "AI for business" introduces prompting, runs some basic analysis through a chat model, and shows the class how to use the tool to write, to brainstorm, to summarise a document. The treatment casts AI as an aid to thinking and a source of raw material. It rarely casts AI as something that executes a whole workflow, because until recently AI could not, and the curriculum has not caught up. Studies of how students actually use these tools bear this out. A study of postgraduate management students at a UK business school found that they used generative AI predominantly for writing, paraphrasing, and rephrasing rather than for critical analysis, with only about one in ten reaching the more demanding stages of reasoning [4]. The graduate this produces can talk to a model with some skill, and has never been taught to design a piece of work for an agent to carry out, or to supervise that agent while it does. They know how to talk to AI. Nobody taught them to direct it.

The expectation they meet at work runs ahead of this. Adoption has moved fast and clusters in exactly the knowledge-intensive, analytical tasks an MBA prepares people for, with usage concentrated where the economic value is highest [5]. Independent survey work finds that a large share of workers now use these tools and report meaningful time savings [6]. Business roles increasingly assume that an employee will hand the repetitive parts of knowledge work

to an autonomous agent, a junior co-worker who does not tire, and spend their own time on judgment and on the client. The assumption is specific. Consolidating a stack of PDFs into the skeleton of an investment memo is the agent's job. Cleaning and structuring a file of transactions is the agent's job. The first draft of a deck is the agent's job. What remains is to frame each task, checks what comes back, and owns the result. A junior hire who can only do these things by hand is slower than a peer who directs an agent, and the gap widens with every task in the week.

Here the competency stops being a nice addition and becomes something a school must grade. The ability to design and supervise a piece of agentic work is structured analytical judgment, the same faculty the MBA already claims as its central product and already examines in every core course. A programme that grades a student's reasoning in a finance case but leaves the student to absorb agentic supervision by accident is treating the second as lesser than the first, when the workplace treats them as the same kind of thing. If graduates are assumed to direct these systems, the competency has to be taught on purpose and assessed formally. It belongs in the graded core, not at the edge of the programme in an elective nobody is required to take. The framework in section 7 and the rubric it points to are what put it there.

4. Where this account comes from

This framework comes from a classroom, not from a survey of the literature. It was developed while teaching ASE-26, a course in agentic software engineering, to third-year computer science students at Holon Institute of Technology. The course ran across 3 semesters and 5 groups, with 47 students in total. It is built from 22 modules: 5 foundational lectures that establish the agentic turn and the anatomy of a workflow, 8 paired modules that each introduce a principle and then practise it, and a closing sequence on the human role, entrepreneurship, and the job market. Three projects of increasing scale run through it, and every project is submitted with an audit trail consisting of the specification, the context pack, the workflow log, and a reflection document. The audit trail is graded alongside the code. That is how the course teaches students to treat the record as part of the engineering, not as paperwork after the fact.

That teaching supplies three things: a designed and repeatedly delivered curriculum, a grading instrument applied to real student work over 3 semesters, and the accumulated experience of watching students learn to direct autonomous agents on projects large enough to fail. It supplies no controlled measurement. No comparison group was run and no instrument was validated. The failure modes named in the next section are derived from the structure of the discipline, not observed and counted in the classroom. They are analytic categories, produced by asking what can go wrong when the division of labour between a human and an agent breaks down, and they earn their place by being coherent and teachable, not merely common. A framework paper that cannot separate what it observed from what it reasoned toward is of little use to anyone who wants to test it later.

The students are not the same students, so the step to the business school needs an argument. What transfers is the structure of the relationship between a person and an autonomous agent, and that structure has nothing to do with code. A human frames the task, supplies the context,

sets the limits, and verifies the result. An agent plans and executes. A record connects the two so that the work can be defended afterwards. Every element of that sequence describes a consultant preparing a client deliverable as accurately as it describes an engineer shipping a feature. The audit trail transfers with particular force, because a business graduate is more likely than an engineer to be asked, by a manager or a client or a regulator, why a number is what it is.

Verification does not transfer, and the difference cuts against the business school. An engineering student checking an agent's work has instruments the machine cannot talk past: a compiler that refuses to build, a test suite that fails, a program that crashes on the edge case. The verification is external to both the student and the agent, and it is unforgiving. A business student has no compiler. A valuation that draws a figure from the wrong line of a statement will still compute, still balance, and still print. A competitive matrix that misclassifies a firm will still look like a competitive matrix. The check has to be performed by judgment against sources, and nothing in the artefact announces that it is needed. This asymmetry makes the competency harder to teach outside engineering and is the reason it deserves a grade. Where the environment does not enforce verification, the curriculum has to.

A second, older piece of practice stands behind the business-school half of the argument. The first author taught strategy execution and performance management at two business schools in Moscow in the mid-2000s, to students who went on to exactly the analytical and managerial roles this paper is concerned with. That experience is too old to carry evidential weight about AI, which did not exist in the classroom then. It supports something narrower: a working knowledge of what business students are like as learners, how they respond to structured analytical method, and which parts of a framework survive contact with a cohort that is impatient with theory and quick to ask what it is for. The pedagogy proposed below is written with that audience in view.

5. Translating ASE-26 pedagogy to MBA students

The framework adapts ASE-26, a discipline-based curriculum for Agentic Software Engineering [7]. ASE-26 was written for engineering students, but its central claim is not about code. It is about how a human should direct an autonomous agent and stand behind the result, and that claim holds in any field where the output is a deliverable rather than a line of software. This section extracts that claim and carries it across to business education.

5.1 A discipline, not a toolset

ASE-26 opens by treating agentic work, which we define as work delegated to an intentional, autonomous system, as a discipline. That word is chosen with care. A discipline has principles, named methods, recognised failure modes, and a standard a practitioner can be held to. It carries from one task to the next because it teaches the structure of a relationship, not the buttons of one tool. A casual user and a trained practitioner can sit at the same screen, use the same model, point at the same task, and produce different work. The gap between them is the discipline.

That reframing is what the paper borrows. Tool training ages out with each release. A discipline does not. If business schools teach Cowork as a set of features, the lesson expires when the next tool arrives. Teach the discipline of directing an agent and the lesson survives the tool. That distinction is what lets agentic AI literacy sit in the graded core instead of a short-lived elective, as section 3 argued.

5.2 The division of labour

The discipline rests on one simple split, the same split a senior takes with a capable junior. A human frames, specifies, and judges. The autonomous agent executes. It acts as a co-worker. The human's real job is to equip the agent well and to verify what it returns. Each half of that job, taken seriously, is a skill in its own right: a clear brief, the right context, sensible limits, and a careful check against a standard.

ASE-26 teaches the split by teaching the ways it breaks. The human can treat the work as their own and redo the agent's output by hand, which loses the speed and the record at once. Or they can hand over judgment along with execution and accept whatever the agent produced. Or they can hand over the framing too, let the agent decide what the problem is, and end with confident work aimed at the wrong question. Each of these is a named failure mode, and each maps onto the business analyst. The analyst who rebuilds the spreadsheet by hand, the one who ships the agent's memo unread, and the one who never asked whether the agent understood the assignment are the same three failures in a different room.

5.3 The anatomy of the work, and the audit trail

ASE-26 breaks a single turn of work with an autonomous agent into parts that recur in every domain: the intent behind the task, the brief that states it, the context the human supplies, the plan the agent forms, the execution, the verification, and the record of all of it. None of these parts is specific to software. A consultant preparing a client deck moves through the same parts as an engineer shipping a feature.

The part that matters most for business is the last one. ASE-26 makes auditability its defining principle: work that cannot be inspected, replayed, or explained after the fact is craft, while professional work is accountable and reproducible. The artefact that delivers this is the audit trail, a record that links intent to brief, to context, to what the agent did, to how the result was checked. In software this guards against silent error. For business it does more, because a manager, a client, a board, or a regulator may later ask why a number is what it is. An agent produces this trail as a byproduct of normal use, in its transcript, its sources, and its versions. The discipline asks only that the practitioner keep it, name it, and be ready to stand on it.

5.4 Verification before trust

From the division of labour follows a rule ASE-26 states plainly: every agent output is a hypothesis until it is checked. The agent works by sampling from a probability distribution, so reasoning about what it should have produced is never enough. Only a test shows what it actually produced. Empirical study of agent failures supports the rule from the other direction. A large annotated analysis of multi-agent systems found that failures of task verification form one of the three dominant categories of breakdown, alongside specification and coordination failures [8], and separate work identifies error propagation, in which an early mistake cascades through every later step, as a central bottleneck in agent reliability [9]. The discipline also names the failure that shadows the rule, verification theatre, the checking that looks rigorous and catches nothing. For MBA students this is the heart of the competency. The value they add lives in the check. Section 6 builds a deliberate planted error into the classroom for exactly this reason.

5.5 Intent that moves: the evolutionary spiral

The deepest idea ASE-26 carries is that intent itself is not fixed at the start. Drawing on Boehm's spiral model and on the co-evolution of problem and solution described by Dorst and Cross, the curriculum holds that the human rarely knows the full shape of what they want before the work begins. The first draft teaches the framer what the brief was missing. Specification and build move together across several turns, and the work converges on what was actually wanted, even though no single turn could have named the destination. Two judgments make this rhythm work: the commit point, where the practitioner decides what to lock and what to revise, and drift detection, the sense for when the work is circling or sliding off its purpose.

A business-school faculty should find this reassuring rather than alarming, because it describes how good analysts and consultants already work. The iterative pass over a deck, the memo that sharpens its own question as it is written, the model rebuilt once the first version exposes the wrong assumption: these are the spiral, already present in the case method and the capstone. No foreign habit is being imported here. It names one the school already teaches and shows how to run it with an agent in the loop.

5.6 What this means for teaching

These commitments change what is taught and what is graded. The student is no longer assessed on a finished artefact that an agent helped produce, because the agent can produce a plausible artefact on its own. Assessment falls instead on the parts the agent cannot supply: the framing, the brief, the context, the verification, and the honest record of what was delegated. ASE-26 grades the discipline, not the artefact, and that is what makes the competency both teachable and examinable. Section 7 turns these commitments into the judgment-led framework, and the assessment claim of section 3 into a working rubric.

6. The classroom, built around a mistake

Students have to do this with their hands, and a tool like Cowork is what lets them. A demonstration that goes smoothly, though, teaches the wrong lesson. If the agent takes a messy folder and returns a clean deliverable while the class watches, the students learn that the tool is impressive, which they already believed, and they learn nothing about their own role. The session has to be built so that the student's judgment is the thing on display. Build the session around a mistake.

The tasks have to be the right ones, because a demonstration only persuades if the students recognise the work as their own future. Several analytical tasks serve well, and each maps onto a course the students already take. In a corporate finance setting, the agent receives a folder of financial statements and analyst reports and is asked to pull the key metrics, build a structured spreadsheet with working formulas, and draft a first valuation summary. In a strategy setting, it takes a set of industry reports and company filings and turns them into a competitive matrix and a short executive briefing. In a consulting setting, it works from scattered interview notes, market data, and internal memos to produce the structure of a client-ready deck. For operations, it organises a body of project documentation, renames and deduplicates the files, and assembles a status report from a pile of meeting notes. One of these should carry the full session as a worked thread. The rest are better named in passing, so the class sees the range of the work without losing the single example it can follow from start to finish.

The session itself runs in three movements, and the order is deliberate. Students first take the noisy data room and structure it by hand, for ten or fifteen minutes, long enough to feel the tedium of the work and to form a private sense of what a good result would look like. They then watch the agent do the same task, faster and more completely than they could. Last, they take the agent's output and critique it, correct it, and extend it. Doing the manual pass first earns the students the standing to judge what the agent produced. A class that has not tried the task has no basis for criticism and will accept whatever the screen shows them, which is the habit the whole exercise exists to break.

The mistake is what turns the session from a demonstration into a lesson, and the case for it rests on more than pedagogy. A preregistered field experiment with consultants at Boston Consulting Group found that AI lifts performance sharply on tasks inside its capability frontier but degrades it on tasks just outside, and that the location of that frontier is not obvious to the worker even on familiar tasks, so consultants over-relied on the tool exactly where closer supervision was needed [10]. The planted error trains a student to find that boundary by hand. At least one of the agent's outputs is arranged in advance to be plausible and wrong. A valuation pulls a figure from the wrong line of a statement and carries it forward through every dependent number, so the result is internally consistent and quietly false. The competitive matrix puts a company in the wrong category, in a way that reads as reasonable until someone who knows the industry looks closely. Finding that error is the students' task, and the discovery is the point of the day. A successful run teaches them that the agent is capable. The caught error teaches them where their own value lives, and that lesson is why the competency belongs in a graded course and not in a tool tutorial. An agent is not there to be admired. It is there to be supervised, and supervision is something the student is learning to do well enough to be graded on.

ISSN 2991-9371

7. A framework built on judgment, not a checklist

A faculty needs a structure to teach against and an examiner needs one to mark against. The obvious move is a checklist of equal parts, task design, oversight, verification, ethics, one box each, and it should be resisted. A list of co-equal pillars tells a student that supervising an agent and keeping a backup are the same size of thing. They are not. The competency is one thing. It is the act of directing an intentional, autonomous system, and the other concerns are dimensions of doing that one thing well. Putting judgment at the centre keeps the framework honest about where the difficulty lies.

7.1 The spine: directing an agent is analytical judgment

At the core of the competency sits the direction of the agent, and it resolves into a handful of real sub-skills that a student can be taught and tested on. The student learns to frame a problem as an outcome rather than as a sequence of micro-instructions, which is the form an agent like Cowork is built to receive: a goal goes in, and a plan with actions comes out. She learns to structure a multi-step inquiry so the agent has a path to follow rather than a vague wish to interpret. She learns to read the intermediate outputs as they appear and to judge them in flight, instead of waiting for a final answer to accept or reject whole. And the student learns, above all, to recognise the plausible-but-wrong result, the kind the planted error in section 6 was designed to train. None of these is an exotic new ability. Each is a near relation of what the case method already drills, which is why a business school is well placed to teach the set and why the set can be examined with the same seriousness as any other analytical skill.

7.2 Oversight as a dimension of direction

An agent like Cowork is built to keep a human in the loop, so that the consequential decisions stay with the person. Inside the spine, this means a particular teachable habit: the student reviews the agent's plan before it acts, watches what it does as it works, and approves or rejects the steps that carry consequences. A simple governance checklist has a place here, as a prompt for the student to look before the agent leaps. It does not have the standing of a separate pillar. Oversight is part of directing the agent well, and it is taught and graded as such.

7.3 Verification as a dimension of direction

Verification is the sub-skill of knowing what to check and how, brought into the open and made examinable. The student cross-checks the numbers against their sources, confirms that a cited document says what the agent claims it says, tests the internal logic of the result, and pushes on the assumptions to see where they give way. Here the assessment principle from section 3 becomes concrete, because the rubric marks how the student verifies, not whether

the final answer happens to be right. Recent work on assessment design points the same way. A controlled study at a UK university found that when an AI-integrated assessment is carefully scaffolded and students are graded partly on how they engage with and reflect on the tool's output, the assessment can build both technical fluency and critical judgment [11]. A student who catches the planted error through a disciplined check has shown the competency in full, even though the artefact reached them flawed. A student who accepts a correct-looking result without testing it has not shown it, even if the result was sound. The grade follows the checking, which is the part the student controls and the part the agent cannot supply.

7.4 Ethics, privacy, and access as a dimension of direction

Giving an agent access to real files, to email, or to institutional data raises questions of data handling and privacy that the student has to learn to weigh, and these belong inside the competency, not beside it. The risks are real. Security researchers showed that a prompt injection hidden in a document could trick the desktop agent into uploading sensitive files to an attacker, and on that basis its maker advised against using it for regulated or highly sensitive data during the preview [3]. The safe operating habits are part of directing a powerful tool responsibly: work in a dedicated sandbox and not on live systems, keep sensitive personal data out of the agent's reach, review any plan that would delete or overwrite files before it runs, and take a backup before a large change. Privacy and access are part of what it means to direct the agent competently. A student who ignores them has directed it badly, which the course can say plainly and grade accordingly. One thing the framework does not settle is how much of this a school can teach before the tools change underneath it. The discipline is meant to outlive any particular agent, but that claim is a bet, and the first cohorts to be graded on it will be the ones who find out whether it was a good one.

8. Two ways to put it in the programme, and the harder problem behind both

A school persuaded that this competency belongs in the core still has to decide where it sits, and the choice is open. Better to lay out the trade-off than to prescribe an answer. The competency can live in a course of its own, a required core module on directing autonomous systems for business work, where students get sustained hands-on practice with a tool like Cowork. Or it can be woven through the existing courses, so that the finance course teaches it on a valuation, the strategy course on a competitive analysis, the operations course on a project archive, and each discipline shows the student how an autonomous co-worker behaves in that discipline's particular work. The standalone module buys depth and a clear place to put the grade, at the cost of looking like a bolt-on that the rest of the curriculum can ignore. Infusion buys integration and the lesson that agentic work is part of every function, at the cost of consistency, since it depends on a dozen instructors teaching to a common standard. A school should choose by its own structure and culture, not by a rule from outside.

The harder problem sits behind both options. Many faculty have not used these tools themselves, and a faculty member who has never directed an agent cannot teach a student to do it, still less grade the doing. Some are wary of handing a system access to their files, and that wariness is a sound instinct the course will have to address and not dismiss. The way through is to start where the risk is low. A first exposure can run on synthetic or public data, inside a folder scoped tightly enough that nothing important is within reach, on a task small enough to see all the way through. The gentle start exists to let a cautious instructor build the experience to teach from. No competency can be handed to students by a faculty that has not first acquired it, and that acquisition is the real precondition for everything this paper recommends.

9. What the graduate becomes, and what the school risks

The change lands, in the end, on the graduate, and it changes what the graduate is for. Work that used to fill the first years of an analytical career, the consolidating and cleaning and first-drafting, is the work an agent now does. What remains for the person is the part the agent cannot reach: to decide what the problem actually is, to frame the task so the agent can act on it, to bring together what several functions know, and to make the judgment that sits on top of the artefact the agent produced. This is the reconfiguration the field experiment on knowledge work anticipated, with human labour moving toward supervisory and curatorial roles rather than disappearing [10]. The graduate moves from executing the steps to directing the work and owning the decision, and the comparative advantage of an MBA, if the degree still has one, lies in that higher work. A degree has to be about the judgment now, because the labour is gone.

For the school, the consequence cuts both ways, and the downside is sharper than the upside is comfortable. The wider economy shows how early the moment is: most firms are investing heavily, yet by one large survey only a tiny fraction of leaders judge their organisations mature in their use of AI, which means the firms hiring these graduates are still working out how to deploy the technology and need people who can help them do it [4]. A programme that leaves agentic AI out, or files it under an elective, sends its graduates into those firms unable to supervise these systems safely or to get the full value from them, and the cost shows up twice, once in the employability of the graduate and once in the reputation of the school that produced them. A programme that builds the competency into the core can say something its competitors cannot, that its graduates arrive ready to direct the tools the work now requires, and that claim is worth something to students choosing a programme and to firms choosing where to recruit. Agentic AI is not one more tool to slot into an existing syllabus. It is the arrival of a new kind of co-worker, intentional and autonomous, in the room where entry-level and mid-level knowledge work happens. A change of that size has to be met at the level of the curriculum, not at the level of a single optional course.

10. Conclusion

One claim carries the paper, and a business school can act on it. Teaching MBA students to hold a skilled conversation with a generative model is no longer enough, because the work they are headed into is being reshaped by intentional, autonomous agents, new co-workers, that do the work instead of answering questions. The skill that work demands is the direction of those agents under a human's judgment. This paper contributes a diagnosis of the gap between how AI is taught and how it will be used, a framework that treats agentic AI literacy as a single judgment-led competency, and a set of teaching patterns, the planted error chief among them, that make the competency something a classroom can practise and an examiner can grade.

This paper cannot yet offer evidence that the framework works. Its claims are arguments from the shape of the technology and from one author's experience teaching the adjacent discipline to engineers, not findings from a controlled study of MBA cohorts. Only empirical work would settle the matter: measuring what students who learn this way can do that their peers cannot, then following graduates into the workplace to see whether the competency holds up against the job, then building shared teaching materials across schools so that the approach can be compared instead of asserted. Until that work is done, the case rests on reasoning. The reasoning seems to us strong enough to act on now. Waiting for evidence means graduates keep walking into a workplace the curriculum did not prepare them for, which is a more expensive way to learn the same lesson. We should also say what would embarrass this argument. If agents improve to the point where they catch their own errors reliably, the planted-error lesson becomes a period piece and the competency we describe shrinks to something a short workshop could cover. We do not think that is close, because the verification problem is not a defect of current models but a consequence of how they work. If we are wrong about that, we would rather have taught a cohort to check carefully and turned out not to need it than to have skipped the lesson and been right for the wrong reason.

References

[1] Anthropic. Building Effective AI Agents. Anthropic, 2025. https://www.anthropic.com/research/building-effective-agents

[2] Anthropic. Effective Context Engineering for AI Agents. Anthropic, 2025. https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents

[3] Anthropic. Claude Cowork: Agentic AI for Knowledge Work. Anthropic, 2026. https://www.anthropic.com/product/claude-cowork

[4] Fischer, I., et al. Making Sense of Generative AI for Assessments: Contrasting Student Claims and Assessor Evaluations. The International Journal of Management Education, 2025.

[5] Anthropic. Anthropic Economic Index: Uneven Geographic and Enterprise AI Adoption (September 2025 Report). Anthropic, 2025. https://www.anthropic.com/research/anthropic-economic-index-september-2025-report

[6] McKinsey & Company. Superagency in the Workplace: Empowering People to Unlock AI's Full Potential at Work (AI in the Workplace: A Report for 2025). McKinsey & Company, 2025.

[7] Gorsky, M. ASE-26: a curriculum for agentic software engineering as a discipline. arXiv:2606.01152 [cs.CY], https://doi.org/10.48550/arXiv.2606.01152

[8] Cemri, M., Pan, M. Z., Yang, S., Agrawal, L. A., Chopra, B., Tiwari, R., Keutzer, K., Parameswaran, A., Klein, D., Ramchandran, K., Zaharia, M., Gonzalez, J. E., and Stoica, I. Why Do Multi-Agent LLM Systems Fail? arXiv preprint arXiv:2503.13657, 2025.

[9] Zhang, G., et al. Where LLM Agents Fail and How They Can Learn from Failures. arXiv preprint arXiv:2509.25370, 2025.

[10] Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality. Organization Science, 37(2): 403–423, 2025.

[11] Martin, A. F., Tubaltseva, S., Harrison, A., and Rubin, G. J. Participatory Co-Design and Evaluation of a Novel Approach to Generative AI-Integrated Coursework Assessment in Higher Education. Behavioral Sciences, 15(6): 808, 2025.

[12] Federal Reserve Bank of St. Louis. The State of Generative AI Adoption in 2025. On the Economy, Federal Reserve Bank of St. Louis, 2025. https://www.stlouisfed.org/on-the-economy/2025/nov/state-generative-ai-adoption-2025

Integrating AI Avatars into Higher Ed: Balancing Technology and Academic Quality

Andrzej Rudko

Koźminski University, Poland

Abstract

Artificial intelligence is rapidly changing how universities create and share learning materials. This presentation outlines the operational framework behind MyAI University—an initiative within the EUonAIR project designed to help institutions jointly produce modular, AI-supported online courses. Coordinating international micro-courses featuring academic expertise, AI avatars, multimedia, and digital assessments presents unique logistical and quality control challenges. To address this, we present a hands-on production workflow tailored for a multi-partner environment. The presentation will walk through the actual steps of course creation: structuring modules, calculating instructional workload, drafting avatar-ready scripts, and implementing standardized quality assurance checkpoints. We will also share tested solutions to common design bottlenecks, such as aligning learning outcomes with assessments, synchronizing AI narration with visual assets, and maintaining transparency regarding generative AI use. Crucially, this case study demonstrates how to position AI avatars as a practical tool within a facultydriven process, ensuring that technology scales delivery without replacing human oversight. Attendees will receive an actionable blueprint and practical recommendations for scaling digital learning resources and building shared, AI-enhanced educational ecosystems across institutions.

Keywords: *AI avatars; artificial intelligence in education; higher education; academic quality; instructional design; interuniversity collaboration; digital learning; quality assurance; EUonAIR; MyAI Universi*

ISSN 2991-9371

Accreditations as Drivers of Organizational Learning and Strategic Development in Higher Education

Edda Apfenthaler[1], Maja Martinović[2] and Zoran Barac[3]

[1]Zagreb School of Economics and Management, Zagreb, Croatia, eapfenth@zsem.hr
[2]Zagreb School of Economics and Management, Zagreb, Croatia, mmartino@zsem.hr
[3]Zagreb School of Economics and Management, Zagreb, Croatia, zbarac@zsem.hr;
Luxembourg School of Business, Luxembourg

Abstract

This study examines how accreditations contribute to the development of quality assurance (QA) in higher education institutions through organizational learning. Drawing on Organizational Learning Theory (OLT), a conceptual model was developed linking accreditation, organizational learning, internal QA systems, quality culture, continuous improvement and strategic development, and stakeholder resistance. The qualitative study was based on semi-structured interviews with four experienced QA and accreditation experts from Croatian higher education institutions. Data were analyzed using Reflexive Thematic Analysis and a hybrid deductive-inductive approach. The findings indicate that accreditation can serve as a development mechanism when evaluation results, recommendations, and stakeholder feedback are transformed into organizational learning and embedded in institutional practices. Organizational learning strengthens internal QA systems, which in turn support the development of quality culture and continuous improvement. International accreditations additionally contribute to benchmarking, international recognition, and strategic positioning. Stakeholder resistance may hinder these processes, although its impact depends on institutional context and change management. Leadership emerged as an additional key enabling factor, influencing whether accreditation recommendations are translated into meaningful organizational change.

Keywords: *accreditation; organizational learning; quality assurance; quality culture; higher education; strategic development*

ISSN 2991-9371

1. Introduction

Quality in higher education has become increasingly important due to internationalization, competition, and growing demands for accountability and continuous improvement [1]–[3]. Accordingly, national and international accreditations are increasingly viewed not only as external evaluation mechanisms but also as drivers of internal quality assurance (internal QA), institutional improvement, and strategic management [3]–[8]. This reflects a broader shift from formal quality control toward continuous improvement and shared quality values [1], [9], [10].

This study is grounded in Organizational Learning Theory (OLT), which explains organizational development through the creation, acquisition, interpretation, sharing, and application of knowledge [11]–[13]. Organizational learning occurs at the individual, group, and organizational levels [13], providing a framework for understanding how knowledge gained through accreditation is embedded in institutional practices. In higher education, accreditation can foster such learning through self-evaluation, data analysis, feedback, stakeholder involvement, and continuous improvement [3], [14[, [15].

The study explores how national and international accreditations contribute to organizational learning, internal QA, quality culture, continuous improvement, and strategic development, and identifies best practices for QA. Based on the theoretical framework and previous research, five theoretical propositions and one sub-proposition were formulated. These propositions guide the conceptual model, semi-structured interviews, and the interpretation of the qualitative findings.

The paper presents the theoretical framework and conceptual model, followed by the methodology, thematic analysis, and conclusion with limitations and recommendations for future research.

2. Theoretical Framework and Research Propositions

Organizational learning represents one of the key theoretical frameworks for understanding the development of QA systems in higher education [11]–[13]. Previous research indicates that national and international accreditations are not merely instruments of external evaluation but can also foster organizational learning, the development of internal quality management mechanisms, and continuous institutional improvement [3], [4]. Contemporary approaches increasingly shift the emphasis from compliance with formal requirements toward the development of a quality culture in which evaluation results are used for decision-making and strategic management [9], [16], [17].

The literature review indicates that international accreditation systems, such as AACSB, EQUIS, and others, encourage institutions to systematically collect and analyze data, involve stakeholders, and align quality processes with strategic objectives [18]–[21]. At the same time, research emphasizes that the effects of accreditation depend on an institution's ability to integrate standards into its day-to-day management rather than treating them solely as administrative requirements [10], [18], [22].

The literature also suggests that organizational learning provides an important foundation for developing an effective internal QA system [11]–[13]. Such systems include self-evaluation, performance indicators, feedback from students and other stakeholders, evidence-based decision-making, and the continuous improvement of study programs and institutional processes [14], [23[–[25]. Numerous authors emphasize that a formally established quality system is not sufficient in itself. Its effectiveness depends on a well-developed quality culture characterized by leadership commitment, employee involvement, a shared understanding of quality objectives, and the continuous use of evaluation results to improve institutional performance [2], [9], [10], [26]. In this context, quality culture represents a link between the QA system and the long-term strategic development of a higher education institution.

Contemporary quality management models emphasize that quality achieves its full impact only when it is linked to strategic management, the institution's mission, and the continuous improvement of educational, research, and management processes [10], [27], [28]. International accreditations can additionally contribute to the international recognition, reputation, and differentiation of higher education institutions, particularly in an environment of increasing international competition [29]. At the same time, research indicates that the effectiveness of quality systems may be constrained by stakeholder resistance, administrative burden, lack of motivation, and limited involvement of faculty and other employees. Therefore, the successful implementation of QA systems requires the active participation of all stakeholders, effective leadership, and an organizational culture that encourages knowledge sharing, collaboration, and organizational learning [9], [15], [30].

Based on the theoretical framework presented above, a conceptual research model was developed, and research propositions P1–P5 and sub-proposition P4a were formulated to represent the expected relationships among national and international accreditations, organizational learning, the internal QA system, quality culture, the strategic development of the institution, and stakeholder resistance:

- *P1: National and international accreditations positively influence the development of organizational learning in higher education institutions.*
- *P2: Organizational learning positively contributes to the development of an effective internal QA system.*
- *P3: A well-developed internal QA system positively influences the development of quality culture in higher education institutions.*
- *P4: A well-developed quality culture positively influences the continuous improvement of educational, research, and management processes, as well as the strategic development of higher education institutions.*
- *P4a (sub-proposition): International accreditations contribute to the differentiation and international recognition of higher education institutions through the development of quality culture.*
- *P5: Stakeholder resistance negatively influences the development of organizational learning, the internal QA system, and quality culture.*

3. Conceptual Model

Based on a review of the relevant literature and the identified research gap, a conceptual research model (Figure 1) was developed to illustrate the proposed relationships among national and international accreditations, organizational learning, the internal QA system, quality culture, and the continuous improvement and strategic development of higher education institutions. Stakeholder resistance is incorporated into the model as a limiting factor that may affect the other constructs. The model also assumes that developing a quality culture can contribute to differentiation and international recognition. All relationships presented in the model derive from the theoretical framework and are linked to the research propositions formulated in this paper.

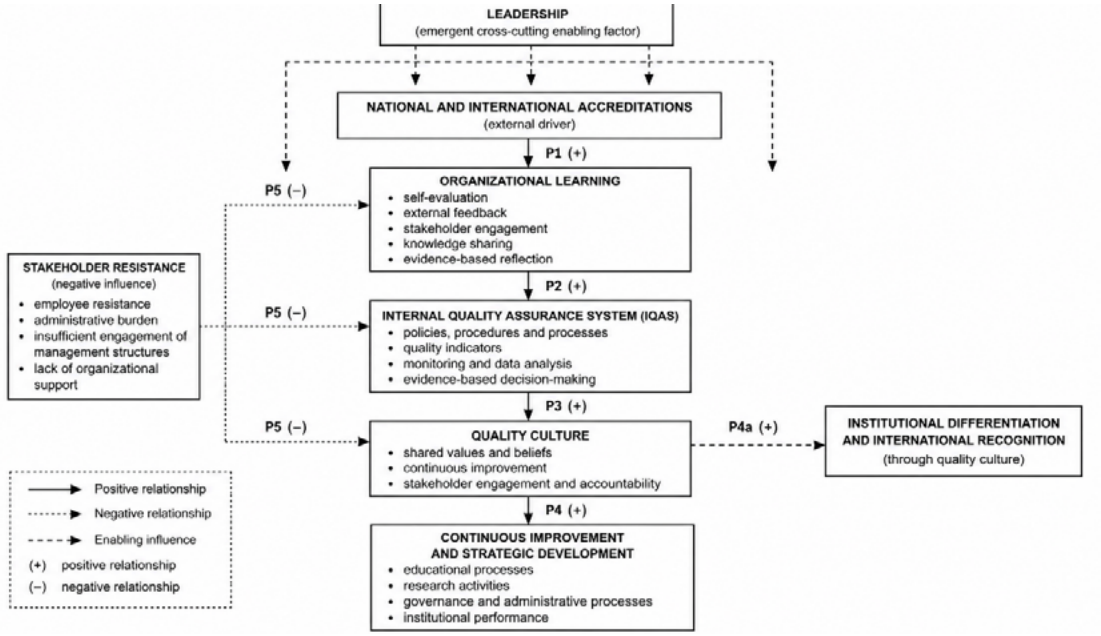

Figure 1. Conceptual Research Model

Source: Authors' conceptualization based on the reviewed literature, prepared with ChatGPT assistance, July 2026.

3.1 Explanation of the Conceptual Research Model

The conceptual research model is grounded in OLT, according to which organizations develop through the acquisition, interpretation, sharing, and institutionalization of knowledge [11]–[13], [31]. Accordingly, national and international accreditations are viewed not only as external evaluation mechanisms but also as drivers of organizational learning and institutional development. The model links accreditation, organizational learning, internal QA, quality culture, and continuous improvement and strategic development. Positive (+) and negative (-) signs indicate the assumed direction of relationships between the constructs.

Following the qualitative analysis, leadership was incorporated into the conceptual model as an emergent cross-cutting enabling factor. Rather than representing an additional research proposition, leadership reflects a contextual mechanism that facilitates the translation of accreditation recommendations and organizational learning into internal QA development, quality culture, and institutional change. P1 (+) proposes that national and international accreditations foster organizational learning through self-evaluation, feedback, stakeholder involvement, knowledge sharing, and evidence-based reflection [11], [12]. P2 (+) assumes that organizational learning strengthens the internal QA system by embedding evaluation results, feedback, and acquired knowledge into institutional practices [11], [13]. P3 (+) proposes that a developed internal QA system contributes to quality culture by integrating systematic monitoring, evidence-based decision-making, stakeholder involvement, and continuous improvement into institutional practices [11], [31]. Quality thus becomes a shared institutional value rather than solely the responsibility of formal QA structures [13], [31]. P4 (+) posits that a quality culture supports continuous improvement and strategic development by linking evaluation results to institutional decision-making and development priorities [13], [31]. P4a (+) further proposes that international accreditations contribute to differentiation and international recognition through the genuine integration of international standards into quality culture and strategic management [13], [31]. Finally, P5 (–) proposes that stakeholder resistance may hinder organizational learning, internal QA, and quality culture by reducing openness to feedback, participation, knowledge sharing, and willingness to change [12], [31]. Thus, accreditation alone does not ensure institutional development; its effects depend on the institution's ability to transform evaluation results and stakeholder feedback into organizational learning and sustainable change.

The conceptual model served as the theoretical basis for developing the semi-structured interview protocol and as an analytical framework for interpreting the findings in relation to the research propositions and OLT.

4. Methodology

The qualitative research was conducted through semi-structured interviews with four experts selected via purposive sampling for their extensive experience in QA and national and/or international accreditation. This approach was appropriate for obtaining information-rich, context-specific insights rather than statistically generalizable results. The interview participants were experts responsible for QA at different higher education institutions in the Republic of Croatia. All participants have held or currently hold senior positions related to the development, management, or evaluation of QA systems in higher education institutions. Their professional experience ranges from more than ten to over fifteen years and includes involvement in national reaccreditation procedures, international institutional accreditation processes, and participation in national and international expert panels. Regarding their key professional experience in the field, the participants reported familiarity with the following systems and frameworks: ASHE, AACSB, EQUIS, EFMD, ESG, EHEA, ASIIN, HAA, NAKVIS, PRME, and ISO. Two participants have also served on the Accreditation Council of ASHE (the Agency for Science and Higher Education), while others are actively involved in international accreditation processes as panel members or mentors. The institutions represented by the participants vary in size in terms of the number of students, employees, and external associates, as well as in their organizational characteristics (public and private sectors) and academic fields (Table 1).

Participants	Description of the Institution
A	Public university faculty in the field of technical sciences, based in Zagreb
B	Public university faculty in the field of technical sciences, located in Eastern Croatia
C	Private university of applied sciences in the field of social sciences, based in Zagreb
D	Public university faculty in the field of social sciences, located in coastal Croatia

Table 1. Description of the Institutions Represented by the Participants in the Sample

The interviews were conducted in July 2026. Following the initial contact, participants provided written responses to a set of pre-prepared open-ended questions, with additional follow-up discussions conducted where necessary to clarify individual responses. The qualitative data were analyzed using Reflexive Thematic Analysis (RTA) [32]–[34], which enables the systematic identification of patterns of meaning within participants' experiences and their connections to the study's theoretical framework.

4.1 Qualitative Data Analysis

Given that the research was based on a pre-developed conceptual model, a hybrid deductive–inductive approach was applied. The deductive component of the analysis was based on the theoretical constructs of the conceptual model, while the inductive component enabled the identification of new patterns emerging from the participants' responses. The analysis was conducted through six phases: familiarization with the data, initial coding, grouping codes into subthemes, development of themes, comparison of responses across all participants, and interpretation of the findings in relation to the research questions and theoretical framework. The credibility of the findings was enhanced through comparisons of responses across participants, multiple reviews of the data during coding, and the use of direct participant quotations to support the interpretation of the findings. The results were organized around common themes identified across all or most participants, with interpretations supported by representative quotations from participants, anonymously identified as A–D.

4.2 Results of the Thematic Analysis

The thematic analysis identified seven main themes describing how national and international accreditations contribute to the development of QA systems in higher education.

4.2.1. National and International Accreditations as a Development Mechanism

The analysis of the interviews shows that participants do not primarily perceive accreditations as regulatory instruments, but rather as a development mechanism that fosters organizational learning, enhances the internal QA system, and guides the institution's strategic development. However, all participants emphasize that the effects of accreditation depend on the institution's willingness and ability to translate recommendations obtained during the evaluation process into concrete development activities. Participants distinguish between the roles of national and international accreditations. National accreditations are associated with the development of the internal QA system, compliance with standards, and continuous process improvement, while international accreditations are associated with internationalization, benchmarking, international recognition, and the strategic positioning of the institution. As one participant stated: *"The greatest value of national accreditation lies in the recommendations which, if accepted by the leadership, become a driver of real change."* (Participant D) Referring to international accreditations, another participant emphasized: *"International accreditation was a means through which we differentiated ourselves from other institutions and defined our strategy more clearly."* (Participant C) Participants also agree that national and international accreditations are not competing but complementary systems. Particular emphasis was placed on the fact that the impact of accreditation arises from how the institution uses the recommendations it receives. As one participant summarized: *"Any accreditation can serve as a development tool, but only if the institution is willing to learn and change."* (Participant C) These findings are consistent with OLT, which suggests that external evaluation can stimulate organizational learning when feedback and acquired knowledge are interpreted and translated into institutional practices [11]–[13], and with previous research emphasizing the developmental role of accreditation beyond formal compliance [3], [4], [18].

Subtheme	Participants	Main Codes
National Accreditations as a Framework for Institutional Development	A, B, C, D	Recommendations, Self-evaluation, ESG, Quality System Development
International Accreditations as an Instrument for Internationalization	A, C, D	International Visibility, Benchmarking, Competitiveness
Accreditations as a Driver of Organizational Learning	A, B, C, D	Reflection, Recommendations, Good Practice Examples
Development-Oriented vs Compliance-Oriented Approaches	A, B, C, D	Implementation of Recommendations, Meaningful Changes
Complementarity of National and International Accreditations	A, C, D	Different Objectives, Mutual Complementarity

Table 2. Thematic Analysis of the Construct "National and International Accreditations as a Development Mechanism"

4.2.2. Organizational Learning as a Driver of Quality Assurance System Development

All participants consider organizational learning a key mechanism by which accreditations contribute to the development of QA systems. The value of accreditation does not arise from the evaluation process itself, but from the institution's ability to transform evaluation results into organizational knowledge that leads to concrete changes. Participants associate organizational learning with the analysis of evaluation results, self-assessment, stakeholder involvement, and data-driven decision-making. As one participant stated: *"The greatest value of accreditation is not the evaluation itself, but what the institution does with the recommendations after the evaluation."* (Participant D) Similarly, another participant emphasized: *"Self-assessment is perhaps the most valuable part of the entire process because it allows you to objectively assess your own strengths and weaknesses for the first time."* (Participant A) The findings also show that organizational learning is not solely the responsibility of the QA office, but rather a collective process involving faculty members, students, management, and administrative staff. The systematic use of evaluation results enables evidence-based decision-making and facilitates the integration of QA activities with the institution's strategic development. This finding directly reflects OLT, according to which organizational development depends on the acquisition, interpretation, sharing, and institutionalization of knowledge [11]–[13], and is consistent with previous research linking accreditation, self-evaluation, and stakeholder feedback to continuous institutional improvement [3], [14], [15].

Subtheme	Participants	Main Codes
Use of Evaluation Results	A, B, C, D	Analysis of Results, Recommendations, Action Plans
Organizational Learning through Accreditations	A, B, C, D	Self-evaluation, Peer Review, Recommendations, Reflection
Stakeholder Engagement	A, B, C, D	Students, Faculty, Management, Administrative Staff
Knowledge Sharing	A, B, D	Communication, Collective Learning, Exchange of Experiences
Data-Informed Decision-Making	B, C, D	KPIs, Evaluations, Indicators, Strategic Decision-making
Continuous Improvement	A, B, C, D	Process Development, Corrective Actions, Quality Culture

Table 3. Thematic Analysis of the Construct "Organizational Learning"

4.2.3. Internal Quality Assurance System

The thematic analysis shows that participants do not perceive the internal QA system as a set of administrative procedures or documentation, but rather as an integrated management system that enables the planning, monitoring, evaluation, and continuous improvement of all key processes within a higher education institution. All participants emphasize that the system's effectiveness depends on how evaluation results are used in day-to-day management. Assessments of the level of development of QA systems vary across institutions, indicating different levels of institutional maturity. One participant stated: *"A developed QA system is not a collection of regulations, but rather the way in which an institution uses information to make better decisions."* (Participant D) Another participant emphasized: *"There is always room for improvement, particularly in the more formal monitoring of the implementation of recommendations and the development of information tools for quality management."* (Participant A) Participants particularly emphasize the importance of the systematic use of evaluation results. Recommendations from internal and external evaluations are a valuable source of information for improving study programs, the teaching process, student support, and institutional management. However, their value is realized only when they are linked to specific action plans, clearly assigned responsibilities, and systematic monitoring of their implementation. The analysis also shows that an effective QA system is inseparably linked to strategic management. As one participant summarized: *"A strategy has little meaning if it is not based on the results of the QA system."* (Participant D) Participants clearly distinguish between the formal and developmental dimensions of the QA system. Formal elements, such as regulations, procedures, reports, and forms, provide a necessary foundation for the system, but are not sufficient to ensure its effectiveness. The developmental dimension includes the use of evaluation results, data-driven decision-making, organizational learning, and continuous process improvement. From an OLT perspective, these findings indicate that

internal QA becomes development-oriented when knowledge generated through evaluation and feedback is embedded in institutional routines and decision-making [11]–[13], consistent with previous research emphasizing systematic evaluation and evidence-based improvement as key elements of effective QA systems [14], [23]–[25].

Subtheme	Participants	Main Codes
System Maturity	A, B, C, D	System Maturity, Institutional Development, Formalization
Use of Evaluation Results	A, B, C, D	Recommendations, Action Plans, Corrective Actions
Systematic Quality Monitoring	A, B, C	KPIs, Performance Indicators, Data Analysis
Integration of Quality into Institutional Management	A, C, D	Strategy, Decision-making, Process Management
Formal and Development-Oriented Elements of the System	A, B, C, D	Procedures, Documentation, Quality Culture, Continuous Improvement

Table 4. Thematic Analysis of the Construct "Internal QA System"

4.2.4. Quality Culture

Participants consider quality culture to represent the highest level of development of the QA system. Quality culture reflects how all members of the academic community think and act, and it implies a shared responsibility for the continuous improvement of the institution. Participants emphasize that a culture of quality cannot be developed solely through accreditation processes or administrative requirements. Rather, it develops gradually through organizational learning, open communication, stakeholder involvement, and consistent leadership support. A developed QA system is a prerequisite for creating a culture in which quality becomes an integral part of everyday work. One participant stated: *"Quality culture means that everyone thinks about how they can do their job better, not because accreditation requires them to do so."* (Participant A) A similar view was expressed by participant B*: "Quality is not the responsibility of the QA office, but of all employees."* Participants also emphasize that a culture of quality develops when employees understand the purpose of QA activities and recognize their contribution to institutional development. In this process, leadership plays a key role by modeling organizational values, fostering collaboration, and creating an environment in which quality becomes a shared responsibility among all stakeholders. These findings are consistent with OLT because they illustrate how shared learning and the institutionalization of knowledge can transform formal QA practices into collective organizational values [11], [13], [31], supporting previous research that emphasizes stakeholder involvement and leadership commitment in developing a sustainable quality culture [2], [9], [10], [26].

Subtheme	Participants	Main Codes
Quality Culture as a Shared Value	A, B, C, D	Responsibility, Shared Values, Everyday Practice
The Role of Leadership in Developing a Quality Culture	A, B, C, D	Leadership Example, Vision, Motivation
Engagement of all Stakeholders	A, B, C, D	Faculty, Students, Administrative Staff, Shared Responsibility
Continuous Improvement	A, B, C, D	Development, Reflection, Learning, Improvement
Quality Culture Beyond Formal Compliance	A, B, C, D	Development-oriented Approach, Meaningful Changes, Sustainability

Table 5. Thematic Analysis of the Construct "Quality Culture"

4.2.5. Continuous Improvement and Strategic Development of the Institution

Participants consider continuous improvement to be a fundamental objective of the QA system. According to their responses, an effective QA system does not end with the completion of an evaluation or accreditation process but involves a continuous cycle of planning, implementation, evaluation, and improvement of all key processes within a higher education institution. Participants emphasize that accreditations are an important driver of development, but their actual value lies in the institution's ability to incorporate expert panel recommendations into strategic planning and translate them into concrete development activities. As one participant stated: *"Accreditation is not the end of the process, but the beginning of a new cycle of development."* (Participant D) Another participant highlighted the importance of data-driven decision-making: *"Quality indicators enable us to make development decisions based on facts rather than assumptions."* (Participant A) Participants associate strategic development with the continuous use of evaluation results, data-driven decision-making, and the long-term development of a quality culture. In this context, the QA system is not perceived as a separate administrative function, but rather as an integral part of the institution's strategic management. This finding is consistent with OLT, which views organizational learning as a process through which acquired knowledge is translated into sustained organizational change [11]–[13], and with previous research linking quality management to continuous improvement and strategic institutional development [10], [27], [28].

Subtheme	Participants	Main Codes
Continuous Improvement as an Ongoing Process	A, B, C, D	Development, Action Plans, Improvements
Strategic Planning	A, B, C, D	Strategy, Development Objectives, Institutional Development
Data-Informed Decision-Making	A, B, C, D	Performance Indicators, Evaluations, KPIs, Analysis of Results
Accreditations as a Driver of Institutional Development	A, B, C, D	Recommendations, Implementation, Development
Sustainability of the QA System	A, B, C, D	Continuity, Long-term Development, Institutional Maturity

Table 6. Thematic Analysis of the Construct "Continuous Improvement and Strategic Development of the Institution"

4.2.6. Stakeholder Resistance in the Quality Assurance System

Participants do not consider stakeholder resistance an unexpected phenomenon or the main obstacle to developing a QA system. Rather, resistance is described as a natural response to organizational changes associated with the introduction of new procedures, standards, and ways of working. However, the findings also show that the intensity of resistance largely depends on how leadership manages change, engages stakeholders in decision-making, and communicates the purpose of QA activities. Participants emphasize that resistance most often occurs when employees perceive quality-related activities as an additional administrative burden or do not understand their purpose. In contrast, in institutions where QA processes are integrated into everyday work and aligned with strategic objectives, the level of resistance is considerably lower. As one participant stated: *"Resistance is a normal part of any change. It is the responsibility of leadership to recognize and manage it."* (Participant D) Another participant emphasized the importance of understanding the purpose of the QA system: *"People are much more likely to accept quality-related activities when they understand that these activities contribute to improving the institution rather than merely representing an administrative obligation."* (Participant A) Participants believe that involving faculty members, students, and administrative staff in planning and evaluation processes increases acceptance of the QA system and fosters a sense of shared responsibility. At the same time, they caution that excessive bureaucratization may undermine the system's developmental function and reduce employees' motivation to participate actively. From an OLT perspective, stakeholder resistance may constrain organizational learning by limiting participation, knowledge sharing, and openness to change [12], [31], which is consistent with previous research identifying resistance, administrative burden, and limited stakeholder involvement as barriers to effective QA development [9], [15], [30].

Subtheme	Participants	Main Codes
Resistance as a Natural Response to Change	A, B, C, D	Change, Uncertainty, Additional Workload
Understanding of the Purpose of the Quality System	A, B, C, D	Communication, Acceptance, Awareness
The Role of Leadership in Reducing Resistance	A, C, D	Change Management, Motivation, Support
Stakeholder Engagement	A, B, C, D	Participation, Shared Decision-making, Responsibility
Administrative Burden	B, C	Bureaucracy, Documentation, Process Efficiency

Table 7. Thematic Analysis of the Construct "Stakeholder Resistance in the QA System"

4.2.7. Leadership as a Driver of the Quality Assurance System

The interview analysis shows that leadership plays a key role in connecting organizational learning, the internal QA system, and quality culture. All participants emphasize that quality does not emerge from formal procedures or accreditation requirements, but rather from management's strategic commitment to institutional development. Participants believe that leadership determines whether recommendations obtained through accreditation processes are translated into concrete development activities or remain merely formal requirements. This is clearly illustrated by one participant's statement: *"National accreditation fosters continuous improvement only if leadership has a vision for implementing the guidelines arising from the accreditation process."* (Participant D) A similar view was expressed by another participant: *"Everything depends on the person at the top of the institution."* (Participant C) In addition to guiding institutional development, leadership plays an important role in ensuring the necessary resources, managing organizational change, and motivating employees to participate in QA activities. Participants emphasize that management, through its own example, shapes organizational culture and encourages the acceptance of quality as a shared value. Leadership plays an enabling role consistent with OLT by fostering shared learning and institutional change [31], while previous research emphasizes its importance for effective QA systems and quality culture [2], [9], [10], [26].

Subtheme	Participants	Main Codes
Strategic Leadership	A, B, C, D	Vision, Mission, Strategic Management
Leadership Support	A, B, C, D	Commitment, Resources, System Development
Change Management	A, C, D	Motivation, Implementation of Recommendations, Changes
Development of a Quality Culture	A, B, C, D	Leadership Example, Shared Values, Responsibility
Leadership and Accreditations	A, B, C, D	Operationalization of Recommendations, Development-oriented Approach

Table 8. Thematic Analysis of the Construct "Leadership as a Driver of the QA System"

5. Conclusion

Based on OLT [11]–[13], [31], the findings are generally consistent with the relationships proposed in the conceptual model. Regarding P1, the findings illustrate how national and international accreditations can foster organizational learning through self-assessment, feedback, stakeholder involvement, and institutional change, consistent with OLT [11]–[13]. The interviews also indicate complementary roles: national accreditation primarily contributes to internal processes and compliance with standards, while international accreditation additionally promotes benchmarking and internationalization. The findings related to P2 illustrate how organizational learning strengthens internal QA through the systematic use of evaluation results and feedback, rather than merely through formal procedures [13], [14]. Regarding P3, participants' accounts indicate that a developed internal QA system can contribute to quality culture by embedding quality into everyday practices and shared responsibility [2], [9]. The findings are also consistent with P4, illustrating how quality culture facilitates continuous improvement and strategic development by linking evaluation results with organizational learning, planning, and institutional development [10], [28]. Regarding P4a, the qualitative findings suggest that international accreditation can contribute to differentiation and international recognition through benchmarking, internationalization, and strategic positioning [18]. However, the interviews indicate that sustainable differentiation requires international standards to be embedded in institutional culture and strategy. Finally, the findings related to P5 indicate a more context-dependent relationship. Stakeholder resistance may hinder organizational learning and quality development, but its effects appear to depend substantially on stakeholder involvement, communication, leadership, and effective change management [30]. Importantly, leadership emerged inductively as an additional cross-cutting enabling factor. Participants emphasized its role in translating accreditation recommendations into organizational learning and institutional change, consistent with Senge [31].

Several limitations should be acknowledged. The small purposive sample of four experienced QA and accreditation experts provided information-rich, context-specific insights but does not allow statistical generalization, and transferability should be considered within institutional and national contexts. The findings are perception-based and cover accreditation systems with different objectives and standards. Future research should test the model on larger samples using quantitative or mixed methods, compare nationally and internationally accredited institutions, and employ longitudinal designs to distinguish short-term accreditation effects from the long-term institutionalization of organizational learning and quality culture.

References

[1] J. Cullen, J. Joyce, T. Hassall and M. Broadbent, "Quality in higher education: From monitoring to management," *Quality Assurance in Education*, vol. 11, no. 1, pp. 5–14, 2003. doi: 10.1108/09684880310462038.

[2] D. Houston, "Rethinking quality and improvement in higher education," *Quality Assurance in Education*, vol. 16, no. 1, pp. 61–79, 2008. doi: 10.1108/09684880810848413.

[3] P. Kumar, B. Shukla and D. Passey, "Impact of accreditation on quality and excellence of higher education institutions," *Investigación Operacional*, vol. 41, no. 2, pp. 151–167, 2021. [Online]. Available: https://www.researchgate.net/publication/350108789_Impact_of_Accreditation_on_Quality_and_Excellence_of_Higher_Education_Institutions

[4] J. J. Lawrence and B. Dangerfield, "Integrating professional reaccreditation and quality award processes," *Quality Assurance in Education*, vol. 9, no. 2, pp. 80–91, 2001. doi: 10.1108/09684880110389654.

[5] AZVO, "Pravilnik o provedbi postupka reakreditacije visokih učilišta," 2023. [Online]. Available: https://www.azvo.hr/wp-azvo-files/uploads/2024/05/Pravilnik-o-provedbi-postupka-reakreditacije.pdf

[6] AZVO, "Reakreditacija visokih učilišta (ESG)," 2023. [Online]. Available: https://www.azvo.hr/vrednovanja/reakreditacija/reakreditacija-visokih-ucilista/

[7] G. J. Alarcón Parra, V. R. Ojeda Silva, G. G. Logroño Alarcón, F. F. Navarrete Chávez and G. G. Merino Naranjo, "Quality management in the university context and its relation to accreditations," *Revista de Gestão e Secretariado*, vol. 16, 2025. doi: 10.7769/gesec.v16i5.4557.

[8] F. A. Puerta-Guardo, A. S. Cantillo-Orozco and J. L. Castillo-Loaiza, "Quality in higher education institutions as a transversal tool in institutional accreditation: A bibliometric review," *European Journal of Educational Research*, vol. 15, no. 1, p. 19, 2026. doi: 10.12973/eu-jer.15.1.19.

[9] R. Telford and R. Masson, "The congruence of quality values in higher education," *Quality Assurance in Education*, vol. 13, no. 2, pp. 107–119, 2005. doi: 10.1108/09684880510594364.

[10] C. Temponi, "Continuous improvement framework: Implications for academia," *Quality Assurance in Education*, vol. 13, no. 1, pp. 17–36, 2005. doi: 10.1108/09684880510578632.

[11] G. P. Huber, "Organizational learning: The contributing processes and the literatures," *Organization Science*, vol. 2, no. 1, pp. 88–115, 1991. doi: 10.1287/orsc.2.1.88.

[12] C. Argyris and D. A. Schön, *Organizational Learning II: Theory, Method, and Practice*. Addison-Wesley, 1996. ISBN: 9780201629835.

[13] M. M. Crossan, H. W. Lane and R. E. White, "An organizational learning framework: From intuition to institution," *Academy of Management Review*, vol. 24, no. 3, pp. 522–537, 1999. doi: 10.5465/AMR.1999.2202135.

[14] J. J. Tari, "Self-assessment processes: The importance of follow-up for success," *Quality Assurance in Education*, vol. 18, no. 1, pp. 19–33, 2010. doi: 10.1108/09684881011015972.

[15] K. O'Mahony and T. N. Garavan, "Implementing a quality management framework in a higher education organisation: A case study," *Quality Assurance in Education*, vol. 20, no. 2, pp. 184–200, 2012. doi: 10.1108/09684881211219767.

[16] D. C. S. Law, "Quality assurance in post-secondary education: Some common approaches," *Quality Assurance in Education*, vol. 18, no. 1, pp. 64–77, 2010. doi: 10.1108/09684881011016007.

[17] H. Thomas and E. Cornuel, "Business school futures: Evaluation and perspectives," *Journal of Management Development*, vol. 30, no. 5, pp. 444–450, 2011. doi: 10.1108/02621711111132957.

[18] J. F. McKenna, C. C. Cotton and S. Van Auken, "The new AACSB accreditation standards: A prospect of tiering?," *Journal of Organizational Change Management*, vol. 10, no. 6, pp. 491–502, 1997. doi: 10.1108/09534819710190166.

[19] AACSB International, "State of Accreditation Report, Business Accreditation as of June 30, 2025," 2025. [Online]. Available: https://www.aacsb.edu/insights/reports/2025/2025-state-of-accreditation-report

[20] AACSB International, "AACSB Global Standards for Business Education," 2026. [Online]. Available: https://www.aacsb.edu/educators/global-standards

[21] M. Cadart, "AACSB 2020 → 2026: What Really Changed in the Standards?," *Headway Advisory*, 2026. [Online]. Available: https://headway-advisory.com/aacsb-2020-%E2%86%92-2026-what-really-changed-in-the-standards/.

[22] O. Serrano, F. Miranda Gonzalez and J. Mourato, "Impact of quality assurance on higher education institution performance: Differences between public and private institutions," *Societies*, vol. 16, no. 3, p. 76, 2026. doi: 10.3390/soc16030076.

[23] F. A. Koslowski, "Quality and assessment in context: A brief review," *Quality Assurance in Education*, vol. 14, no. 3, pp. 277–288, 2006. doi: 10.1108/09684880610678586.

[24] H. Thomas, "Business school strategy and the metrics for success," *Journal of Management Development*, vol. 26, no. 1, 2007. doi: 10.1108/02621710710720068.

[25] R. Ardi, A. Hidayatno and T. Y. M. Zagloel, "Investigating relationships among quality dimensions in higher education," *Quality Assurance in Education*, vol. 20, no. 4, pp. 408–428, 2012. doi: 10.1108/09684881211264028.

[26] M. Hodgkinson and M. Kelly, "Quality management and enhancement processes in UK business schools: A review," *Quality Assurance in Education*, vol. 15, no. 1, pp. 77–91, 2007. doi: 10.1108/09684880710723043.

[27] R. L. Calingo, "The evolution of strategic quality management," *International Journal of Quality & Reliability Management*, vol. 13, no. 9, pp. 19–37, 1996. doi: 10.1108/02656719610150597.

[28] J. Kettunen, "A conceptual framework to help evaluate the quality of institutional performance," *Quality Assurance in Education*, vol. 16, no. 4, pp. 322–332, 2008. doi: 10.1108/09684880810906472.

[29] P. Kotler and K. F. A. Fox, *Strategic Marketing for Educational Institutions*, 2nd ed. Prentice Hall, 1995. ISBN: 978-0136689898.

[30] E. Mainardes, H. Alves and M. Raposo, "A model for stakeholder classification and stakeholder relationships," *Management Decision*, vol. 50, no. 10, pp. 1861–1879, 2012. doi: 10.1108/00251741211279648.

[31] P. M. Senge, *The Fifth Discipline: The Art and Practice of the Learning Organization*, rev. ed. Doubleday, 2006. ISBN: 9781905211203.

[32] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006. doi: 10.1191/1478088706qp063oa.

[33] V. Braun and V. Clarke, *Thematic Analysis: A Practical Guide*. Sage, 2021. ISBN: 978-1473953246.

[34] V. Braun and V. Clarke, "Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches," *Counselling and Psychotherapy Research*, vol. 21, no. 1, pp. 37–47, 2021. doi: 10.1002/capr.12360.

THE USE OF AI TOOLS

The authors used ChatGPT (OpenAI) for language editing and translation of selected parts of the manuscript, and to assist in the preparation and visualization of the conceptual research model. All content was reviewed and verified by the authors.

The Impact of Artificial Intelligence on the Economic Professions

But Dedaj

University of Prishtina, Kosovo

Abstract

The study explores the transformative impact of artificial intelligence (AI) on the economic professions, with particular attention to its effects on labour markets, professional skills, and education systems. The emergence and rapid diffusion of artificial intelligence (AI) as a general purpose technology are fundamentally reshaping the organization of work, decision making processes, and the generation of economic value across a broad range of economic activities. Its rapid development and widespread adoption are also creating uncertainty about the future of employment and the types of knowledge and competencies that professionals will need. In this context, preparing the workforce for continuously evolving job requirements has become increasingly challenging, as technological innovation is reshaping occupations and professional tasks at an unprecedented pace. The study therefore emphasizes the need for education and training systems to respond more effectively to these changes. Rather than focusing solely on specialized and technical knowledge, traditional education should place greater emphasis on adaptability, critical thinking, problem solving, creativity, digital competencies, and lifelong learning. These capabilities can help economic professionals and future workers remain resilient, adaptable, and competitive in an increasingly AI driven labour market.

Keywords: *Artificial Intelligence (AI); Economic Professions; Labour Markets; Future of Work; Professional Skills; Education and Lifelong Learning; Digital Transformation*

Beyond Screen Time: Smartphone Salience as a Cognitive Predictor of Problematic Smartphone Use

Matej Nakić[1], Ivan David Dogan[2] and Goran Oblaković[3]

[1]Zagreb School of Economics and Management, Zagreb, Croatia, mnakic@zsem.hr
[2]Zagreb School of Economics and Management, Zagreb, Croatia, iddogan@zsem.hr
[3]Luxembourg School of Business, Luxembourg, goran.oblakovic@luxsb.lu

Abstract

Research on problematic smartphone use has predominantly focused on screen time and usage frequency, implicitly treating digital engagement as a quantitative phenomenon. Adopting a cognitive perspective, the present study examines smartphone salience—the extent to which smartphones remain cognitively prominent and mentally accessible—as a proximal predictor of problematic smartphone use beyond usage frequency and individual differences. In a cross-sectional sample of 198 adult university students, we assessed smartphone salience, problematic smartphone use, perceived social relatedness, usage frequency, and demographic variables. Hierarchical regression analyses indicate that salience is positively associated with problematic smartphone use after controlling for age, gender, and usage frequency, and that it explains additional variance beyond usage frequency alone. Contrary to common assumptions, neither perceived social relatedness nor age moderated this association. These findings suggest that, alongside established predictors, smartphone salience represents an additional cognitive aspect that warrants consideration in research on problematic smartphone use, with implications for digital well-being in organizations and higher education.

Keywords: *smartphone salience, problematic smartphone use, attention, digital well-being, cognitive preoccupation, higher education*

ISSN 2991-9371

1. Introduction

Digital technologies have become deeply embedded in contemporary work and everyday life [1]. Smartphones, in particular, function as continuous access points to communication, information, and organizational systems. While such technologies enable flexibility and responsiveness, they also introduce new challenges for the sustainable allocation of attention [2]. Increasingly, concerns about digital well-being in organizations are framed not in terms of access or availability, but in terms of persistent cognitive engagement that interferes with task focus, recovery, and long-term performance [3], [4]. These dynamics are especially consequential in higher-education settings, where sustained attention is central to learning and academic performance and where students are among the most intensive smartphone users. Importantly, these concerns are not limited to younger individuals, as smartphone adoption and everyday use are now widespread across adult age groups [5].

Much of the existing discourse implicitly equates problematic technology use with the amount of time spent using digital devices [6]. However, this perspective risks oversimplifying the phenomenon by treating digital engagement as a purely quantitative issue. Psychological accounts increasingly conceptualize problematic technology use not as excessive engagement per se, but as a compensatory behavior aimed at regulating unmet psychological needs, such as social connectedness or belonging [7], [8]. From a self-determination theory perspective, such compensatory tendencies can be understood as reflecting frustration of the basic psychological need for relatedness—the need to feel socially connected, valued, and meaningfully engaged with others [9]. However, at the organizational level, sustainability concerns arise not merely when technologies are used frequently or serve compensatory functions, but when they become cognitively dominant or salient—persistently capturing attention in ways that undermine focused work and self-regulation [10]. In this context, smartphone salience refers to the extent to which smartphones remain cognitively prominent and mentally accessible, capturing attention even in the absence of active use [10], [11]. Crucially, salience is conceptually distinct both from the behavioral compulsion emphasized in addiction and dependence accounts and from problematic use itself: rather than an index of impairment, it represents a proximal cognitive antecedent—the mental prominence of the device—that may give rise to problematic engagement. This distinguishes salience from adjacent constructs such as habitual checking, cognitive preoccupation, or compulsive use, which either describe overt behavior or already presuppose dysfunction.

Emerging discussions within the attention economy suggest that digital technologies increasingly compete for cognitive resources not only through time demands, but through their mental availability [10]. In an economy in which human attention is the scarce and monetized resource, smartphones and their applications are deliberately engineered to maximize engagement, and their most durable effect may be to remain mentally present rather than merely to accumulate usage time. Smartphones may therefore remain salient even when not actively used, drawing attention through anticipation and habitual checking behaviors that operate largely automatically [12]. Such cognitive salience represents a qualitatively different form of engagement than frequency of use, with potentially stronger implications for attentional sustainability and self-regulation in work and study contexts [11]. From an organizational perspective, distinguishing

between frequency and salience is critical: two employees may use their smartphones equally often, yet differ substantially in how much the device occupies their attention during task execution. If problematic outcomes are driven primarily by salience rather than frequency, interventions focused solely on reducing screen time may fail to address the core mechanism underlying impaired focus and recovery.

Despite growing interest in digital well-being, little is known about whether cognitive preoccupation with smartphones—beyond sheer usage frequency or psychological overcompensation—functions as a proximal mechanism of problematic smartphone use. Two boundary conditions are theoretically salient. First, if salience partly reflects a compensatory response to unmet relatedness [7], [9], its association with problematic use should be stronger among individuals whose need for relatedness is frustrated and weaker among those who feel socially connected; perceived social relatedness might therefore attenuate the salience–problematic use link. Second, common assumptions about heightened vulnerability among younger, "digitally native" users imply that age should moderate this association, yet empirical evidence remains limited and inconsistent [13], [14]. Addressing these gaps, the present study examines smartphone salience as a central cognitive mechanism underlying problematic smartphone use, while considering perceived social relatedness and age as boundary conditions across adult age groups.

Accordingly, we tested four hypotheses. Smartphone salience is positively associated with problematic smartphone use after controlling for demographics and usage frequency (H1). Salience explains incremental variance in problematic smartphone use beyond usage frequency (H2). Perceived social relatedness moderates (attenuates) the salience–problematic use association (H3). Age moderates this association (H4).

2. Methods

2.1 Participants and Procedure

A total of 198 adult university students participated in the study. Participants were enrolled in undergraduate (BA) and graduate (MA) programs at universities in Zagreb (Croatia) and Luxembourg. The sample ranged in age from 18 to 51 years (M = 23.67, SD = 5.38) and was gender-balanced, with 97 participants identifying as male (49%) and 99 as female (50%); 2 participants (1%) did not report their gender. All participants owned a smartphone and completed an online questionnaire after providing informed consent. Participation was voluntary and anonymous, and the study protocol was approved by the ethics committee of Zagreb School of Economics and Management.

Of the 198 participants, 16 had incomplete data on one or more analysis variables (primarily age, n = 11) and were excluded via listwise deletion, yielding an analytic sample of N = 182 for all regression models.

2.2 Measures

Mobile Phone Salience (MPS-7). Smartphone salience was assessed using a 7-item Mobile Phone Salience Index (MPS-7), developed for the present study to capture the cognitive and behavioral prominence of the smartphone in everyday life (see Appendix). Items were rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), with higher scores indicating greater salience. Example items include "I regularly check my phone even if it does not ring" and "My phone is within my reach even at night." Prior to scale construction, factorability diagnostics supported adequacy for exploratory factor analysis (KMO =.75; Bartlett's test $\chi^2(45) = 589.36$, $p < .001$). Exploratory factor analysis (principal axis factoring) was conducted on the initial 10 items. In the one-factor solution, two items (items 9–10) failed to load above.40 (h^2 =.15 and.13, respectively), and a third item (item 8) displayed a weak communality (h^2 =.24). In a two-factor solution, items 9 and 10 loaded strongly on a separate factor (.79 and.75), whereas item 8 did not load above.40 on either factor. Substantively, items 9 and 10 tapped a preference for mediated over face-to-face communication (e.g., "I prefer settling a dispute on the phone rather than face to face") rather than the cognitive prominence of the device, and item 8 captured interpersonal worry rather than device salience. All three were therefore excluded, and the remaining seven items formed a coherent single factor with acceptable internal consistency (Cronbach's α =.78). A composite salience score was computed as the mean of the retained items.

Problematic Smartphone Use (PSU). Problematic smartphone use was assessed using a 10-item self-report scale designed to capture functional impairment and maladaptive patterns of smartphone engagement in everyday contexts (see Appendix). Although several validated scales for problematic smartphone use exist (e.g., the Smartphone Addiction Scale [15]; the Mobile Phone Problem Use Scale [16]), the present study developed a context-specific 10-item measure to ensure conceptual alignment with its focus on everyday functioning and attentional outcomes rather than clinical addiction indicators. Items were rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Example items include "I have difficulty controlling the amount of time I spend on my smartphone" and "I feel a strong compulsion to check my smartphone even when I know there is nothing urgent or important." Factorability diagnostics indicated suitability for factor analysis (KMO =.88; Bartlett's test $\chi^2(45) = 898.14$, $p < .001$). Exploratory factor analysis using principal axis factoring supported a unidimensional structure, with all items loading strongly on a single factor (loadings ≥.53), explaining 45% of the variance. The scale demonstrated excellent internal consistency (Cronbach's α =.89). A composite PSU score was computed by averaging item responses.

Perceived Social Relatedness. Perceived social relatedness was assessed using the 8-item relatedness subscale of the Basic Psychological Need Satisfaction Scale [17], capturing the extent to which participants felt socially connected, accepted, and supported by others. Items were rated on a 5-point Likert scale, with higher scores indicating greater satisfaction of the need for relatedness [9]. The scale demonstrated acceptable internal consistency (Cronbach's α =.71). A composite score was computed by averaging item responses.

Smartphone Usage Frequency. Self-reported smartphone usage frequency was assessed using a single-item measure asking participants to indicate how frequently they use their smartphone in daily life, on a 5-point Likert scale, with higher scores reflecting more frequent

use. This measure captured overall usage intensity and served to distinguish frequency of use from cognitive salience in subsequent analyses.

2.3 Analytic Strategy

Analyses were conducted in R. Descriptive statistics and zero-order correlations were first examined for all study variables. To examine the association between smartphone salience and problematic smartphone use beyond usage frequency and individual differences, a series of regression analyses with sequential entry of predictors was conducted. In baseline models, age, gender, and smartphone usage frequency were included as covariates, and smartphone salience was then added as the focal predictor. To explore potential boundary conditions, interaction terms between salience and perceived social relatedness as well as age were examined in separate models. Continuous predictors were mean-centered prior to computing interaction terms, and interaction effects were further examined using simple slope analyses where applicable.

We acknowledge partial content overlap between the MPS-7 and PSU scales: both include items referencing phone checking and discomfort when the phone is unavailable. However, the two instruments differ in their target construct. MPS-7 items capture the cognitive prominence and accessibility of the device (e.g., keeping the phone within reach, checking without a prompt), whereas PSU items capture functional impairment and loss of control (e.g., neglecting responsibilities, difficulty controlling usage time). At the empirical level, the zero-order correlation between the two composites was moderate (r =.41), indicating approximately 17% shared variance—consistent with conceptually related but distinct constructs. Furthermore, variance inflation factors in all regression models remained low (VIF ≤ 1.32), suggesting that multicollinearity did not compromise the estimates.

3. Results

Variable	M	SD	α	1	2	3	4	5
1. PSU	2.70	0.83	.89	—				
2. MPS-7	3.59	0.79	.78	.41***	—			
3. Usage frequency	4.21	0.75	—	.52***	.48***	—		
4. Relatedness	5.06	0.90	.71	−.12	−.03	−.10	—	
5. Age	23.78	5.41	—	−.10	−.05	.04	−.03	—

Table 1. Descriptive statistics, reliability coefficients, and zero-order correlations (N = 182). *Note:* ***p <.001.

Table 1 presents descriptive statistics, internal consistency coefficients, and zero-order correlations for all study variables. To examine whether smartphone salience was associated with problematic smartphone use beyond demographic factors and usage frequency, a regression model including salience, usage frequency, age, and gender was estimated. Smartphone salience emerged as a significant positive predictor of problematic smartphone use (B = 0.20, p =.008), controlling for all covariates. Usage frequency was also positively associated with problematic use (B = 0.49, p <.001). Age showed a marginal negative association (p =.088), whereas gender was not significantly related to the outcome. The model explained approximately 32% of the variance (R^2 =.32), supporting H1.

A hierarchical regression tested whether salience explained additional variance beyond usage frequency and demographics (Table 2). In the baseline model, usage frequency was a strong positive predictor (B = 0.59, p <.001; R^2 =.29). When salience was added, it accounted for a significant increment in explained variance (ΔR^2 =.028), F(1, 177) = 7.19, p =.008. In the full model, both usage frequency (B = 0.49, p <.001) and salience (B = 0.20, p =.008) remained significant, indicating that salience contributes uniquely to problematic smartphone use beyond frequency of use, supporting H2. In practical terms, salience accounted for an additional 2.8 percentage points of explained variance after usage frequency and demographics were already in the model. Although modest in absolute magnitude, this increment was statistically significant and indicates that salience captures a dimension of problematic engagement not reducible to usage frequency. Figure 1 displays the standardized regression coefficients for the full model, illustrating the relative contribution of each predictor: usage frequency showed the largest standardized effect (β =.43), followed by smartphone salience (β =.19).

To test whether perceived social relatedness moderated the salience–problematic use association, a moderation model including salience, relatedness, their interaction, and covariates was estimated. Salience remained a significant positive predictor (B = 0.20, p =.007), whereas relatedness was not significantly associated with the outcome (p =.21). The salience × relatedness interaction was not significant (B = −0.02, p =.79), so H3 was not supported. To test age as a moderator, a further model was estimated; salience again predicted the outcome (B = 0.20, p =.008), and the salience × age interaction was not significant (B = −0.004, p =.75), so H4 was not supported. Across all models, salience consistently predicted problematic smartphone use independently of usage frequency and demographics (Table 3).

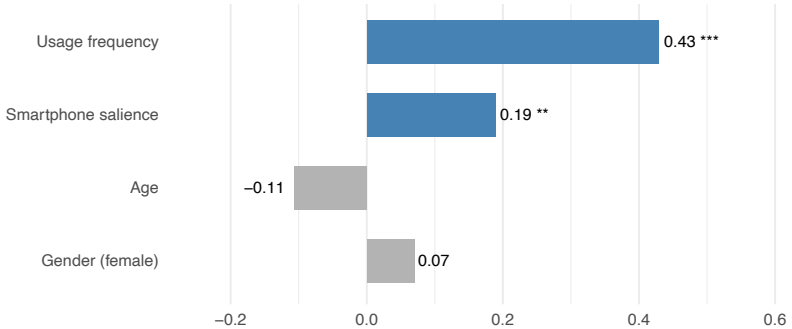

Figure 1. Standardized regression coefficients (β) for the full model predicting problematic smartphone use (N = 182). Blue bars indicate significant predictors. **p <.01, ***p <.001.

Predictor	Step 1: B	Step 1: p	Step 2: B	Step 2: p
Smartphone salience	—	—	0.20	.008
Smartphone usage frequency	0.59	<.001	0.49	<.001
Age	−0.02	.059	−0.02	.088
Gender (female vs. male)	0.07	.530	0.06	.574
R^2	.289		.317	
ΔR^2	—		.028	.008

Table 2. Hierarchical regression predicting problematic smartphone use (N = 182). *Note.* Step 1 R^2 =.289, F(3, 178) = 24.15, p <.001; Step 2 R^2 =.317, F(4, 177) = 20.54, p <.001; ΔF(1, 177) = 7.19, p =.008.

Hypothesis	Description	Result	Key statistics
H1	Salience predicts PSU	Supported	B =.20, p =.008
H2	Salience adds beyond frequency	Supported	ΔR^2 =.028, p =.008
H3	Relatedness moderates salience–PSU	Not supported	B = −.02, p =.79
H4	Age moderates salience–PSU	Not supported	B = −.004, p =.75

Table 3. Summary of hypothesis testing.

4. Discussion

The present study examined smartphone salience as a cognitive mechanism underlying problematic smartphone use, beyond usage frequency and individual differences. Across analyses, salience emerged as a consistent positive predictor of problematic smartphone use, even when controlling for age, gender, and usage frequency. When perceived social relatedness was additionally included as a covariate in the moderation model, salience remained a significant predictor (B = 0.20, p =.007). Problematic engagement with smartphones is thus not solely a function of how often devices are used, but also of the degree to which they remain cognitively salient and mentally accessible.

Importantly, salience explained additional variance beyond self-reported usage frequency, supporting a theoretical distinction between quantitative indicators of digital engagement and qualitative forms of cognitive involvement. Consistent with work emphasizing the cognitive costs of smartphone salience and mental availability [10], these findings position salience as

a proximal cognitive antecedent of problematic use rather than a mere symptom of it. We emphasize, however, that the cross-sectional design cannot establish this ordering: salience may precede problematic use, constitute one facet of it, or emerge as a consequence, and disentangling these possibilities is an important task for longitudinal research.

Contrary to the compensatory-use rationale, perceived social relatedness did not moderate the salience–problematic use association. Although we reasoned that salience would matter less for individuals whose relatedness needs are met, its effect appeared to operate largely independently of perceived social connectedness. Similarly, and against the "digital native" assumption, age did not moderate the association. Together, these null findings challenge the view that problematic smartphone engagement is primarily driven by social deficits or disproportionately affects younger users, echoing evidence that age-related differences in problematic use are often modest in adult samples [13].

4.1 Implications

These findings carry practical implications wherever sustained attention is at a premium. Because problematic use was associated with cognitive salience rather than usage frequency alone, interventions that target only screen time may miss the underlying mechanism. More promising are measures that reduce the mental prominence of the device—for example, norms that limit notifications, removing the phone from the visual field during focused work (in our data, keeping the phone "within reach" is a core salience marker), and training in attentional self-regulation. In organizations, this speaks to the design of recovery periods and interruption norms. In higher education specifically, from which our sample was drawn, the same logic suggests concrete steps for educators: designing learning environments and classroom policies that lower device salience, and helping students develop metacognitive strategies to manage the pull of the smartphone during study and lectures. Framing digital well-being around salience rather than screen time may thus be more actionable for both employers and educators.

4.2 Limitations and Future Directions

Several limitations qualify these conclusions. First, the cross-sectional design precludes causal inference and leaves the temporal ordering of salience and problematic use unresolved. Second, all measures were self-report, raising the possibility of common-method variance. Because both focal constructs were assessed with newly developed, context-specific indices, some content overlap between salience and problematic-use items is acknowledged (r =.41); future work should establish their discriminant validity against established problematic-use and preoccupation measures, ideally with confirmatory factor analysis in independent samples. In particular, a formal test of whether salience is redundant with the cognitive-preoccupation items of problematic-use scales—which the present data did not allow—would strengthen the construct claim. Third, usage frequency was captured with a single, coarse self-report item; objective usage logs and experience-sampling would provide finer-grained and less biased estimates. Fourth, the sample consisted of university students; although this aligns the study

with higher-education concerns and spans a meaningful adult age range, replication in broader occupational populations is needed. Finally, future research could incorporate additional contextual factors—work demands, personality, and technology-design features—that may shape the development and consequences of smartphone salience.

5. Conclusion

Moving beyond screen time, this study identifies smartphone salience—the cognitive prominence and mental accessibility of the device—as a predictor of problematic smartphone use that contributes uniquely beyond usage frequency. We found no evidence that this association varied as a function of perceived relatedness or age, suggesting that cognitive salience may represent a general vulnerability rather than one confined to socially disconnected or younger individuals.

For researchers, these findings underscore the need to complement frequency-based measures with indicators of cognitive involvement when studying digital well-being. For educators in higher education, the practical implication is that policies targeting screen time alone—such as phone bans—may be insufficient; more effective approaches may involve reducing the mental prominence of the device through environmental design (e.g., phone-free zones, notification management) and fostering students' metacognitive awareness of attentional capture. For organizations more broadly, the results suggest that digital well-being interventions should attend not only to how much employees use their devices, but to how persistently those devices occupy their attention during focused work and recovery.

Future research should examine these dynamics longitudinally and in more diverse populations, and establish the discriminant validity of smartphone salience against related constructs using confirmatory approaches.

References

[1] V. Cirillo, L. Fanti, A. Mina and A. Ricci, "The adoption of digital technologies: Investment, skills, work organisation", Structural Change and Economic Dynamics, 66, pg. 89-105, 2023.

[2] J. Skowronek, A. Seifert and S. Lindberg, "The mere presence of a smartphone reduces basal attentional performance", Scientific Reports, 13(1), 9363, 2023.

[3] J. D. Upshaw, C. E. Stevens Jr, G. Ganis and D. L. Zabelina, "The hidden cost of a smartphone: The effects of smartphone notifications on cognitive control", PLoS One, 17(11), e0277220, 2022.

[4] S. Ohly and L. Bastin, "Effects of task interruptions caused by notifications from communication applications on strain and performance", Journal of Occupational Health, 65(1), e12408, 2023.

[5] A. Perrin, "Mobile technology and home broadband 2021", Pew Research Center, 559, 2021.

[6] R. M. S. Santos et al., "The associations between screen time and mental health in adults: A systematic review", Journal of Technology in Behavioral Science, 9(4), pg. 825-845, 2024.

[7] D. Kardefelt-Winther, "A conceptual and methodological critique of internet addiction research: Towards a model of compensatory internet use", Computers in Human Behavior, 31, pg. 351-354, 2014.

[8] J. D. Elhai, R. D. Dvorak, J. C. Levine and B. J. Hall, "Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology", Journal of Affective Disorders, 207, pg. 251-259, 2017.

[9] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions", Contemporary Educational Psychology, 25(1), pg. 54-67, 2000.

[10] A. F. Ward, K. Duke, A. Gneezy and M. W. Bos, "Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity", Journal of the Association for Consumer Research, 2(2), pg. 140-154, 2017.

[11] A. Hadar et al., "Answering the missed call: Initial exploration of cognitive and electrophysiological changes associated with smartphone use and abuse", PLoS One, 12(7), e0180094, 2017.

[12] A. Oulasvirta, T. Rattenbury, L. Ma and E. Raita, "Habits make smartphone use more pervasive", Personal and Ubiquitous Computing, 16(1), pg. 105-114, 2012.

[13] P. A. Busch and S. McCarthy, "Antecedents and consequences of problematic smartphone use: A systematic literature review", Computers in Human Behavior, 114, 106414, 2021.

[14] M. Nahas, S. Hlais, C. Saberian and J. Antoun, "Problematic smartphone use among Lebanese adults aged 18-65 years using MPPUS-10", Computers in Human Behavior, 87, pg. 348-353, 2018.

[15] M. Kwon, D. J. Kim, H. Cho and S. Yang, "The smartphone addiction scale: Development and validation of a short version for adolescents", PLoS One, 8(12), e83558, 2013.

[16] A. Bianchi and J. G. Phillips, "Psychological predictors of problem mobile phone use", Cyberpsychology & Behavior, 8(1), pg. 39-51, 2005.

[17] R. M. Ryan and E. L. Deci, "The darker and brighter sides of human existence: Basic psychological needs as a unifying concept", Psychological Inquiry, 11(4), pg. 319-338, 2000.

6. Appendix. Scale Items

Problematic Smartphone Use (PSU) items

1. I feel anxious or distressed when I am unable to use my smartphone.
2. I prioritize using my smartphone over other important activities in my life.
3. I have difficulty controlling the amount of time I spend on my smartphone.
4. I neglect responsibilities, such as work or school, due to my smartphone use.
5. I experience withdrawal symptoms, such as irritability or mood swings, when I am unable to use my smartphone.
6. I find myself using my smartphone excessively, even when it is not necessary or appropriate.

7. My smartphone use interferes with my ability to concentrate or focus on other tasks.
8. I have strained relationships with others due to my excessive attachment to my smartphone.
9. I feel a strong compulsion to check my smartphone even when I know there is nothing urgent or important.
10. I experience distress or negative emotions when I am not able to use my smartphone for an extended period.

Mobile Phone Salience Index (MPS-7) items — retained items in bold

1. **I regularly check my phone even if it does not ring.**
2. **I feel bad when I leave my phone at home / when it runs out of battery.**
3. **My phone is within my reach even at night.**
4. **If I left my phone at home, I would be willing to go home for it.**
5. **When I sit down somewhere, I put my phone at a visible place, within my reach.**
6. **I'm nervous if I cannot be reached on the phone.**
7. **I'm nervous if I cannot get through to somebody immediately.**
8. If a close friend / family member doesn't pick up the phone, I start to worry.
9. I prefer talking about awkward things on the phone rather than face to face.
10. I prefer settling a dispute on the phone rather than face to face.

AI, Future Skills and Business Education

SESSION CHAIR:

Karmela Aleksić-Maslać

What Knowledge-Intensive Entrepreneurs Still Need to Learn: Skill Gaps in the Age of AI

Vitaliy Roud

Zagreb School of Economics and Management, Zagreb, Croatia, vitaliy.roud@zsem.hr

Abstract

Founders in an advanced Croatian technology incubator rate sales and marketing as the gaps that most limit their growth, while close to half call technical and artificial-intelligence skills irrelevant, and most access AI through shared platform interfaces rather than building it in-house. We ask whether the skill gaps that constrain knowledge-intensive entrepreneurial firms shift with the firm's AI strategy, and what the answer means for the future of business education. From a survey of 56 startups at the Zagreb Innovation Centre, we group firms by their digital and AI strategy and relate skill-gap severity to how deeply firms engage AI. Commercial gaps dominate every profile and, far from narrowing as firms commit to AI, they deepen with it: the firms that build their own AI report the largest business-development gap, and the more a firm prioritises AI the larger its own data and AI gap. AI does not close these gaps; engaging it more deeply raises the demand for the commercial and business-model skills they describe. Because advantage in the age of AI accrues to firms whose business models and processes are organised around knowledge, the competence these founders most lack is the design of a knowledge-centric business model that uses cheap, widely available AI. Teaching that competence, in degree programmes and in the continuing education of working founders, is the direction in which business education must develop.

Keywords: *knowledge-intensive entrepreneurship, skill gaps, artificial intelligence strategy, latent class analysis, business education, incubator.*

ISSN 2991-9371

1. Introduction

Knowledge-intensive entrepreneurial firms turn advanced scientific and technical knowledge into new products and markets, and the economies that climb the technology ladder build ecosystems to grow them [1], [2]. What limits these firms is a first-order policy question, and access to skills, competence and knowledge sits at its centre [3]. For one advanced population the answer is direct. Inside a leading incubator the technical foundation is already strong and cheaply available, so the binding constraint is commercial competence, and a new general-purpose technology is sharpening that constraint rather than dissolving it.

Artificial intelligence is spreading through entrepreneurship the way earlier general-purpose technologies spread through the wider economy: cheap to access, improving fast, and shared across competitors [4], [5], [6]. When the same capability is cheap and available to every firm, owning the technology ceases to confer advantage, and the business model that puts it to work becomes the differentiator. Whether AI closes the competence gaps founders face or opens new ones remains unsettled, and it is an active question for research [7].

We study 56 knowledge-intensive startups hosted by the Zagreb Innovation Centre (ZICER), a Croatian incubator, and ask a question with two parts: do the skill gaps that most limit growth shift with the firm's AI strategy, and what does the resulting pattern mean for the future of business education? The contribution is to cross two dimensions the descriptive evidence has so far kept apart. We group firms by their AI strategy, test how skill-gap severity differs across those groups, and read the result as a question of what business education must teach. The competence the data single out is the design of a knowledge-centric business model on commodity AI, which falls within the established remit of business education.

2. Knowledge-intensive entrepreneurship and the skill-gap question

2.1 Knowledge-intensive entrepreneurial firms

Knowledge-intensive entrepreneurship denotes new ventures that draw on, and add to, advanced bodies of knowledge, joining Schumpeterian novelty with the cumulative learning that evolutionary economics and innovation-systems research describe [1]. These firms matter beyond their own balance sheets: they diffuse knowledge, attract complementary investment, and pull regional ecosystems toward higher-value activity [8]. In Central and Eastern Europe the orientation of such firms shapes how fast economies catch up technologically [2]. Incubators concentrate this population, which makes them a useful site for studying the constraints that bind it. University entrepreneurial ecosystems, in particular, depend on the presence and growth of knowledge-intensive firms [9].

Knowledge intensity describes how a firm competes, not the sector it sits in. A knowledge-centric venture builds its advantage on what it knows and on how its model and processes turn that knowledge into value, rather than on cost, scale, or a protected market [1], [3]. The assets that carry the advantage are intangible: skills, routines, data and relationships that are hard to buy and hard to copy [10]. The policy interest follows. Economies that want to move up the value chain need more firms that compete on knowledge, and they need the institutions that build the human capital behind them [2], [11]. Which part of that human capital is now scarce, and who can supply it, is the question this study pursues.

2.2 Where the gaps sit: commercial, not only technical

The competence a knowledge-intensive firm lacks is rarely a single thing. Performance in these firms depends on a bundle of capabilities, and managerial and commercial competences gate growth at least as tightly as technical depth [3]. There is more than one route to high performance, and firms reach it through different combinations of capability, which is why the gaps a firm reports are heterogeneous rather than uniform [12]. The composition of the founding team explains part of that heterogeneity: variety in founder experience shapes which functions a firm can staff from within and which it must learn [13]. A firm's existing knowledge base also sets the ceiling on what new competence it can absorb, so a technically strong team does not automatically become commercially capable [10].

Two kinds of competence sit behind these gaps, and they are not symmetric. Technical competence, meaning engineering, product and data skills, is what founders in a technology incubator usually bring with them, because it is the reason the venture exists. Commercial competence is learned later, if at all. It covers the work of finding customers, pricing, positioning, financing, and assembling the business model that connects them, and it is the competence that converts a working technology into a growing firm [3], [12]. In a population selected for technical strength, the commercial side is where the binding constraint is most likely to sit. The AI literature gives a reason to expect that constraint to tighten rather than ease.

2.3 AI as a general-purpose technology: closing gaps or opening them

General-purpose technologies create value only when user firms invent the complementary practices that put them to work [4]. AI fits this pattern, and recent work treats it as a method of invention whose returns depend on recombination inside the adopting firm rather than on the technology itself [5]. The economics of AI sharpen the point: as machine prediction becomes cheap, the scarce and valuable inputs become judgement, problem framing, and business-model design [6]. For small and medium firms, the depth of AI adoption interacts with the firm's knowledge base to determine which constraints actually bind [14]. Startups already reach for generative AI inside commercial functions such as growth and marketing, not only inside engineering [15]. The mechanism predicts a shift rather than a closure. AI lowers the technical and data barrier and raises the value of the commercial competence around it.

A second consequence concerns the distribution of advantage. When a general-purpose technology becomes cheap and shared, the firms that gain the most are those whose business model and processes are organised around knowledge, because the technology rewards knowledge-centric designs more than knowledge-light ones [4], [5]. Cheap prediction adds the most value where a firm understands the problem deeply, holds a distinctive view of its data, and has a clear account of how the prediction creates value [6]. The returns to AI are therefore uneven. They accrue to firms organised around knowledge and widen the distance between them and firms that add AI to a knowledge-light operation [14].

The scarce competence shifts accordingly. Operating the technology is inexpensive and becoming more so; the demanding task is the design of a knowledge-centric business, a business model and set of processes that create, capture and compound knowledge with AI built into them, so that the firm's advantage grows rather than erodes. This competence is commercial and strategic, and it becomes scarcer and more valuable as AI spreads.

2.4 Research question

This logic is testable in a population that has already adopted AI to varying degrees. If AI mainly lowers technical barriers, then firms with deeper AI strategies should report smaller technical and data gaps. If the commercial gap is the binding one, it should persist across AI strategies, including among the most technically advanced firms. We test both expectations and then ask what the answer means for the future of business education. Behind both lies a larger question: if AI rewards knowledge-centric firms, what should business education teach founders to build?

3. Data and methods

We use survey evidence from 56 startups hosted by ZICER. The survey covers seven strategic dimensions of the firm. These are young, technology-oriented ventures, which makes the sample an advanced rather than a representative cross-section of small firms.

Skill gaps. Founders rated how much a gap in each of eight functions limits the firm's growth, on a four-point scale from 0 (not relevant) to 3 (severely limiting growth): technical or engineering, sales and business development, marketing and communications, finance and accounting, operations and logistics, leadership and management, product development, and data analysis or AI.

AI strategy. The survey records each firm's digital and AI strategy through four blocks: digital-transformation initiatives, the form of cooperation with AI providers, the AI hardware and infrastructure in use, and the rated importance of seven AI application areas. We summarise these into profiles using latent class analysis, an unsupervised model that assigns each firm to the strategy class it most resembles (estimated with the R package poLCA). A three-class solution

gave the clearest and most interpretable profiles; class-assignment certainty was high (entropy $R^2 = 0.99$). The model used 56 firms with complete strategy data.

Three AI-strategy profiles. We order the three classes by their AI build-depth: dedicated hardware, proprietary models, and strategic or licensing relationships with providers. The deepest profile is **AI-Native** (n = 10); the shallowest is **Minimal-AI** (n = 26); the middle profile, which reaches AI through interfaces and the cloud, is **Adopters** (n = 20). The labels follow the class profiles, not arbitrary numbering.

Tests. We join skill gaps and AI-strategy class firm by firm (n = 56). Our primary measure of AI engagement is continuous. The AI-build index counts a firm's own-AI commitments across hardware, infrastructure and provider relationships, from dedicated processors and in-house infrastructure to proprietary models and strategic or licensing partnerships (range 0–6); the AI-ambition index is the mean rated importance of the seven AI applications (0–3). We relate each of the eight gap scores to these two indices with Spearman rank correlations. As a class-level check we also compare each gap across the three AI-strategy classes with a one-way Welch analysis of variance, which does not assume equal variances across the unequal-sized classes, and a Kruskal–Wallis test confirms the same conclusion. We treat the four-point gap items as interval scores, a common simplification, and read all results as descriptive patterns: with 10 firms in the smallest class the class comparisons have little power, so we report relationships and effect sizes rather than leaning on significance, and report p-values uncorrected with the caveat that an isolated significant result would warrant caution. The sample is one incubator at one time, so the numbers are local.

4. Results

Sales and marketing are the gaps that most often limit growth (Figure 1). 61% of founders rate sales and business development as moderately or severely limiting, and 54% say the same of marketing. Technical and data skills run the other way: 34% call the technical gap not relevant and 46% call the data/AI gap not relevant. The most limiting function on average is sales / business development. The pattern matches the composition of the population: strong engineering foundations alongside limited commercial capability.

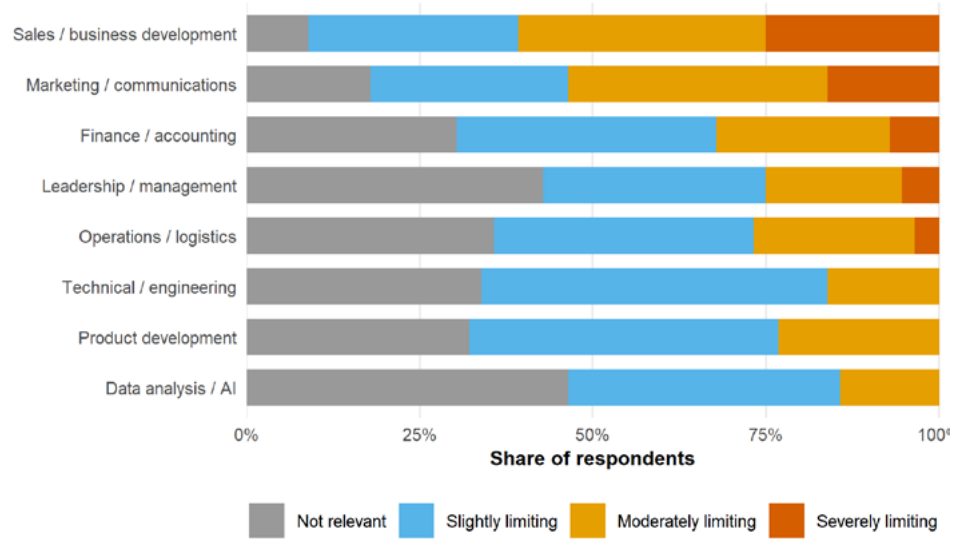

Figure 1. Impact of skill gaps on firm growth across the sample. Bars show the share of respondents at each level; commercial functions carry the most weight on the right.

Firms differ sharply in how they approach AI (Figure 2). Most reach AI through external providers: 64% via platform interfaces and 32% by buying software or services, while 46% use no specialised hardware. A minority builds in-house: the AI-Native tier concentrates the firms with their own infrastructure, proprietary models, and strategic provider relationships.

Each bar in Figure 2 shows one strategy indicator as a share of its maximum, so a longer bar means the class scores higher on that indicator, and the three shades run from Minimal-AI (light) to AI-Native (dark). Reading down the panel, the classes separate on the build items — own hardware, in-house infrastructure, proprietary models — and differ far less on whether AI matters to them. What distinguishes the classes is build depth, not enthusiasm.

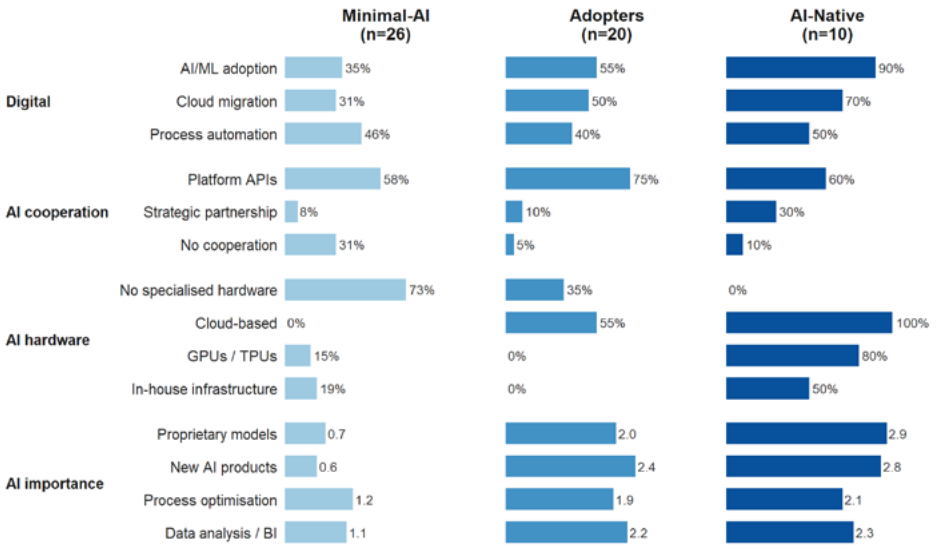

Figure 2. AI-strategy profiles from the three-class latent class model, ordered by build depth (light to dark: Minimal-AI to AI-Native). Bars show each indicator as a share of its maximum (firm share for adoption items; class mean on the 0–3 scale for AI-importance items), with the value printed at the bar end.

The deeper a firm goes into AI, the larger its business gap, not the smaller (Table 1). The AI-build index, which counts a firm's own-AI commitments from dedicated hardware to proprietary models and strategic provider partnerships, correlates positively with the sales and business-development gap (ρ = +0.28, p =.037). The firms that commit most to building AI report the greatest need for the competence to turn it into a business. AI ambition tracks the data and AI gap in the same direction: the more a firm rates AI as important across its applications, the larger its own data and AI skill gap (ρ = +0.24, p =.073). Engaging AI more deeply raises the demand for skills rather than removing it; the remaining gaps show no consistent relationship with either measure.

Table 1 reports rank correlations, so each entry asks whether firms that engage AI more deeply also report a larger gap in that function. A positive value means the gap grows with AI engagement; a negative value means it shrinks. Two entries stand out from zero, and both are positive: sales and business development against the build index, and data and AI against the ambition index.

The class view tells the same story and adds a second point: the gap is universal (Figure 3). Commercial functions carry the largest gaps in every AI-strategy class, and they are heaviest among the AI-Native builders, whose commercial-gap score (6.1 out of 12) runs above both the Adopters (4.6) and the Minimal-AI firms (5.4). No single class differs significantly on any gap (Welch ANOVA, all p >.05; a Kruskal–Wallis check agrees), which at 10 firms in the smallest class is what one expects. The commercial gap does not belong to one type of firm. It is the common condition of the population, and it is sharpest where AI commitment is deepest.

Figure 3 plots the mean severity of each gap within each AI-strategy class, on the same 0–3 scale founders used. Comparing the three bars within a row shows whether a gap belongs to one type of firm or to all of them; comparing rows shows which functions bind hardest. The commercial rows sit furthest right in all three classes, and the AI-Native bar leads them.

Skill gap	ρ with AI-build	ρ with AI-ambition
Sales / business development	+0.28*	-0.11
Marketing / communications	-0.05	-0.08
Finance / accounting	+0.04	+0.09
Leadership / management	+0.20	+0.18
Operations / logistics	-0.13	-0.02
Product development	-0.04	+0.03
Technical / engineering	-0.11	+0.03
Data analysis / AI	+0.15	+0.24†

Table 1. Spearman correlation of each skill gap with two continuous measures of AI engagement: the AI-build index (count of a firm's own-AI commitments, from dedicated hardware to proprietary models and strategic or licensing partnerships; 0–6) and the AI-ambition index (mean rated importance of seven AI applications; 0–3). A positive value means the gap is larger where AI engagement is deeper. Rows are ordered by overall severity. * p <.05, † p <.10; n = 56.

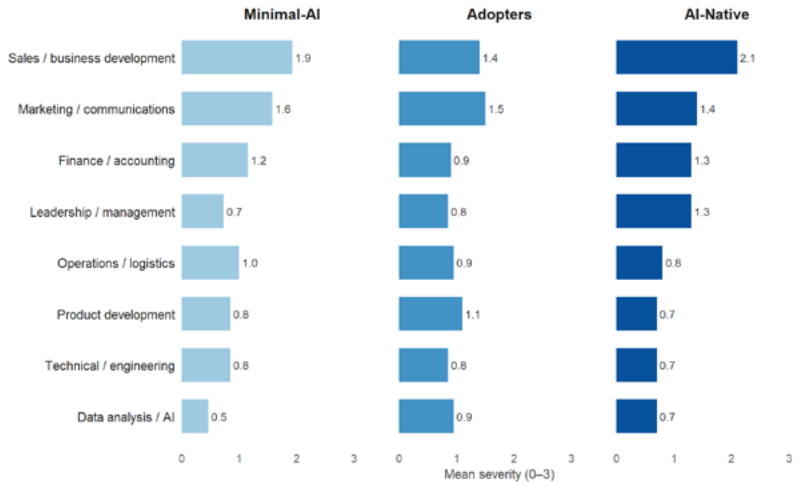

Figure 3. Mean severity of each skill gap by AI-strategy class (0 = not relevant, 3 = severely limiting). Gaps are ordered by overall severity.

The results converge on a single reading. The binding constraint in this population is commercial competence rather than technical capacity, and it binds hardest where AI runs deepest. As the technology grows cheaper and more capable, the scarce competence is the design of a competitive business, and the firms furthest into AI feel that scarcity most acutely.

5. Conclusion

The competence AI leaves scarce is the one business education exists to supply, and the firms that go furthest into AI need it most. In this sample the technical layer is inexpensive and shared: most firms reach AI through the same interfaces their competitors use, and the commercial functions carry the largest gaps in every AI-strategy class (Figure 3). The gap does not ease as firms commit to AI; it deepens. The firms that build their own AI report the largest business-development gap ($\rho = +0.28$), and the more a firm prioritises AI the larger its own data and AI gap ($\rho = +0.24$). Deeper AI engagement raises the demand for these competences rather than meeting it, which is what absorptive capacity [10] and the economics of cheap prediction [6] imply. Technical sophistication does not produce commercial sophistication.

The broader implication concerns the distribution of advantage. A cheap, shared technology does not raise every firm equally; it rewards businesses whose business models and processes are organised around knowledge and widens their lead over knowledge-light competitors [4], [5]. Advantage in the age of AI accrues to knowledge-centric firms, and it grows as more of the business comes to run on knowledge [6], [14]. The present data are consistent with this reading. The firms that engage AI most deeply report more skill needs rather than fewer, and the needs they report are commercial: how to design a business model and processes that create and compound knowledge, and how to build AI into them so the firm's advantage widens rather than erodes.

These findings set a direction for business education. If advantage flows to knowledge-centric firms, the skill worth teaching is how to make a business knowledge-centric: how to identify and build the proprietary knowledge, data and relationships that generic AI cannot supply, and how to design a business model that compounds them with AI. That curriculum is commercial and strategic, and it lies within the historical core of management and entrepreneurship education rather than alongside a technical course [16]. Three implications follow. The design of knowledge-centric, AI-empowered business models belongs in the required core of entrepreneurship and management programmes, taught as a concrete skill: how to price, position and defend a product whose visible capability is a commodity. The people who carry the gap are working founders rather than enrolled students, which places executive and continuing education at the centre of business education rather than at its margin. The incubator becomes a natural site of instruction, since the distinctive contribution of business education to firms adopting widely available AI is the business-model layer, a capability that recent work on AI and business-talent development places at the centre of its agenda [14], [17]. What these ventures most need is the competence to build a viable business on technology that is now cheap and shared [11].

The study has clear limits. It covers one Croatian incubator at one moment, with 56 firms and 10 in the smallest AI-strategy class, so the tests have low power, the effect-size intervals are wide, and a null result is uninformative rather than evidence of no difference. Latent class assignments are treated as fixed, which understates classification uncertainty. The findings are associations, not causal effects. What generalises beyond this sample is the mechanism rather than the magnitudes. Where AI is cheap and widely available, which the 64% reaching it through external interfaces here suggests is common, advantage shifts toward the knowledge a firm designs into its business, and supplying the competence to design it falls to business education.

The conclusion is straightforward. Founders in an advanced incubator are strong on technology and short on the commercial competence required to build a business with it, and AI does not remove this constraint; it raises the premium on it. The firms that advance will be the knowledge-centric ones, and their founders need to learn how to design a knowledge-centric business, which the technology itself does not teach. That competence is teachable, and teaching it is the task that now falls to business education. A natural next step is to test the pattern in other incubators and to follow founders over time, examining whether those trained to design knowledge-centric, AI-empowered business models grow faster than those who acquire the competence informally.

6. Acknowledgements

We thank the Zagreb Innovation Centre (ZICER) and the participating startups for their cooperation, and colleagues at Economic Institute Zagreb for comments on the survey design.

The author used a large language model to assist with language editing and the drafting of prose. The research design, the analysis, the interpretation of results and all conclusions are the author's own, and the author takes full responsibility for the content of the paper.

References

[1] F. Malerba and M. McKelvey, "Knowledge-intensive innovative entrepreneurship integrating Schumpeter, evolutionary economics, and innovation systems," *Small Business Economics*, vol. 54, pp. 503–522, 2020, doi: 10.1007/s11187-018-0060-2.

[2] S. Radosevic and E. Yoruk, "Entrepreneurial orientation of knowledge-based enterprises in Central and East Europe," in *Dynamics of knowledge-intensive entrepreneurship: Business strategy and public policy*, F. Malerba, Y. Caloghirou, M. McKelvey, and S. Radosevic, Eds., London: Routledge, 2016.

[3] B. Fischer, S. Salles-Filho, C. Zeitoum, and F. Colugnati, "Performance drivers in knowledge-intensive entrepreneurial firms: A multidimensional perspective," *Journal of Knowledge Management*, vol. 26, no. 5, pp. 1342–1367, 2022, doi: 10.1108/JKM-03-2021-0264.

[4] T. F. Bresnahan and M. Trajtenberg, "General purpose technologies: 'Engines of growth'?" *Journal of Econometrics*, vol. 65, no. 1, pp. 83–108, 1995, doi: 10.1016/0304-4076(94)01598-T.

[5] I. M. Cockburn, R. Henderson, and S. Stern, "The impact of artificial intelligence on innovation: An exploratory analysis," in *The economics of artificial intelligence: An agenda*, A. Agrawal, J. Gans, and A. Goldfarb, Eds., Chicago: University of Chicago Press, 2018, pp. 115–146.

[6] A. Agrawal, J. Gans, and A. Goldfarb, *Prediction machines: The simple economics of artificial intelligence*. Boston, MA: Harvard Business Review Press, 2018.

[7] M. Obschonka *et al.*, "Artificial intelligence and entrepreneurship: A call for research to prospect and establish the scholarly AI frontiers," *Entrepreneurship Theory and Practice*, vol. 49, no. 3, pp. 620–641, 2025, doi: 10.1177/10422587241304676.

[8] M. McKelvey and A. H. Lassen, *Managing knowledge intensive entrepreneurship*. Cheltenham, UK: Edward Elgar, 2013. doi: 10.4337/9781781005521.

[9] A. N. Link and R. M. Sarala, "Advancing conceptualisation of university entrepreneurial ecosystems: The role of knowledge-intensive entrepreneurial firms," *International Small Business Journal*, vol. 37, no. 3, pp. 289–310, 2019, doi: 10.1177/0266242618821720.

[10] W. M. Cohen and D. A. Levinthal, "Absorptive capacity: A new perspective on learning and innovation," *Administrative Science Quarterly*, vol. 35, no. 1, pp. 128–152, 1990, doi: 10.2307/2393553.

[11] D. Kotsopoulos, A. Karagianaki, and S. Baloutsos, "The effect of human capital, innovation capacity, and Covid-19 crisis on knowledge-intensive enterprises' growth within a VC-driven innovation ecosystem," *Journal of Business Research*, vol. 139, pp. 1177–1191, 2022, doi: 10.1016/j.jbusres.2021.10.055.

[12] G. H. S. M. de Moraes, B. Fischer, S. Sales Filho, and M. Dabić, "More than one way to get there: A configurational view on performance drivers in knowledge-intensive entrepreneurship," *Journal of Knowledge Management*, vol. 27, no. 11, pp. 206–230, 2023, doi: 10.1108/JKM-03-2023-0267.

[13] E. Gifford, G. Buenstorf, D. Ljungberg, M. McKelvey, and O. Zaring, "Variety in founder experience and the performance of knowledge-intensive innovative firms," *Journal of Evolutionary Economics*, vol. 31, pp. 677–707, 2021, doi: 10.1007/s00191-020-00692-6.

[14] E. D'Amico, D. B. Audretsch, M. Belitski, and A. Colombelli, "Adopting artificial intelligence in small and medium businesses: The knowledge-based perspective," *Small Business Economics*, 2026, doi: 10.1007/s11187-026-01214-7.

[15] A. Rezazadeh, M. Kohns, R. Bohnsack, N. António, and P. Rita, "Generative AI for growth hacking: How startups use generative AI in their growth strategies," *Journal of Business Research*, vol. 192, p. 115320, 2025, doi: 10.1016/j.jbusres.2025.115320.

[16] G. Nabi, F. Liñán, A. Fayolle, N. Krueger, and A. Walmsley, "The impact of entrepreneurship education in higher education: A systematic review and research agenda," *Academy of Management Learning & Education*, vol. 16, no. 2, pp. 277–299, 2017, doi: 10.5465/amle.2015.0026.

[17] Q. Wu, L. Chen, M. Chen, and Y. Huang, "Exploring the impact of artificial intelligence on business talent development in higher education: A systematic literature review and research agenda," *The International Journal of Management Education*, vol. 24, p. 101287, 2026, doi: 10.1016/j.ijme.2025.101287.

Artificial Intelligence in Policing Technologies, Ethical Challenges, and Implications for Democratic Governance

Krunoslav Antoliš

University of Applied Sciences in Criminal Investigation and Public Security, Ministry of Internal Affairs, Republic of Croatia, Zagreb, Croatia, kantolis@mup.hr

Abstract

The integration of artificial intelligence into policing marks a shift toward datadriven operational models. Predictive policing applies machine learning and spatial analytics to historical crime data, while computer vision supports facial recognition and licenseplate analysis. Natural language processing enables largescale monitoring of online communication, and drones expand continuous data collection. These tools improve resource allocation and automate laborintensive tasks, yet introduce substantial ethical and technical concerns. Central among them is algorithmic bias: models trained on historically skewed data risk reinforcing social disparities, reflected in higher falsepositive rates for facial recognition among minority groups. Additional issues include the opacity of blackbox systems, reduced accountability, and privacy risks associated with mass surveillance. Public opinion research shows conditional support dependent on strong oversight. The future of AIenabled policing therefore relies on ethical frameworks, rigorous auditing, and evidencebased regulation. The paper also considers how highereducation institutions can embed AIrelated policing competencies into criminal justice curricula.

Keywords: *Bias, Privacy, Accountability.*

ISSN 2991-9371

1. Introduction

The intersection of police work and artificial intelligence (AI) is an evolving and controversial topic. AI technologies are being integrated into law enforcement to improve efficiency, enhance crime prediction and prevention, and automate certain tasks. However, this also raises ethical concerns about privacy, bias, and accountability.

The following are key areas where AI is being used in policing: predictive policing, facial recognition, license plate recognition, automated reporting and documentation, social media monitoring, robotics and drones.

AI systems can analyze large datasets, such as crime reports, and use algorithms to predict where crimes are most likely to occur in the future. These predictions can help police allocate resources more effectively. For example, PredPol (Predictive Policing) is a widely known tool that analyzes crime patterns and directs officers to areas with a higher likelihood of crime [1]. Predictive policing models typically rely on two dominant approaches: placebased prediction (forecasting highrisk microlocations using spatialtemporal statistics) and personbased prediction (identifying individuals at elevated risk of offending or victimization). These models often employ regression analysis, random forests, or neural networks trained on historical crime datasets.

AI-driven facial recognition technology helps law enforcement identify suspects in real-time using surveillance footage or images. It is particularly useful for tracking missing persons or identifying criminals. However, concerns about privacy violations and the potential for wrongful arrests have sparked debates. Critics argue that facial recognition systems can be inaccurate and disproportionately target people of color [2].

AI-powered license plate recognition systems scan vehicles and identify them as they pass through checkpoints or surveillance zones. This technology is used for tracking stolen vehicles, finding wanted suspects, or monitoring traffic.

AI is being used to automate routine administrative tasks such as filing reports, processing documents, and managing case files. This reduces the time officers spend on paperwork, allowing them to focus more on policing [3].

AI can analyze social media platforms to identify potential threats, track criminal activity, or gather intelligence on individuals. By scanning online conversations, AI helps law enforcement spot emerging trends or criminal behavior [1].

AI-powered drones are used in surveillance, crowd control, and searching difficult-to-reach areas. These drones provide officers with real-time data and video feeds during operations. Robotic systems, such as bomb disposal robots, reduce risks to officers and civilians by handling dangerous tasks.

ISSN 2991-9371

This article is based on a narrative literature review drawing on key academic sources in policing, surveillance studies, and AI ethics. The selected works [1], [2], [3] were chosen due to their influence, methodological rigor, and relevance to contemporary debates on AI in law enforcement. [3] shows how data-driven surveillance "deepens discretionary power through digital infrastructures embedded in policing."

AI in policing is a teaching and learning challenge for universities that educate future police leaders, digitalforensics experts, criminologists, publicsecurity analysts, and communicationmanagement professionals. Highereducation institutions must prepare students to understand, evaluate, and ethically govern AI systems used in policing.

1.1 Methodological Approach

Although the paper draws on empirical case studies to illustrate the operational and ethical implications of AI deployment in policing, it is important to clarify the underlying methodological approach. This study is based on a narrative qualitative research design, combining document analysis, conceptual synthesis, and comparative case examination of influential works in policing, surveillance studies, and AI ethics. The empirical examples referenced throughout the paper are derived from secondary qualitative data, including peerreviewed publications, policy reports, and documented realworld deployments of AI systems in law enforcement. The methodological approach therefore relies on qualitative content analysis, through which key themes—such as algorithmic bias, transparency, accountability, and public perception—were systematically identified and compared across cases. This interpretive framework enables the paper to synthesize diverse empirical findings, highlight crosscontextual patterns, and evaluate the implications of AI technologies for democratic governance and ethical policing practices. While not based on primary fieldwork, the study employs a comparative qualitative methodology that integrates empirical evidence with theoretical analysis, providing a robust foundation for pedagogical and curricular recommendations in highereducation settings.

2. Ethical concerns and challenges

AI integration in policing presents significant ethical and operational challenges that must be addressed to ensure fairness, transparency, and accountability. [2] argues that "algorithmic systems often reproduce and amplify existing social hierarchies," raising concerns about racialized outcomes in policing. [4] critically examines "structural problems inherent in algorithmic policing systems." [5] notes that algorithmic harms can occur "even when systems appear to be designed correctly."

Bias and discrimination: AI systems trained on biased data can reinforce existing inequalities, leading to disproportionate targeting of certain communities.Example: Predictive policing algorithms have been criticized for contributing to over-policing in minority neighborhoods [2].

Transparency and accountability, many AI models operate as "black boxes," making it difficult to assess how decisions are made or challenge them when errors occur. As noted by [6], "AIdriven criminal justice systems risk undermining fundamental rights when deployed without adequate transparency, oversight, and humanrightsbased safeguards" (p. 568). Establishing clear accountability mechanisms is essential. Example: When AI-driven policing tools make mistakes, it is often unclear who should be held responsible—developers, law enforcement, or policymakers [1].

Privacy concerns, the widespread use of AI-driven surveillance tools, such as facial recognition and social media monitoring, raises concerns about mass surveillance and potential abuse. Example: Some facial recognition systems have been deployed without public consent, infringing on privacy rights [3].

Over-reliance on technology, while AI enhances efficiency, excessive dependence on automated systems may reduce human judgment in critical policing decisions. Example: Law enforcement may over-trust AI predictions without considering context, leading to flawed decision-making.

Legal and regulatory gaps, existing laws may not adequately address the ethical and operational complexities of AI in policing, underscoring the need for clearer guidelines on data usage, oversight, and accountability. [7] analyze mechanisms for "transparency and accountability in public safety algorithms." Balancing technological advancements with ethical safeguards is crucial to ensuring AI serves as a tool for justice rather than perpetuating systemic issues. These ethical challenges directly inform the central research questions that guide contemporary academic and policy debates.

3. Research questions

Here are several potential research questions based on the intersection of police work and AI, focusing on the key areas and ethical concerns we've mentioned: predictive policing and bias, ethics of facial recognition technology, accountability in AI systems, privacy concerns in AI surveillance, effectiveness and efficiency of AI in automated documentation, public perception of AI in policing, ethical regulation of AI in policing, impact of AI on police resource allocation, AI's role in combating crime vs. reinforcing inequality, the future of AI in policing, trends and challenges. [8] document how AI "reshapes policing and immigration control practices."

Research questions are:

- How can AI-based predictive policing tools be optimized to minimize racial bias and prevent disproportionate targeting of marginalized communities? *Reference*: [1]. *The rise of big data policing: Surveillance, race, and the future of law enforcement*. NYU Press.

- What are the ethical implications of using AI-driven facial recognition technology in law enforcement, and how can police departments ensure the accuracy and fairness of such systems? *Reference*: [2]. *Race after technology: Abolitionist tools for the new Jim Code*. Polity Press.

- Who should be held accountable when AI systems used in policing make mistakes, and how can transparency be ensured in the development and deployment of these algorithms? *Reference*: [3]. *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press.

- How does the automation of reporting and documentation through AI impact the quality and integrity of law enforcement data, and what are the potential risks involved? *Reference*: [1]. *The rise of big data policing: Surveillance, race, and the future of law enforcement*. NYU Press. [9] argue that AI policing in the Global South can "accelerate democratic erosion."

- How do citizens perceive the use of AI technologies in law enforcement, and what measures can be taken to address concerns about surveillance, bias, and privacy violations? *Reference*: [3]. *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press. [10] introduce new empirical methods for "quantifying racial bias in policing algorithms."

- What regulatory frameworks should be established to govern the ethical use of AI technologies in law enforcement, and how can these frameworks balance security and civil liberties? *Reference*: [2] *Race after technology: Abolitionist tools for the new Jim Code*. Polity Press.

- Does the integration of AI in policing disproportionately reinforce existing societal inequalities, and if so, what steps can be taken to ensure AI systems are used for social good? *Reference*: [1]. *The rise of big data policing: Surveillance, race, and the future of law enforcement*. NYU Press.

- How might the role of AI in policing evolve in the next decade, and what challenges will law enforcement face in implementing AI technologies while maintaining public trust? *Reference*: [3]. *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press.

Each of these questions could lead to in-depth research, examining both the technical capabilities of AI and the broader social, ethical, and legal concerns surrounding its use in law enforcement.

4. Findings & examples

Here are some key findings and insights based on the research areas related to the intersection of AI and policing. These findings explore both the technological capabilities and the ethical concerns highlighted in the initial request.

4.1 Predictive policing and bias

Findings: Studies have shown that AI-driven predictive policing tools like PredPol can be useful in directing police resources more efficiently by predicting where crimes are likely to occur. However, these systems have been criticized for perpetuating existing biases. The algorithms often rely on historical crime data, which may already reflect biased policing practices (e.g., over-policing in certain neighborhoods). This can lead to biased predictions, disproportionately targeting minority communities. Researchers emphasize the need for more diverse and representative data sets to improve the fairness of these tools [1]. **Example**: In cities like Los Angeles, predictive policing systems have been used to allocate police resources, but civil rights groups have raised concerns about racial profiling and the system's reliance on biased historical data [2].

4.2 Ethics of facial recognition technology

Findings: AI-powered facial recognition systems have become increasingly prevalent in law enforcement for tasks such as identifying suspects or tracking missing persons. However, there is growing concern about their accuracy and the potential for misuse. Studies have shown that facial recognition systems are less accurate at identifying people of color, particularly Black and Asian individuals, leading to false positives and wrongful arrests. As demonstrated in one of the most influential empirical studies on algorithmic bias in computer vision, "darkerskinned females are the most misclassified group, with error rates of up to 34.7%" [11] empirically demonstrate that "darker-skinned females are the most misclassified group," highlighting severe demographic disparities. Additionally, the widespread use of facial recognition without public consent raises significant privacy issues [3]. **Example**: In the UK, the use of facial recognition by police has faced public backlash due to concerns about its accuracy and potential misuse, prompting calls for stricter regulation. [12] show that facial recognition regulation "varies significantly across the EU, U.S., and Asia-Pacific."

4.3 Accountability in AI systems

Findings: Accountability in the use of AI in policing remains a significant challenge. When AI systems make mistakes (e.g., false arrests due to flawed algorithms or predictive policing errors), it can be difficult to determine who is responsible. Legal experts argue that the lack of transparency in the design and functioning of algorithms makes it difficult to hold any party accountable. Calls for clear policies and oversight mechanisms are growing to ensure that police departments using AI technologies are held to high standards of accountability [1]. **Example**: In the U.S., advocacy groups have been pushing for laws that would require police departments to disclose their use of AI technologies and allow independent audits of these systems [2].

4.4 Privacy concerns in AI surveillance

Findings: The use of AI in surveillance, including facial recognition and social media monitoring, raises significant privacy concerns. Law enforcement agencies often use these technologies to monitor public spaces or track online activity, sometimes without clear consent from the public. This can lead to mass surveillance, where individuals' activities are continuously monitored without their knowledge. Critics argue that this infringes on civil liberties and could lead to a "surveillance state" [3]. **Example**: The controversy over police use of facial recognition during protests, such as the Black Lives Matter protests in the U.S., sparked widespread debate about the balance between security and privacy.

4.5 Effectiveness and efficiency of AI in automated documentation

Findings: AI-driven tools for automating routine tasks like report writing and document management have been shown to improve efficiency within police departments. By automating administrative work, officers can focus more on community policing and crime prevention. However, concerns have been raised about the potential for errors in automated reports, as AI systems may lack the context or understanding needed to handle complex situations. **Example**: In some police departments, AI systems are being used to automatically generate incident reports based on officer input, but some officers have expressed concerns that these systems may miss critical details or misinterpret data [2].

4.6 Public perception of AI in policing

Findings: Public opinion on the use of AI in law enforcement is mixed. While some citizens see AI tools as helpful for improving crime prevention and police efficiency, others are deeply concerned about the implications for privacy and civil rights. Research indicates that public trust in AI systems is often contingent on transparency, accountability, and the demonstration of clear benefits without compromising civil liberties [3]. **Example**: Polls conducted in various countries show that while a majority of people may support AI tools for enhancing law enforcement's effectiveness, a significant portion also calls for more oversight and safeguards against potential abuse.

4.7 Ethical regulation of AI in policing

Findings: Regulatory approaches vary significantly across jurisdictions. The European Union's AI Act proposes strict limitations on realtime biometric surveillance, while the United States relies on decentralized municipal bans and transparency ordinances. [13] warn of "societal risks associated with AI-driven biometric surveillance." The [14] warns that facial recognition poses "serious fundamental rights challenges." In contrast, China employs expansive AIdriven

surveillance as part of its national security strategy. These divergent models illustrate the tension between security, civil liberties, and state governance philosophies.

The rapid adoption of AI in policing has outpaced the development of legal frameworks to regulate its use. Ethical concerns, including issues of bias, privacy, and accountability, have prompted calls for more robust regulations. Policymakers and civil rights groups argue for the creation of national and international standards to govern AI in policing, ensuring that these technologies are used responsibly and in a way that respects individual rights. Example: Some European countries, including the UK and France, are exploring regulations to limit the use of facial recognition in public spaces, while other countries like China have adopted more pervasive surveillance practices, sparking international concerns about privacy. [15] classifies real-time biometric surveillance as a "high-risk AI practice subject to strict regulatory constraints." [16] propose "legitimate surveillance" as an ethical framework for AI use in policing.

4.8 Impact of AI on police resource allocation

Findings: AI-powered tools have been shown to improve police resource allocation by providing insights into crime patterns and predicting where resources should be deployed. However, there are concerns about over-reliance on AI predictions, which could lead to misallocation of resources and reinforce existing policing strategies that disproportionately affect certain communities. Critics warn that predictive tools should not be used as the sole basis for decision-making but should be combined with human oversight and local context. **Example**: The city of Chicago has implemented AI-based resource allocation systems that help direct police patrols based on crime data, but concerns have been raised that this may lead to over-policing in certain neighborhoods.

4.9 AI's role in combating crime vs. reinforcing inequality

Findings: While AI technologies have the potential to reduce crime and improve law enforcement efficiency, there is a risk that these tools could reinforce existing social inequalities. If AI systems are trained on biased data, they may perpetuate discriminatory practices in policing, leading to higher rates of surveillance and arrest for certain racial or socioeconomic groups. Researchers argue that addressing this issue requires a careful examination of how AI is deployed and a commitment to ensuring that these technologies do not exacerbate inequality.

[1] warns that predictive policing can "institutionalize historical patterns of surveillance embedded in crime data." **Example**: The controversy surrounding the use of predictive policing in minority neighborhoods highlights how AI, while designed to enhance crime prevention, may end up reinforcing systemic biases in law enforcement.

To synthesize the operational and ethical dimensions of AI deployment in policing, Table 1 provides a comparative overview of key technologies, their benefits, and associated risks.

AI Technology	Operational Benefits	Ethical and Technical Risks
Predictive policing (PredPol, HunchLab; placebased & personbased models)	Improves hotspot forecasting; enhances patrol allocation; supports strategic deployment of resources	Reinforces historical data bias; risk of overpolicing marginalized communities; limited algorithmic transparency
Facial Recognition Technology (FRT)	Rapid suspect identification; supports missingperson investigations; aids forensic analysis	Higher falsepositive rates for minority groups; potential for mass surveillance; lack of public consent
Automatic License Plate Recognition (ALPR)	Efficient vehicle tracking; automated identification of stolen or wanted vehicles; supports traffic enforcement	Longterm retention of movement data; surveillance misuse risks; unclear dataretention policies
NLPbased social media analysis	Detects emerging threats; identifies criminal networks; monitors online radicalization trends	Largescale data collection without consent; contextual misinterpretation; potential infringement on free expression
AIdriven administrative automation (report generation, document processing)	Reduces administrative workload; accelerates case processing; standardizes documentation	Risk of misinterpretation of complex cases; reduced human judgment; overreliance on automated outputs
Drones and robotic systems	Enhances officer safety; improves situational awareness; enables access to hardtoreach areas	Potential for disproportionate surveillance; regulatory gaps; risk of police militarization
Computer vision for video analytics	Accelerates review of large video datasets; detects suspicious activities; supports forensic investigations	False positives; dependence on video quality; embedded dataset biases

Table 1. Comparative Overview of AI Technologies in Policing, Their Operational Benefits, and Ethical Risks

5. Final reflections

The integration of artificial intelligence (AI) into policing represents a transformative shift in law enforcement, offering significant potential to enhance efficiency, improve crime prediction and prevention, and automate routine tasks. Tools such as predictive policing, facial recognition, license plate recognition, and AI-driven drones have demonstrated their utility in optimizing resource allocation, identifying suspects, and reducing risks to officers and civilians. However, the adoption of these technologies is not without controversy, as it raises profound ethical concerns related to bias, privacy, and accountability.

AI systems, such as predictive policing algorithms, have been criticized for perpetuating racial and socioeconomic biases due to their reliance on historically skewed data. Similarly,

facial recognition technologies, while effective in identifying suspects, have been shown to disproportionately misidentify people of color, leading to wrongful arrests and privacy violations. The lack of transparency and accountability in AI decision-making further complicates these issues, as it becomes challenging to determine responsibility when errors occur.

The widespread use of AI in surveillance, including social media monitoring and facial recognition, has sparked debates about the balance between public safety and individual privacy. Critics warn that unchecked surveillance could lead to a "surveillance state," where civil liberties are compromised. Additionally, while AI-driven automation of administrative tasks improves efficiency, concerns remain about the potential for errors and the loss of nuanced human judgment in complex situations.

Public perception of AI in policing is mixed, with some citizens appreciating the potential for improved safety and others expressing concerns about bias and privacy violations. To address these challenges, there is a growing call for robust regulatory frameworks that ensure the ethical use of AI in law enforcement. Policymakers, civil rights organizations, and technology developers must collaborate to establish guidelines that balance the benefits of AI with the protection of individual rights.

Looking ahead, the future of AI in policing will likely involve more advanced technologies, such as explainable AI, which aims to provide greater transparency in decision-making processes. [17] argue that explainable AI can "strengthen democratic accountability in policing." However, the challenge will be to ensure that these advancements do not exacerbate existing inequalities or undermine public trust. By addressing ethical concerns and fostering accountability, AI can be a powerful tool for enhancing public safety while respecting civil liberties.

6. Relevance to Higher Education

Artificial intelligence in policing has become a central topic for highereducation institutions that train future professionals in criminal justice, public security, digital forensics, and information sciences. Universities increasingly face the responsibility of equipping students with competencies needed to understand algorithmic decisionmaking, evaluate risks of bias, and critically assess the societal implications of AIsupported policing practices. [18] argues that AI produces "political, economic, and ecological costs that are often obscured."As policing agencies adopt predictive analytics, facial recognition, automated surveillance, and riskassessment tools, highereducation curricula must evolve to provide theoretical foundations, methodological skills, and ethical frameworks for analyzing these technologies. The relevance of this paper to higher education lies in its contribution to curriculum development, pedagogical design, and academic discourse. By examining AI in policing through conceptual, operational, and ethical lenses, the paper offers educators structured content that can be integrated into lectures, seminars, and casebased learning activities. This supports universities in preparing students for technologically advanced policing environments and strengthens academic programs that address digital transformation in publicsecurity institutions.

6.1 Pedagogical Contribution

The integration of artificial intelligence into policing practices creates new pedagogical demands for highereducation institutions that prepare future professionals in criminal justice, public security, digital forensics, and communication management. This paper contributes to pedagogy by offering a structured framework for teaching AIenabled policing within university curricula. The proposed framework supports educators in designing learning activities that develop students' competencies in algorithmic reasoning, ethical evaluation, risk assessment, and evidencebased decisionmaking.

The paper provides conceptual models and casebased examples that can be used to construct lectures, seminars, and practical exercises. These instructional elements help students critically examine AI tools such as predictive analytics, facialrecognition systems, automated surveillance platforms, and riskassessment algorithms. By engaging with realworld policing scenarios, students learn to identify technological limitations, sources of bias, and governance challenges associated with AI deployment. [19] describe AI as part of an "everything crisis," requiring new governance models.

In addition, the paper outlines pedagogical strategies that promote active learning, including scenario simulations, guided debates, problembased learning, and reflective analysis. These approaches enable educators to cultivate analytical thinking, ethical awareness, and interdisciplinary understanding—competencies essential for responsible use of AI in policing. The pedagogical contribution therefore lies in translating complex technological and operational issues into accessible, academically rigorous teaching materials that enhance curriculum quality and support educator effectiveness.

6.2 Usefulness for Educators

The insights presented in this paper are designed to support educators who teach courses in criminal justice, public security, digital forensics, communication management, and information sciences. By translating complex AIdriven policing technologies into structured conceptual modules, the paper provides instructors with academically grounded teaching materials that can be incorporated into lectures, seminars, and practical exercises.

Educators can use the analytical frameworks and case studies offered here to facilitate student understanding of predictive analytics, facialrecognition systems, automated surveillance, and algorithmic riskassessment tools. These materials enable instructors to guide students through critical evaluation of technological capabilities, limitations, and ethical implications. The paper also proposes pedagogical strategies—such as scenariobased simulations, guided debates, and problembased learning—that help educators foster analytical reasoning, ethical awareness, and interdisciplinary thinking.

Furthermore, the paper supports curriculum development by outlining learning outcomes relevant to AIsupported policing, including digital literacy, responsible technology governance, and evidencebased decisionmaking. This makes the paper directly useful for educators seeking to update or expand their courses to reflect contemporary technological transformations in policing and publicsecurity institutions.

7. Future directions

Emerging mitigation strategies include fairnessaware machine learning (e.g., equalized odds, demographic parity), dataset debiasing techniques, differential privacy for sensitive data, and model interpretability tools such as SHAP or LIME. These approaches aim to reduce algorithmic discrimination and increase transparency in operational decisionmaking.

As AI technology continues to evolve, its role in policing is expected to expand, incorporating more advanced machine learning models and autonomous systems. While AI has the potential to enhance efficiency, accuracy, and crime prevention, its use also raises significant ethical concerns. Ensuring responsible AI deployment requires a balanced approach that prioritizes fairness, transparency, accountability, and civil liberties. Ethical development of AI systems must be designed with fairness and transparency in mind. Regular audits of algorithms are essential to detect and mitigate bias.Community involvement in discussions about AI use in policing fosters trust, addresses public concerns, and promotes transparency.Regulation and oversight, governments and law enforcement agencies must establish clear policies and oversight mechanisms to ensure AI is used responsibly and does not infringe on individual rights. Policymakers and civil rights organizations continue to push for stricter regulations. Training for law enforcement, educating officers on AI's capabilities and limitations ensures responsible and effective use in policing.Focus on human rights, AI deployment must prioritize the protection of civil liberties, preventing misuse and potential abuses of power.Advancements in AI transparency, research is increasingly focused on developing "explainable AI" systems that provide clear, understandable decision-making processes, particularly in areas like predictive policing.

These findings reflect both the promise and risks associated with AI in policing. While AI technologies offer significant advantages in law enforcement, they also present challenges related to fairness, privacy, and accountability. The future of AI in policing will likely involve stronger collaboration between technology developers, law enforcement agencies, civil rights organizations, and policymakers to ensure that AI systems are used ethically and effectively. Balancing technological innovation with human rights, transparency, and accountability is essential for building public trust and ensuring fair and just law enforcement practices.

7.1 Curricular Integration

The rapid adoption of artificial intelligence in policing requires highereducation institutions to integrate AIrelated competencies into existing curricula in criminal justice, public security, digital forensics, communication management, and information sciences. This paper supports curricular integration by identifying the knowledge domains, skills, and learning outcomes that university programs must address to prepare students for technologically advanced policing environments.

The conceptual frameworks and case analyses presented in this paper can be incorporated into core and elective courses dealing with policing strategies, digital transformation in public institutions, forensic technologies, and ethical governance. By embedding AIsupported policing topics into lectures, seminars, and laboratory exercises, educators can help students understand algorithmic decisionmaking, evaluate risks of bias, and critically assess the operational and societal implications of AI deployment.

The paper also provides guidance for curriculum designers seeking to update academic programs. It outlines how modules on predictive analytics, facial recognition, automated surveillance, and riskassessment algorithms can be integrated into existing course structures without requiring major program redesign. These modules can be aligned with competencybased education frameworks, ensuring that students develop digital literacy, ethical reasoning, and evidencebased decisionmaking skills.

7.2 Learning Outcomes

Within highereducation programs focused on policing, public security, digital forensics, and information sciences, students must develop a comprehensive understanding of the fundamental principles of artificial intelligence, machine learning, and algorithmic decisionmaking, as well as their application in policing. This includes the ability to explain how AI is used in predictive analytics, facialrecognition systems, automated surveillance, and algorithmic riskassessment tools, and to recognize the technological limitations, dataquality requirements, and potential sources of error that may influence policing outcomes. [20] show that “dirty data produce bad predictions,” undermining fairness.

In addition to conceptual understanding, students are expected to acquire analytical and criticalthinking skills that enable them to evaluate the accuracy, reliability, and validity of AIbased policing systems using evidencebased criteria. Through the analysis of case studies, they should learn to identify operational risks, unintended consequences, and governance challenges associated with AI deployment, as well as assess algorithmic bias, discriminatory outcomes, and fairness issues in automated decisionmaking processes.

Ethical and legal reasoning represents another essential dimension. Students must understand ethical frameworks relevant to AI in policing, including human rights, privacy, transparency, and accountability. They should be able to evaluate AIsupported policing practices in relation

to national legislation, EU regulations, and international standards, and formulate ethically responsible approaches to the design, implementation, and oversight of AI technologies.

Operational and professional competencies are equally important. Students should be capable of applying riskassessment and decisionsupport models in simulated policing scenarios involving AI tools, interpreting algorithmic outputs, and integrating them into professional decisionmaking. They must also be able to communicate AIrelated findings clearly and precisely to diverse audiences, including police leadership, policymakers, and the general public.

Developing digital literacy and data competence involves understanding dataquality requirements, principles of data governance, and dataprotection obligations in AIenabled systems. Students should be able to evaluate datasets for representativeness and potential bias and demonstrate foundational skills in interpreting algorithmic models, visualizations, and analytical reports.

Interdisciplinary integration forms an important part of the educational process. Students need to connect AIsupported policing practices with fields such as criminology, communication management, public administration, and information sciences, and analyze how AI reshapes organizational culture, communication flows, and publicsecurity strategies. This fosters the ability to integrate socialscientific, legal, and technological perspectives when evaluating AI in policing.

Finally, students should develop reflective and responsible professional practice. This includes considering the societal implications of AIenabled policing, such as issues of trust, legitimacy, and public perception. They should be able to propose responsible innovation strategies for the adoption of AI technologies in policing institutions and demonstrate awareness of the ethical responsibilities of future practitioners working with such systems. [1] identifies “predictive policing 2.0” as a new generation of more advanced yet less transparent systems.

8. Conclusion

AI is transforming policing by enhancing efficiency, crime prediction, and resource allocation, yet its adoption raises persistent ethical concerns involving **bias, privacy,** and **accountability**. While AI can strengthen public safety, responsible implementation requires transparency, oversight, and regulation to prevent reinforcing systemic disparities or infringing on civil rights. A balanced approach—technologically advanced yet ethically grounded—is essential to uphold justice and equality. The findings emphasize the importance of integrating **AIinpolicing competencies** into university curricula so future practitioners understand both operational capabilities and ethical constraints of emerging technologies. Future research should prioritize longitudinal evaluations of AI tools in real policing environments, crossnational analyses of regulatory effectiveness, and empirical testing of algorithmic fairness interventions. Evidencebased assessment is crucial for ensuring that AI supports policing in ways that enhance safety while protecting civil liberties. Ultimately, the trajectory of AI in law enforcement will depend on aligning innovation with rigorous ethical standards and sustained institutional oversight.

References

[1] A. G. Ferguson, "*Policing Predictive Policing*", Washington University Law Review, vol. 94, no. 5, pp. 1109–1189, 2017. https://wustllawreview.org/wp-content/uploads/2021/10/94.5.1.pdf?utm_source=copilot.com

[2] R. Benjamin, *Race after technology: Abolitionist tools for the new Jim Code*. Cambridge, UK: Polity Press, 2019.

[3] S. Brayne, *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press, 2021.

[4] E. Bietti, "The ethics of algorithmic policing: A critical review," *Philosophy & Technology*, vol. 35, no. 3, pp. 1–28, 2022.

[5] A. L. Hoffmann, "Even when you are doing everything right: Algorithmic harms and the limits of AI fairness," *Information, Communication & Society*, vol. 25, no. 14, pp. 1919–1935, 2022.

[6] A. Završnik, "Criminal justice, artificial intelligence systems, and human rights," *ERA Forum*, vol. 20, no. 4, pp. 567–583, 2020.

[7] R. Kellogg and J. Zhao, "Algorithmic governance in public safety: Evaluating transparency and accountability mechanisms," *Government Information Quarterly*, vol. 40, no. 2, p. 101–124, 2023.

[8] P. Molnar and L. Gill, "Bots at the border: The use of AI in immigration and policing contexts," *European Journal of Migration and Law*, vol. 24, no. 2, pp. 133–160, 2022.

[9] V. Marda and S. Narayan, "Data-driven policing in the Global South: Algorithmic harms and democratic erosion," *Big Data & Society*, vol. 9, no. 1, pp. 1–12, 2022.

[10] S. Khan and P. Goff, "Measuring racial bias in policing algorithms: New empirical approaches," *Journal of Quantitative Criminology*, vol. 38, no. 4, pp. 1007–1032, 2022.

[11] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," *Proc. Mach. Learn. Res.*, vol. 81, pp. 1–15, 2018.

[12] M. Mann, A. Daly, and M. Wilson, "The limits of facial recognition regulation: A comparative analysis of EU, US, and Asia-Pacific frameworks," *International Data Privacy Law*, vol. 12, no. 3, pp. 215–230, 2022.

[13] R. Williams and E. Kind, "The future of biometric surveillance: Assessing the societal risks of AI-driven policing," *Surveillance & Society*, vol. 20, no. 4, pp. 480–499, 2022.

[14] FRA, *Facial recognition technology: Fundamental rights considerations*. European Union Agency for Fundamental Rights, 2022.

[15] European Commission, *Artificial Intelligence Act*. Official Journal of the European Union, 2024.

[16] E. Aizenberg and J. van den Hoven, "Designing for legitimate surveillance: A conceptual framework for the ethical use of AI in policing," *AI and Ethics*, vol. 2, no. 1, pp. 45–60, 2022.

[17] K. Crawford, *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press, 2021.

[18] G. Smith and J. Miller, "Explainable AI in policing: Opportunities and risks for democratic accountability," *Policing: A Journal of Policy and Practice*, vol. 17, no. 1, pp. 45–62, 2023.

[19] M. Sloane and E. Moss, "AI and the everything crisis: Rethinking ethics, governance, and accountability," *Patterns*, vol. 3, no. 7, p. 100550, 2022.

[20] R. Richardson, J. Schultz, and K. Crawford, "Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice," *NYU Law Review*, vol. 96, no. 1, pp. 192–235, 2021.

ISSN 2991-9371

From Knowing to Doing: Redesigning Business Education for AI

Monika Sońta[1], Lyubomyr Matsekh-Ukrayinskyy[2] and Maciej Mąkosa[3]

[1]Kozminski University, Warsaw, Poland, msonta@kozminski.edu.pl
[2]Reutlingen University, Reutlingen, Germany,
Lyubomyr.Matsekh-Ukrayinskyy@Reutlingen-University.de
[3]Kozminski University, Warsaw, Poland, 49844@kozminski.edu.pl

Abstract

As artificial intelligence transforms managerial practice, business schools face growing pressure to equip graduates with artificial intelligence literacy while keeping curricula accessible to non-technical learners. In this paper, artificial intelligence refers primarily to generative tools such as large language model assistants, the technology driving current classroom and workplace adoption; the broader term artificial intelligence is retained for consistency with the education-literacy literature. Despite emerging educational initiatives, limited evidence exists on effectively designing such courses within actionable pedagogical frameworks. This paper addresses this gap through two complementary pain-point analyses grounded in thematic inquiry: a corpus of 838 unique sources from academic, enterprise, and practitioner platforms, and a longitudinal dataset of 607 student evaluation records collected across five academic years from 1,236 enrolled management students. Findings reveal five systemic obstacles: entry-point paralysis, acute anxiety regarding self-efficacy, shallow enterprise adoption, workplace stigma, and institutional incoherence. Practical application, active tool integration, contemporary case studies, and collaborative pitch projects consistently outperformed traditional lecture-based instruction. Grounded in these empirical insights, this study proposes actionable guidelines for course redesign, alongside core pedagogical principles supporting agile, continuously evolving curricula for management education.

Keywords: *Artificial Intelligence in Education, Generative AI, Management Education, Curriculum Design, AI Literacy, Experiential Learning.*

1. Introduction

As artificial intelligence (AI) transforms managerial practice, business schools face growing pressure to prepare graduates with AI literacy while ensuring that curricula remain accessible to students without technical backgrounds. Throughout this paper, "AI" is used as the umbrella term established in the education-literacy literature, while "generative AI" is used specifically where findings concern large language model tools (e.g., ChatGPT, Copilot) — the dominant form of AI students and instructors encounter in practice today. This shift necessitates a departure from static pedagogical models toward dynamic frameworks that align AI literacy with evolving employment demands and institutional policy support [1]. The study proposes a structured taxonomy of pedagogical interventions designed to navigate the tension between rapid technological disruption and the foundational requirements of academic integrity [2]. By leveraging this taxonomy, educators can better reconcile student expectations for practical, future-oriented training with the institutional necessity for clear guidance on responsible and ethical AI adoption [3].

Although numerous AI education initiatives have emerged, limited evidence is available on how AI courses can be effectively designed for management students within a practical framework [4]. This framework identifies major pedagogical bottlenecks, including the misalignment between theoretical instruction and industry application, and advocates modular curricula [5] that develop core AI competencies while supporting institutional restructuring for long-term sustainability [4].

Integrating structured AI interactions also lets educators balance technological innovation with human-centric soft skills [6], so that students cultivate not only technical proficiency but also the critical evaluation and metacognitive engagement needed to navigate the AI-driven workplace [7]. Accordingly, this study offers a roadmap for transitioning from passive knowledge dissemination to active, andragogical models that mitigate overreliance while fostering professional autonomy [8], with AI literacy embedded across domain-specific management subjects rather than treated as a peripheral technical skill [9].

2. Theoretical Background

The rapid integration of generative AI into professional workflows has fundamentally transformed the skills expected of contemporary business graduates, shifting pedagogical priorities toward the critical evaluation and ethical application of these technologies [10]. Emerging scholarship grounded in social constructivism and connectivism seeks to make AI instruction more interdisciplinary by integrating technical competencies with ethical reasoning and informed AI systems management [11].

2.1 AI Literacy as a Conceptual Foundation

A first family of approaches centers on cultivating AI literacy: a consolidated set of competencies that equip nonexperts to understand, use, and critically evaluate AI without requiring programming skills. Laupichler et al. conclude that AI literacy research in higher and adult education is still in its infancy, with no settled definition of what content should be taught to non-expert audiences [12]. Ng et al. provide one of the most influential conceptual anchors by proposing four intertwined aspects of AI literacy — know and understand, use and apply, evaluate and create, and ethical issues — which reframe the goal of teaching from technical proficiency toward informed, responsible engagement with AI [13]. Institutions have operationalized this differently: the University of Florida's "AI Across the Curriculum" initiative infuses AI education into every discipline [14], while Locke and Rainer propose a dedicated three-course sequence moved out of computer science and into the mainstream business curriculum [15].

2.2 Experiential and Project-Based Learning with AI

A second family of approaches insists that literacy alone is insufficient and pairs conceptual grounding with handson, project-based experience. Wunderlich et al. demonstrate that machine learning is particularly amenable to experiential learning for business students: community-based tools and public datasets let non-programmers adapt open ML examples and interpret business implications without writing code from scratch [16]. An interdisciplinary project-based course combining creativity methods with AI prototype development across two editions with 32 students reports enhanced student confidence and a more nuanced grasp of AI technologies [17]. Survey evidence from business and economics students points the same way: students are aware of AI's impact and willing to learn, but their knowledge is limited by lack of training, and they benefit most from realistic use cases presented with the technology's real limitations rather than abstract theory [18]. These findings align with the recommendation that AI curricula for non-technical fields should embed case studies relevant to students' own disciplines, accompanied by explicit discussion of ethical implications such as algorithmic bias and data privacy [19]. This experiential turn addresses a documented failure mode: traditional coding education methods have consistently proven inadequate for non-technical learners, which is why low-code platforms and scaffolded, project-driven designs are increasingly positioned as the bridge between technological capability and practical professional application [20]. Adopting these modular, hands-on strategies lets educators demystify complex algorithms while grounding theory in the real-world scenarios essential for managerial decision-making [21].

2.3 Education as Sustainability

Beyond experiential and project-based designs, low-code and no-code development offers a complementary pathway for non-IT learners. Sońta and Przegalińska argue that students without programming backgrounds can co-create working IT solutions when educators articulate entry points into low-code learning and keep the technology "in the background" while students solve

authentic problems from their own experience [22]. Because traditional coding education has consistently proven inadequate for non-technical learners [23], low-code platforms lower the perceived access barrier and operate as a "building block approach" [20], conceptualized as "education as sustainability" [20]. This shift aligns with broader efforts to reimagine management education, where students leverage AI-powered tools not merely for technical proficiency but as a mechanism for addressing complex global challenges [24], [25]. Across these three frameworks, existing research has largely focused on single-course implementations and small cohorts, limiting generalizability across student populations, delivery formats, and institutional settings [17], [26]. Broader implementation also depends on instructor competence, infrastructure, and mechanisms for updating course content as AI capabilities evolve [27]. Against this background, the originality of the present study lies in its two-stage analytical design, combining a web-scraping analysis with a longitudinal student-evaluation dataset, establishing an open framework that educators can adapt to their own institutional contexts.

3. Methods

The study is structured around two complementary data sources that together capture AI pedagogy in business education from both a global and a classroom-level perspective. To structure the empirical inquiry, this study addresses three interrelated research questions: (1) What are the main challenges in teaching a general “AI in Business” course to management students without technical backgrounds? (2) Which elements of the learning process do students rate highest, and which lowest? (3) Which tools and pedagogical approaches are most effective?

We define short introductory ‘AI in Business’ courses as classes comprising between 12 and 24 teaching hours and typically not exceeding 3 ECTS credits, most often introduced by universities to rapidly enhance students’ digital competencies in response to labor-market demand [28]. To answer the research questions, we analyzed two study components separately — an analysis of online sources and an evaluative survey study — with the third question answered in the discussion section through data triangulation. Taking into account the possibility of systematic bias arising from the selection method and the nature of web-scraped data, we focused our recommendations primarily on solutions identified within corresponding thematic categories across both analyses, limiting potential discrepancies between theoretical discourse and practical educational needs.

3.1 Web-Scraping Analysis: Multi-Source Data Collection Framework

The first component is a web-scraping analysis: 838 unique raw sources collected from ten academic, enterprise, and informal community platforms — arXiv, OpenAlex, ERIC, Semantic Scholar, Exa web search, YouTube, Reddit, Forrester, and PwC — acquired through 15 targeted query clusters and 10 continuous data streams. Platforms were selected to deliberately triangulate three source types: peer-reviewed academic repositories (arXiv, OpenAlex, ERIC, Semantic Scholar), for methodological rigor and established pedagogical frameworks;

enterprise/industry research (Forrester, PwC), for workforce AI-skill demand and adoption-ROI data unavailable in academic literature; and informal community platforms (Reddit, YouTube, Exa web search), for unfiltered instructor and practitioner sentiment not captured by either of the above. This multi-source architecture balances rigorous peer-reviewed literature with real-time practitioner feedback, ensuring that the identified pain points reflect both published research and emerging classroom workarounds.

For YouTube specifically, the harvester used yt-dlp to collect public video title and description metadata together with the top viewer comments on each video — not the video or audio content itself — on the premise that comments on AI-education videos capture authentic learner reactions comparable in kind to Reddit discussion threads.

To capture high-signal qualitative insights while mitigating noise from unmoderated platforms, version 4 (v4) of the scraping pipeline scoped informal community extraction to high-relevance domain hubs (r/Professors, r/artificial, r/education).

Platform Category	Data Streams Included	Signal Density Rating	Methodological Contribution
Academic Repositories	arXiv, OpenAlex, ERIC, Semantic Scholar	High	Peerreviewed pedagogical frameworks, empirical interventions, and foundational theory.
Enterprise Insights	Forrester, PwC	High	Industry alignment metrics, workforce AI skill requirements, and institutional benchmarks.
Community & Web	Reddit, YouTube, Exa Web Search	Moderate–High	Unfiltered instructor sentiment, emerging classroom workarounds, and student feedback.

Table 1. The multi-source web-scraping data collection framework applied in this study (author's own elaboration).

Eight hundred thirty-eight unique raw sources were collected across nine working sources; the majority not cited above were off-topic false positives from generic queries (physics education, unrelated business-modeling papers, general tech news, etc.) — normal noise for keyword-based search, excluded rather than force-fit into the painpoint framing. ERIC drove most new findings in v2; Exa and YouTube drove v3; Reddit (once fixed) drove v4 and was the highest-signal single source for real, unprompted complaints.

Two limitations follow from this design. First, keyword-based retrieval and platform accessibility bias the corpus toward English-language, publicly indexed content; private communities (e.g., closed Slack/Discord groups) were unreachable and are under-represented. Second, community-sourced material (Reddit, YouTube comments) reflects self-selected, unverified posters rather than a representative sample, so these sources are treated as illustrative qualitative signal — not generalizable prevalence estimates — throughout.

3.2 Student Feedback Analysis

The second source comprises a longitudinal analysis of 607 student evaluation records collected over five academic years from the graduate course "Artificial Intelligence in Business." The course is offered within the Master in Management programme at Kozminski University in full-time, part-time, online, and English-track formats and has enrolled 1,236 students in total.

To evaluate instructional quality and learning outcomes, a standardized evaluation framework comprising two instruments was administered across academic years 2021/2022 through 2025/2026: a Course Quality Evaluation Survey and an Assessment Quality Evaluation Survey. Quantitative items were measured primarily using a 5-point Likert agreement scale (strongly disagree to strongly agree), discrete ordinal frequency scales for attendance and instruction mode (in-person, hybrid, remote), and a single-item overall course rating scale (1 = lowest to 5 = highest). Qualitative constructs gathered open-text responses on effective active-learning methods, elements requiring pedagogical improvement, and contextual feedback.

Across the five academic years from 2021/2022 through 2025/2026, the evaluation dataset comprised 607 survey responses from a total population of 1,236 enrolled students, yielding an overall mean course-satisfaction score of 4.71 out of 5.00. Annual mean evaluation scores remained consistently high, recorded at 4.88 in 2021/2022, 4.73 in 2022/2023, 4.83 in 2023/2024, 4.75 in 2024/2025, and 4.69 in 2025/2026. Mean instructor-evaluation scores were 4.91/5.00, 4.84/5.00, and 4.43/5.00 across the teaching staff. Regarding assessment and examination metrics, objective and fair exam administration received a mean score of 4.88/5.00 across 16 evaluated group instances. This matched the 4.88/5.00 rating assigned across 14 instances for the clarity of grading criteria presented during the first class.

For the thematic analysis, Saldaña's approach to thematic coding was employed [29], [47]. This methodological choice enabled a systematic transition from raw narrative data to recurring educational themes and facilitated the identification of latent patterns in student sentiment [30], [31]. AI tools were used to support the thematic analysis and the development of the coding map; the themes and concepts themselves were derived manually.

4. Results

Regarding the first research question — what are the main challenges involved in teaching a general "AI in Business" course to management students without technical backgrounds? — the findings revealed key structural obstacles.

4.1 Web-Scraping Analysis Pain Points

Table 2 highlights the key thematic obstacles identified across academic discourse and practitioner forums.

No.	Pain Point / Challenge	Evidence and Implications	Source / Research Confirmation
1	Knowing–doing gap	Learners express persistent uncertainty regarding enterprise use cases, starting points, and workflow alignment.	Lim et al. [32]
2	Broad but shallow application	High tool usage breadth (e.g., ChatGPT, DeepSeek) with minimal comprehension of mechanics, yielding dependency.	Vâlcea et al. [33]
3	Reduced struggle & cognitive atrophy	Unregulated GenAI bypasses problemsolving, risking critical thinking degradation and agency loss.	Klein & Klein [34]
4	Social punishment & concealed usage	Workers conceal AI use to avoid lowcompetence perception, destroying peerlearning networks.	Li et al. [35]
5	Faculty & infrastructure capacity gap	Acute shortages in faculty expertise, reskilling time, compute infrastructure, and curriculum updates.	Charow et al. [36]
6	Shallow Copilot tool adoption	Telemetry (5.5M sessions) shows enterprise use clusters on lowstakes writing tasks rather than complex analytics.	Dillon et al. [37]
7	Structural pretertiary equity gap	Uneven early computer science access presorts nontechnical students away from foundational AI concepts.	Bulathwela et al. [38]

Table 2. Key pain points and challenges identified in the web-scraping analysis (author's own elaboration).

Synthesizing these findings into student-centric dimensions reveals five core pain points: Overwhelming Entry Barrier (an “I don't know where to start” paralysis due to tool proliferation); AI Anxiety Suppressing Engagement (fear of job displacement and of appearing incompetent hinders adoption); Superficial Adoption vs. Unproven ROI (wide tool availability rarely yields deep implementation; industry estimates cited in the corpus put the share of enterprise GenAI projects with no measurable return at roughly 95% [39]); Social Stigma and Hidden Workflows (visible AI usage is socially penalized, so competent users hide it); and Corrosive Institutional Incoherence (institutions teach AI while building systems to detect and prevent its use).

4.2 Student Feedback Analysis Pain Points

The principal pain points identified in the student survey concerned course duration, delivery format, practical depth, and assessment transparency. Given the rapid evolution of AI, students described the course as compressed, expressing a strong need for additional contact hours and advanced follow-up modules. Students demonstrated stronger critical engagement when instructors moved beyond generic tool demonstrations toward critical validation of AI-generated analytical outputs [40]. From a delivery standpoint, students indicated the need for a sharper pivot away from theoretical lecturing toward hands-on technical exercises; general reflections on new technologies occasionally took an overly abstract, "philosophizing" form that limited time devoted to practical tool mechanics. Data also revealed friction concerning assessment clarity, showing misalignment between traditional lecture methods and practical management demands [41].

4.3 Web-Scraping Analysis Solutions

Students value curricula emphasizing discipline-specific applications, supporting autonomous skill development and problem-solving [42].

Core Pedagogical Enabler	Operational Mechanism & Practical Impact
Enabler A: NoCode Experiential Design	Replaces passive lecture delivery with active, nocode application building. Students test, iterate, and deploy functional prototypes directly within class workflows [43].
Enabler B: Metacognitive Regulation	Implements targeted interventions (e.g., 2hour workshops) focused on algorithmic questioning, hallucination auditing, and output critique to boost selfefficacy [44].
Enabler C: MultiSource Feedback (PAIRR)	Combines Peer review, AI evaluation, Integrated Reflection, and Revision. AI feedback complements human peer review to safeguard critical thinking [45].
Enabler D: Syntax Bridge Framework	Offloads syntactic and mechanical execution (e.g., code generation) to AI while requiring learners to focus strictly on business logic and interpretation.
Enabler E: AACSB Institutional Alignment	Moves institutional evaluation from isolated tool mastery toward crosscurricular output assessment, preventing siloed policies [46].

Table 3. Core Pedagogical Enablers (author's own elaboration).

4.4 Student Feedback Analysis Solutions

Satisfaction peaks when curricula prioritize scaffolded technical application over abstract conceptualization. Key teaching techniques include hands-on tool integration (direct in-class deployment, e.g., Chatfuel, ChatGPT, Moral Machine), no-code implementation, real-world case method, project pitching and practical assessment, and guest experts from industry. Students appreciated high content currency, clear concept translation, and professional delivery; recommendations for enhancement emphasized shifting the instructional ratio toward guided workshops, expanding contact hours, and communicating final grading rubrics earlier in the term.

4.5 Integrated Approach

To answer the third research question — which tools and pedagogical approaches are most effective? — we synthesized findings across both datasets; real-world case studies emerged as a primary interconnected theme [43].

Thematic Domain / Dimension	WebScraping Corpus Findings	Student Evaluation Survey Findings	Integrated Synthesis & CrossAnalysis	Targeted Pedagogical Enabler
1. Execution & Practical Application	Persistent knowing–doing disconnect; broad but shallow tool usage.	High satisfaction with handson tool testing; critique of abstract lecturing.	Passive conceptual instruction creates execution paralysis.	Enabler A & D: Nocode prototyping & Syntax Bridge Framework.
2. Cognitive Engagement & Learning	Unregulated GenAI risks cognitive atrophy and loss of agency.	Positive engagement with output verification exercises.	Tool overreliance is mitigated by output auditing.	Enabler B: Metacognitive Regulation Module (Week 1).
3. Assessment Design & Evaluation	Contradictory institutional mandates; traditional assignments obsolete.	Highly rated capstone pitches; pain point around late rubrics.	Traditional exam formats fail; pitch evaluations succeed with early rubrics.	Enabler C: PAIRR Framework & OutputAuditing Rubrics.
4. Social Dynamics & Culture	Stigma leads users to conceal usage, destroying peer networks.	High appreciation for guest speakers; need for clear boundaries.	Concealed usage damages collaborative learning.	Enabler E: AACSB Alignment & Disclosure Protocols.
5. Course Scope & Delivery	Faculty shortages, compute bottlenecks, update burdens.	High satisfaction with currency; friction over short contact hours.	Rapid AI evolution creates course compression strain.	RealWorld Case Method: Agile delivery & workshop focus.

Table 4. Master Integrated Thematic & Evaluation Matrix (author's own elaboration).

5. Discussion

The conclusions drawn from both thematic analyses reflect broader scholarly literature. Phenomena such as “cognitive atrophy,” the “credibility gap,” the “syntactic facade,” and AI stigma are increasingly recognized in academic and enterprise settings [48]. This convergence highlights three recurring problem categories: Adoption Dynamics (shift from generic text generation toward specialized no-code modeling and industry workflows — learning through empirical experience); Integrity & Perception (dissatisfaction with automated detection tools, driving demand for transparent usage rubrics — disclosing AI use); and Instructional Friction (balancing conceptual mastery with automated efficiency in assessment — assessing human work). Given the empirical dataset, connectivism in Seymour Papert's formulation serves as a compelling theoretical lens for understanding these dynamics [49].

5.1 Practical Implications for Educators

Grounded in these findings, we outline three practical guidelines for course redesign: (1) Integrate a Week 1 metacognitive regulation module — a mandatory two-hour intervention addressing AI anxiety, establishing cognitive boundaries, and equipping non-technical learners with strategies for questioning AI outputs. (2) Rebalance assessment criteria toward output auditing and the Peer + AI Review + Reflection (PAIRR) model [50], [51], emphasizing logical validation and hallucination detection over prompt syntax. (3) Establish transparent AI disclosure protocols to eliminate social stigma and normalize AI use as a peer-learning opportunity [52], supported by iterative design-thinking to deepen conceptual understanding [53]. Deep-dive real-world case studies further bridge theoretical frameworks and operational realities, providing empirical grounding and actionable benchmarks for curriculum design.

5.2 Study Limitations

The reliance on a single-institution cohort for student evaluation data limits the generalizability of these pedagogical findings across diverse sociocultural contexts. The web-scraping corpus, as noted in Section 3.1, is similarly bounded by platform accessibility and self-selection among posters. Future research should adopt multi-institutional quantitative designs paired with qualitative educator interviews to test robustness across varying higher-education settings.

6. Conclusion

The integration of generative AI into business education necessitates a transition from passive consumption to an "execution mindset," where students learn to contextualize and implement AI-derived insights to avoid abstract "philosophizing." Future iterations of management curricula must prioritize adaptive assessment models that evaluate analytical reasoning, critical auditing, and strategic synthesis over raw technical execution.

7. Acknowledgement

This research was funded by the 4InnoPipe2 – Strengthening Innovation Pipelines for Impactful Universities 2.0, EIT Cohort 4.

THE USE OF AI TOOLS

During the preparation of this manuscript, the authors used Gemini 2.0 Flash (Google DeepMind; accessed August 2026) for literature screening and supporting thematic coding visualizations. The authors independently evaluated all analytical outputs and take full responsibility for all final themes, conceptual models, and manuscript content.

References

[1] A. B. M. M. Rahman, J. Yang, Z. Zhan, M. K. Hossain, Y. Xu, and W. Feng, "Pedagogical Transformation Through Artificial Intelligence: Implementation Barriers and Adoption Strategies in Business Education," Journal of the Association for Information Systems, 2025.

[2] H. Rai, N. Malik, and A. Varma, "Generative AI and management education: a knowledge management perspective," Journal of Knowledge Management, pp. 1–28, 2026.

[3] M. Drummond and G. Dale, "Generating a Student-Informed Teaching and Learning Conceptual Framework for GenAI in Business Schools: A Case Study," Journal of University Teaching and Learning Practice, 2026.

[4] Q. Wu, L. Chen, M. Chen, and Y. Huang, "Exploring the impact of artificial intelligence on business talent development in higher education: A systematic literature review and research agenda," International Journal of Management Education, 2025.

[5] Y. Ma, S. Yufang, L. Ming-da, Y. Zhang, W. Chai, A. Huang, and X. Zhao, "Preparing Students for an AI-Driven World: Generative AI and Curriculum Reform in Higher Education," Frontiers of Digital Education, vol. 2, no. 4, 2025.

[6] L. C. Troy, "Enhancing Marketing Student Success with AI-Guided Experiential Learning: A Case Study," AMTP, 2025.

[7] M. Adamakis and T. Rachiotis, "Artificial Intelligence in Higher Education: A State-of-the-Art Overview of Pedagogical Integrity, Artificial Intelligence Literacy, and Policy Integration," Encyclopedia, vol. 5, no. 4, p. 180, 2025.

[8] S. J. Hyde, A. Busby, and R. L. Bonner, "Tools or Fools: Are We Educating Managers or Creating Tool-Dependent Robots?" Organizational Behavior Teaching Review, vol. 48, no. 4, pp. 708–734, 2024.

[9] J. Kim, S. Yu, R. Detrick, and N. Li, "Exploring students' perspectives on Generative AI-assisted academic writing," Education and Information Technologies, vol. 30, no. 1, pp. 1265–1300, 2024.

[10] C. K. Y. Chan, "A comprehensive AI policy education framework for university teaching and learning," International Journal of Educational Technology in Higher Education, vol. 20, no. 1, 2023.

[11] G. Eboh, "Educating for Responsible AI Governance - Building Interdisciplinary Frameworks for the Management of Artificial Intelligence," Digital Scholarship - Texas Southern University, 2026.

[12] M. C. Laupichler, A. Aster, J. Schirch, and T. Raupach, "Artificial intelligence literacy in higher and adult education: A scoping literature review," Computers and Education Artificial Intelligence, vol. 3, p. 100101, 2022.

[13] D. T. K. Ng, J. K. L. Leung, S. K. W. Chu, and S. Qiao, "Conceptualizing AI literacy: An exploratory review," Computers and Education Artificial Intelligence, vol. 2, p. 100041, 2021.

[14] J. Southworth, K. W. Migliaccio, J. Glover, D. M. Reed, C. McCarty, J. H. Brendemuhl, and A. Thomas, "Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy," Computers and Education Artificial Intelligence, vol. 4, p. 100127, 2023.

[15] H. S. Wisniewski, "What is The Business with AI? Preparing Future Decision Makers and Leaders," Technology & Innovation, vol. 21, no. 4, pp. 337–350, 2020.

[16] L. Wunderlich, A. Higgins, and Y. Lichtenstein, "Machine Learning for Business Students," pp. 512–518, 2021.

[17] A. Azamnouri, D. Hörauf, B. Schönberge, J. Bogner, and S. Wagner, "Teaching AI competencies: Experiences from coaching interdisciplinary teams to develop AI prototypes," Computers and Education Artificial Intelligence, vol. 10, p. 100580, 2026.

[18] C. Almaraz-López, F. A. Menéndez, and C. López-Esteban, "Comparative Study of the Attitudes and Perceptions of University Students in Business Administration and Management and in Education toward Artificial Intelligence," Education Sciences, vol. 13, no. 6, p. 609, 2023.

[19] C. K. Y. Chan and W. Hu, "Students' voices on generative AI: perceptions, benefits, and challenges in higher education," International Journal of Educational Technology in Higher Education, vol. 20, no. 1, 2023.

[20] M. Sońta and D. Kedziora, "Democratizing Software Development: Low-Code as an Educational Pathway towards Sustainable and Collaborative Organizations," 2023.

[21] K. Lucas and M. Elder, "Making the Cut: An Exercise in AI-Aided Decision Making," Management Teaching Review, 2025.

[22] M. Sońta and A. Przegalińska, "Say 'yes' to 'no-code' solutions: how to teach low-code and no-code competencies to non-IT students," in Edward Elgar Publishing eBooks, pp. 330–342, Edward Elgar Publishing, 2024.

[23] D. Sun, A. Boudouaia, C. Zhu, and Y. Li, "Would ChatGPT-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? An empirical study," International Journal of Educational Technology in Higher Education, vol. 21, no. 1, 2024.

[24] S. M. Ghouse, "AI-Enhanced Curriculum for Sustainable Business," pp. 105–124, 2025.

[25] J. Roos, "The Next Management Education: Charting a New Path Forward," pp. 69–82, 2025.

[26] H. S. E. Samaty and N. Albadi, "From policy frameworks to AI mentoring practice: A structured approach to responsible innovation in architectural education," Alexandria Engineering Journal, 2026.

[27] T. I. Park, J.-M. Seo, H.-J. Yoon, and K. E. Lee, "Institutionalizing convergence education for medical artificial intelligence," Biomedical Engineering Letters, vol. 15, no. 6, pp. 1073–1083, 2025.

[28] M. Benmamoun, "Reinventing International Business Education: Integrating the Power of Generative AI," AIB Insights, vol. 24, no. 1, 2023.

[29] J. Saldaña, The Coding Manual for Qualitative Researchers, Hôpital Rivière-des-Prairies, 2025.

[30] N. Bunda, E. D. P. Jr., and G. Dizon, "Understanding student viewpoints: sentiment and thematic insights from course feedback," Journal of Education and Learning (EduLearn), vol. 20, no. 2, pp. 687–699, 2026.

[31] M. Hujala, A. Knutas, T. Hynninen, and H. Arminen, "Improving the quality of teaching by utilising written student feedback: A streamlined process," Computers & Education, vol. 157, p. 103965, 2020.

[32] W. M. Lim, A. Gunasekara, J. L. Pallant, J. Pallant, and E. Pechenkina, "Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators," The International Journal of Management Education, vol. 21, no. 2, p. 100790, 2023.

[33] S. Vâlcea, M. R. Hamdani, and S. Wang, "Exploring the Impact of ChatGPT on Business School Education: Prospects, Boundaries, and Paradoxes," Organizational Behavior Teaching Review, vol. 48, no. 5, pp. 915–947, 2024.

[34] C. R. Klein and R. Klein, "The extended hollowed mind: why foundational knowledge is indispensable in the age of AI," Frontiers in Artificial Intelligence, vol. 8, p. 1719019, 2025.

[35] P. Li, L. Ma, N. Dong, L. Xing, and S. Yuan, "When Hiding Hurts: How Stealth Use of Generative AI Resources Impairs Task Performance," Academy of Management Proceedings, vol. 2025, no. 1, 2025.

[36] R. Charow et al., "Artificial Intelligence Education Programs for Health Care Professionals: Scoping Review," JMIR Medical Education, vol. 7, no. 4, 2021.

[37] E. W. Dillon, S. Jaffe, S. Peng, and A. Cambon, "Early Impacts of M365 Copilot," arXiv preprint arXiv:2504.11443, 2025.

[38] S. Bulathwela, M. Pérez-Ortiz, C. Holloway, M. Cukurova, and J. Shawe-Taylor, "Artificial Intelligence Alone Will Not Democratise Education: On Educational Inequality, Techno-Solutionism and Inclusive Tools," Sustainability, vol. 16, no. 2, p. 781, 2024.

[39] J. Westover, "The GenAI Divide: Why 95% of Enterprise AI Investments Fail—and How the 5% Succeed," Human Capital Leadership, vol. 28, no. 3, 2025.

[40] S. McPhee and M. Jerowsky, "Beyond technical skills: a pedagogical perspective on fostering critical engagement with generative AI in university classrooms," Frontiers in Education, vol. 10, 2025.

[41] A. G. Ossai, M. B. Asabor, J. O. Oduma, and M. N. Oseji, "Reshaping management education for a globalized world," Frontiers in Education, vol. 10, 2025.

[42] X. Zhou, J. Zhang, and C. Chan, "Unveiling Students' Experiences and Perceptions of Artificial Intelligence Usage in Higher Education," Journal of University Teaching and Learning Practice, vol. 21, no. 6, 2024.

[43] S. Matook, Y. M. Wang, and M. Axelsen, "Experiential Learning for Citizen Developers," Business & Information Systems Engineering, vol. 67, no. 1, pp. 7–30, 2025.

[44] Y. Walter, "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, vol. 21, no. 1, 2024.

[45] M. MacArthur, S. Minnillo, L. Sperber, C. Whithaus, and N. Stillman, "GIFT-AI: teaching the game and leveling the field: Peer and AI Review + Reflection in a business writing course," Frontiers in Communication, vol. 10, 2025.

[46] A. A. S. Mohammad and S. Muhammad, "The Role of Artificial Intelligence Integration in Achieving AACSB Accreditation Standards," in Advances in computational intelligence and robotics, pp. 223–242, IGI Global, 2025.

[47] M. Lungu, "The Coding Manual for Qualitative Researchers," American Journal of Qualitative Research, vol. 6, no. 1, pp. 232–237, 2022.

[48] S. S. Ballal, "THE SYNTACTIC FACADE: A THEORETICAL PERSPECTIVE ON COGNITIVE ATROPHY IN AI-AUGMENTED LEARNING," Zenodo, 2026.

[49] R. Kop and A. R. Hill, "Connectivism: Learning theory of the future or vestige of the past?" The International Review of Research in Open and Distributed Learning, vol. 9, no. 3, 2008.

[50] T. Mzili, I. Mzili, A. Abatal, Z. Oughannou, A. K. Arya, and M. Kurdi, "AI Hallucinations in Retrieval- Augmented and Generative Systems: A Rigorous Review of Definitions, Failure Mechanisms, Evaluation, and Mitigation Strategies," EDRAAK, pp. 41–51, 2026.

[51] L. Sperber, M. MacArthur, S. Minnillo, N. Stillman, and C. Whithaus, "Peer and AI Review + Reflection (PAIRR): A human-centered approach to formative assessment," eScholarship, vol. 76, p. 102921, 2025.

[52] M. Peters, "Leveraging Generative Artificial Intelligence with Transparency: Enhancing Academic Integrity in Higher Education," Journal of Scholarly Publishing, vol. 56, no. 2, pp. 510–528, 2025.

[53] M. Sarıtepeci and H. Y. Durak, "Effectiveness of artificial intelligence integration in design-based learning on design thinking mindset, creative and reflective thinking skills: An experimental study," Education and Information Technologies, vol. 29, no. 18, pp. 25175–25209, 2024.

From Entrepreneurial Practice to Entrepreneurship Education: A Situated Experiential Framework for Critical Digital Literacy Development

Thao Laavunen

Åbo Akademi University, Turku, Finland, thao.laavunen@abo.fi

Abstract

This research explores how entrepreneurs develop critical literacy to manage the complexities of modern digital information environments. Drawing on qualitative interviews with local and immigrant business owners in Finland, the study identifies that while formal training provides structure, most entrepreneurs develop these essential skills through self-directed, practice-based learning. Entrepreneurs value education when it is practical, context-sensitive, timely, and directly applicable to their venture, industry, and stage of development. Based on these findings, the study proposes a Situated Experiential Framework for Critical Digital Literacy, linking a recurring entrepreneurial learning cycle with the situational dimensions that shape the critical literacy practices and development. The framework contributes to higher-education entrepreneurship programs by suggesting that critical literacy should be embedded in specific business problems and environments rather than taught only as a transferable technical skill.

Keywords: *critical literacy, information literacy, digital literacy, literacy development, literacy training, entrepreneurs, digital information, digital environment, entrepreneurial education, higher education*

ISSN 2991-9371

1. Introduction

Entrepreneurship education prepares learners to identify opportunities, make decisions under uncertainty, and navigate abundant, fragmented, and uneven-quality digital information [1]-[4]. Future entrepreneurs need not only access information about markets, customers, competitors, technologies, regulations, and financing, but also to evaluate its credibility, interpret its relevance, and apply it to particular business contexts [5]-[7]. These capabilities can be understood through the concept of critical literacy (CL). In this study, critical literacy encompasses information and digital literacy in digital environments and emphasizes the critical understanding of information environments and appropriate action within them [8], [9], [10]. Critical literacy is therefore relevant not only to entrepreneurs' current business activities, but also to the design of entrepreneurship education intended to prepare learners for information-intensive professional practice. Existing research has emphasized the need to incorporate information and digital literacy into entrepreneurship education and training programs [2]-[3]. Studies have also associated such education with entrepreneurial competence, readiness, and intention [11-12]. However, much of this literature focuses on students, educational interventions, or program outcomes. Less attention has been paid to what higher-education entrepreneurship programs can learn from practicing entrepreneurs' experiences of developing critical literacy through everyday business activity. As a result, two important gaps remain. First, although prior work establishes the importance of critical literacies, limited attention has been paid to how entrepreneurs develop CL and enact it in practice [13]-[14]. Second, while training is frequently proposed as a practical implication in the literature, entrepreneurs' own perspectives on the relevance and expectations of CL training remain underexplored. By examining these gaps, this study aims to understand how entrepreneurs develop CL and to translate the empirical results into implications for the design of higher-education entrepreneurship programs. The study, therefore, addresses the following research question: *How do entrepreneurs develop critical digital literacy, and how can their learning experiences and perceptions of formal training inform the design of higher-education entrepreneurship programs?*

2. Literature Review

Researchers generally agree that critical literacy does not develop in isolation. It is shaped by the environment in which people encounter information and by what they need to do with that information. Bawden [15], for example, views literacy as a context-dependent capability rather than a fixed set of technical skills. Choi et al. [16] similarly describe information literacy as "learning how to learn." From this perspective, people need to understand the expectations of their field, the resources available to them, and the situation in which information will be used before they can decide whether it is reliable, relevant, or useful. Critical literacy, therefore, develops as people learn to make sense of information within particular social, professional, and digital settings [17]-[19], [26]. This view is consistent with practice-based approaches to information behaviors. Jarrahi and Thomson [20] argued that information practices emerge through people's interactions with their environment. Lloyd's [21] concept of "literacies of

information" further develops this idea by showing that people become information-literate through participation in what she calls "information landscapes." Within these landscapes, individuals gradually learn which sources, activities, forms of knowledge, and ways of knowing are recognized as useful or trustworthy. This learning involves both individual interpretation and shared understanding. People draw on their own experiences, but they also learn from others' practices, expectations, and judgments. In entrepreneurship, this kind of learning is often closely connected to everyday practice. Politis [31] conceptualizes entrepreneurial learning as the transformation of prior experience into knowledge that informs subsequent entrepreneurial action. Entrepreneurs regularly face unfamiliar problems, incomplete information, and changing digital tools. As a result, they may develop CL by trying different approaches, comparing sources, seeking advice from others, and learning from the consequences of their previous decisions. Reviewing research on experiential learning in entrepreneurship education, Motta and Galina [32] found that experiential approaches can support the development of entrepreneurial skills and competencies. Studies have linked the development of CL with lifelong learning [14], [22]. Formal education may provide useful concepts and methods, but practical judgment is often built over time through problem-solving, experimentation, interaction, and reflection. Entrepreneurial education research also showed that entrepreneurs' CL are better developed through practical tasks than through general instruction alone [2], [4].

3. Methodology

This study employs a qualitative research design, using semi-structured interviews with entrepreneurs in Finland. By doing so, the study examines how entrepreneurs developed their critical literacy (CL) capabilities over time, and how training such capabilities is perceived. The design of the interview questions was guided by the study's research questions to examine the development of CL in information-seeking activities and also collect entrepreneurs' perceptions of the role of training in CL development [23]. This study includes both native Finnish and immigration entrepreneurs to highlight the shared experiences among business owners with different background but operating within the same Finnish entrepreneurial environment. We recruited our interviewees through the snowball sampling method in which initial participants referred others from their networks [24]. The initial referral process resulted in an excess number of Vietnamese entrepreneurs, thereby, this group became the focal immigrant cohort of this study. The interview candidates were selected based on these criteria: (i) they are either native Finnish or Vietnamese immigrants living in Finland; (ii) they must own or co-own a business in Finland, which includes both sole traders and various forms of business entities. The final sample included 19 entrepreneurs. The interview lasted approximately 90 minutes via Zoom or Google Meet. The interview length provided a comprehensive exploration of the research questions. The interviews are conducted mainly in English. On rare occasions, Vietnamese words or expressions were used to facilitate the discussions. Each interview was audio-recorded (with consent) and transcribed. Ethical transparency was maintained, with participants fully informed about data handling. Table 1 presents the demographic information of the participants.

Category	Immigrant Entrepreneurs
Number Of Respondents	n=19.
Location	Helsinki (n=12); Turku (n=4); Tampere (n=1); Joensuu (n=1); Hyvinkä (n=1).
Gender	Female (n=8); Male (n=11).
Age	28–34 years (n=7); 35–44 years (n=10); 45 years and above (n=2).
Educational Level	High school graduate (n=4); Bachelor's degree (n=8); Master's degree (n=7).
Language Skill	English (n=19); basic to intermediate Finnish (n=4); native Finnish (n=9).
Business Sector	Architecture photography (n=1); Education (n=1); Farming (n=1); Fashion (n=1); IT[1]/Educational game design (n=5); Marketing (n=4); Music (n=2); Real estate (n=2); Recycling (n=1); Restaurant (n=1).
Business Type	Limited Company (n=11); Sole trader (n=7); Still deciding (n=1).
Operational Duration	Less than 1 year (n=3); 1-5 years (n=8); more than 5 years -10 years (n=5); more than 10 years (n=3).

Table 1. Demographic information of the respondents

´We began the data analysis by cleaning the data to remove extraneous words and noise that might have been inadvertently captured during transcription. Vietnamese interview sections were translated where necessary to improve the readability and the translation is confirmed by the interviewees. The qualitative data collected from the interviews were uploaded to NVivo 15 for systematic organization. We followed Braun and Clarke's [25] six-phase framework for thematic analysis. Coding was conducted by one researcher under the supervision of the research supervisor. The analysis was inductive, and no predefined codebook was used. Initial codes were developed from participants' responses and iteratively reviewed and refined as the analysis progressed. Related codes were subsequently grouped into subthemes and broader themes, which were repeatedly checked against the coded extracts and the full dataset for coherence and consistency. Coding decisions and the developing thematic structure were discussed with the supervisor throughout the analytical process. The study also followed established criteria for qualitative rigor, including credibility, transferability, dependability, and confirmability [30] (see Table 2).

Credibility was supported by grounding the analysis closely in participants' own accounts of their entrepreneurial experiences and information practices.
Transferability was supported by providing contextual information about the study setting, participant groups, and business backgrounds. This allows readers to assess whether and how the findings may be relevant to other entrepreneurial or digital information contexts.
Dependability was supported by documenting the research process, including participant recruitment, interview procedures, transcription, coding, and theme development. This provided a clear record of how the data was collected and analyzed.
Confirmability was supported by maintaining a close connection between the data and the interpretations presented. Codes and themes were reviewed against the interview extracts, and representative quotations were used to ground the findings in participants' accounts. Reflexive attention was also given to the researcher's role in interpretation, with supervisory discussions used to check and refine analytical decisions.

Table 2. Strategies for Establishing Trustworthiness in Data Analysis

4. Results

4.1 Entrepreneurs' Critical Literacy Development

Participants described two main pathways through which they developed critical digital literacy: mixed learning, which combined formal education with practical experience, and predominantly self-directed learning. Nine of the 19 participants followed a mixed pathway, while the remaining participants relied mainly on learning through everyday entrepreneurial practice. Participants following a mixed pathway referred to university education, entrepreneurship workshops, incubator programs, and professional or industry-specific courses. Some believed that formal education had helped them organize their approach to information and navigate complex information environments more efficiently. Interviewee 15 explained that information seeking was often included in the training they attended: "*Every time I participate in training, a part of it is about finding information on that topic*." Interviewee 3 similarly noted that training could reduce confusion and wasted effort when dealing with numerous information sources, *"There are so many sources... it can be very confusing... without proper training and knowing how to search properly, you waste a lot of time and still don't find the right answers.*"

Training was also considered useful for developing the ability to interpret and apply information. Interviewee 9 explained that professional education had taught her not only how to access data, but also how to analyze reports and apply their findings to business decisions. Incubator and start-up programs were valued for similar reasons by other interviewees. In addition to providing information, they helped participants identify reliable sources, avoid information overload, and gain access to relevant professional networks. However, the influence of formal education varied across participants. Some reported that previous courses had contributed little to their everyday information practices. Interviewee 1, for example, found that university courses focused more on formal research methods than on the information problems she

encountered in her business. Interviewee 17 also reported that attending several courses had not substantially changed his established ways of working. This issue seems to be often connected to a perceived mismatch between the courses' content and the practical demands of running a business. Furthermore, participants described developing their abilities through repeated searching, trial-and-error, problem-solving, and continuous interaction with digital information. Interviewee 6 explained how this process had gradually shaped both his searching and evaluation practices: "I've *gained experience from doing research and Google search…I list out what the problem is, identify the problem or the topic that I want to have information about, and then just list out the keywords, try those keywords on Google, and get the information I need. [To validate the information], I go to a website and see how the information is distributed… do they have a source, what year is it from, who is the author… I've gained this through experience… just trying keywords on Google and finding what I need."*

Participants refined their understanding through interaction with peers, professional networks, and online communities. These interactions helped them clarify uncertain information, compare interpretations, and validate whether information was applicable to their situation. Interviewee 3, for example, emphasized the value of speaking with "people who have done the same thing," describing this as the "*best conversation*" because it allowed him to ask an "*unlimited number of questions*." Similarly, Interviewee 8 explained that trust-based communities were valuable because meeting members in person helped her identify "*like-minded people*" with whom she could exchange knowledge and even develop business relationships. These networks are used for validation or to produce shared interpretation, not merely to collect information. Interviewee 11 observed a growing culture of collaboration among entrepreneurs, "*CEO level, directors, and business owners, we tend to collaborate more…. We have talked about, for example, this concept type of thing, and other people have looked at our numbers and said that we have an idea what to do with you"*. For many participants, learning was triggered by an immediate business need. Interviewee 4 described learning as something that occurred "only when I have a demand to find the answer." Similarly, other participants emphasized that their literacy was largely self-developed through practice and experience, describing how they had *"pretty much learned everything by [themselves]"* (Interviewee 16) and how they had developed their approach *"through years of experience,"* with a focus on fact-based decision-making (Interviewee 14). Across both learning pathways, experiential and self-directed learning played a central role. For those who valued formal education, learning remained closely linked to experimentation, especially when encountering emerging technologies. Interviewee 9 explained that using tools such as ChatGPT required active experimentation, repeated questioning, and critical refinement of the output before it could be applied. This suggests that engagement with a new digital tool does not automatically produce literacy. Rather, CL develops through the iterative process in which entrepreneurs actively learn from their experiences.

4.2 Perceptions of Training for Developing CL

In this study, participants expressed different views about the essentials of formal education for developing CL. Some regarded training as necessary, others were skeptical, and several believed its usefulness was conditional. These views were shaped not only by participants' general attitudes toward education, but also by their previous training experiences and the extent to

which the content was relevant to their business activities. A group of participants highlighted the role of training in improving information practices. Particularly, it provides guidance and support for more structured approaches to information searching. It is particularly helpful for entrepreneurs to avoid time-consuming trial-and-error processes. As Interviewee 15 explained, training helped him to "*save a lot of time*", enabling them to perform tasks that would otherwise be difficult without "*tips and training in that space*". Similarly, Interviewee 3 emphasized that training can help individuals "*search better and more efficiently… and validate the information*". Participants also suggested that training could introduce them to more effective information-seeking methods. As Interviewee 12 noted, although his skills were largely self-developed, he believed others "have done much better" and expressed willingness to learn from structured courses. Entrepreneurs' view on the necessity of formal training is presented from Interviewee 9's statement: "*I always believe that a person without good education, we don't know how to do a piece properly…. And when you don't know, you make it wrong, you have to take more time to research again, do everything again, and it may be wrong again. So, I think that training and knowledge are really needed.*". Interviewee 15 also emphasized the importance of training when comparing it to self-learning practices, *"Self-learning is not always a very good way to learn everything, especially you might not even know what you should be looking for, what kind of information you should be looking for"*. In contrast, several participants perceived training as unnecessary, particularly those who relied heavily on experience-based approaches. For example, Interviewee 17 stated that they could manage information "*quite well… no problems at all*," relying on the structure of online information environments. Similarly, Interviewee 18 expressed skepticism about the relevance of training, arguing that general courses would not address his specific needs and that he could "*do it myself*", or strongly rejected the idea of training, as Interviewee 4 explained, "*…only when I have a question in mind… otherwise I am not interested*." These responses suggest that skepticism was not always directed toward learning itself. Instead, participants often questioned whether formal education offered sufficient value for the time and effort required. Meanwhile, some of them present a more neutral view, suggesting that the value of training depends on timing, relevance, and individual needs. In particular, they consider that training is more useful for early-stage entrepreneurs than for those with more experience. Interviewee 7, for example, noted that training can be helpful "*when you're starting your business…and without any starting point"* to understand unfamiliar concepts and navigate information for the first time.

Meanwhile, others questioned the time and effort constraints given the demands of running a business, as Interviewee 2's remark, "*So even I know that I will learn something, but then is that worth spending one day studying it? I see the value, but still I always have other important stuff to do, and to focus on the day and the time is limited.*" Participants' expectations of future education were shaped particularly by relevance, practicality, and time efficiency. Training was seen as more useful when it addressed a specific industry, business stage, regulatory issue, or current information needs. Interviewee 14 summarized the importance of timing by stating, "*If training is not useful at the moment, it's forgettable.*" Opportunities for interaction were another valued feature. Some participants viewed training not only as a source of knowledge but also as a setting for meeting peers, discussing information problems, and developing possible collaborations. As a result, participants generally preferred flexible, topic-specific, and practice-oriented formats rather than standardized courses. Interviewee 15 explained: "*It should be divided into different topics, and we learn about how to find the proper information and then discuss or share our experience of how you get information…because different topics require different types of information sources, and the way I seek information is different based on the topic."*

5. Discussion

This study examines how entrepreneurs develop critical digital literacy through everyday business practice and how they perceive formal training intended to support their development. The study shows that CL is enacted through self-directed, experiential learning or a mixed approach merging self-directed and training-guided learning. CL development is a contextual and practice-based process. Entrepreneurs actively acquire CL skills through trial-and-error, problem-solving, and real business challenges [13]-[14], [22]. Entrepreneurs in this study also reveal the role of the information environment in shaping their CL development, as they employ different tactics to overcome challenges and exploit the advantages of each environment [17]-[18], [21]. The information environment was therefore not merely the background to learning. It acted as an informal curriculum that continuously presented entrepreneurs with problems, feedback, risks, and opportunities. Such findings suggest that the importance of CL development lies in both "learning by doing" and in understanding the conditions under which experience becomes critical judgment [14], [16], [22]. Furthermore, the findings reveal the social dimensions of CL development, as it often depends on interactions with social networks, such as observing peers' discussions, seeking peers' validation, and participating in collaborative environments. It is also material because information is accessed through tools, platforms, documents, analytics systems, financial records, and digital media [17], [20], [21]. However, entrepreneurs' accounts in this study challenge the discussion on the importance of training for the development of CL. Although some of them acknowledged that training helped them structure and improve their methods for finding information, and adopt new technologies [4], [11]. They still favor self-directed study over official training and question the need for training for their personal and entrepreneurial growth. The main concern entrepreneurs have is the content of the training, whether it matches their needs and is worth the effort and time spent. Choi [26] identified a tension between information literacy as a set of transferable competencies and context-specific practice; this study clearly leaned more toward context-specific practice among entrepreneurs. Entrepreneurs expect the training to reflect real-world situations, and, through practice, they can easily apply it to their business practices and outcomes. Generally, this study's findings argue that CL cannot be reduced to a stable set of searching, technical, or evaluation skills that can be transferred unchanged across situations [12]-[13], [19]. Participants' accounts show that judgments about reliability and usefulness depend on the type of business problem, the source of information, the industry, the venture stage, and the consequences of acting on the information. What counted as an appropriate source or method in one situation did not necessarily work in another. Based on these findings, this study proposes a Situated Experiential Framework for examining the development of Critical Literacy and informing designs for the literacy courses in Entrepreneurship Education (Figure 1). The framework integrates two related dimensions of entrepreneurial learning. The inner layer represents the recurring process through which entrepreneurs develop critical judgment. The outer layer identifies the situational conditions that shape this process and influence whether experience becomes meaningful learning. These dimensions emerged inductively from recurring patterns across the interviews. Table 3 illustrates how the two layers of the proposed framework are grounded in the coded interview data by linking each component to representative codes and participant extracts.

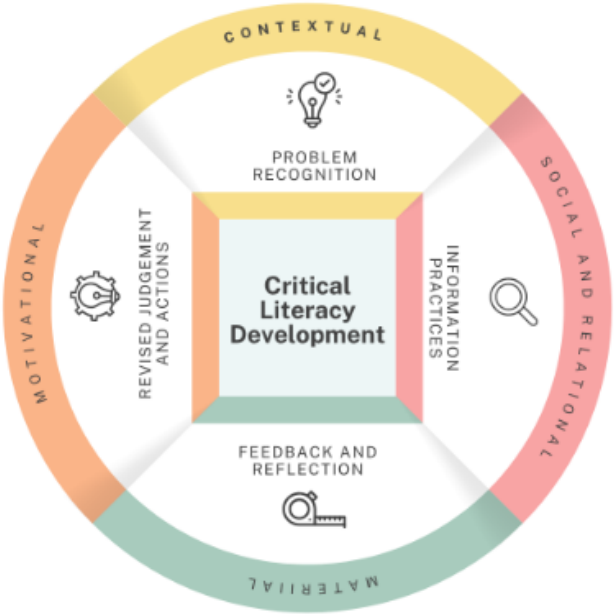

Figure 1. Situated Experimental Framework of Critical Literarcy Development

The entrepreneurial learning process begins when an entrepreneur encounters a specific information problem. This is followed by information practices, like information seeking, source comparison, contextual interpretation, and the application or testing of information through action [11]-[13]. The outcomes of the decision provide feedback, prompting reflection, and finally revision of previous judgments and actions. Through repeated engagement in this cycle, entrepreneurs gradually learn which sources are credible, which information is relevant, how to adapt information to a particular context, and what risks may result from acting on it [14, 18, 20-21, 27, 31-32]. However, experience does not automatically produce critical literacy. The cycle is shaped by contextual, social, material, and motivational conditions [8-10, 17, 21]. Judgments depend on the industry, location, venture stage, and nature of the business problem. They are also developed through interaction with peers, mentors, professional communities, and other trusted actors. At the same time, information practices are shaped by the surrounding environments and emerging technologies. It is also need-driven and becomes especially meaningful when entrepreneurs face immediate decisions with practical consequences.

Dimensions	Illustrative Codes from the Analysis	Representative Quotes
Experiential Process		
Problem Recognition	problem identification, verifying stakeholders (customers, employees, partners), planning business strategy, business operations (legal, forming the business, taxation), monitoring the market	*"I list out what the problem is, identify the problem or the topic that I want to have information about" (Interviewee 6)* *"We search on the internet for Google, supplier, or partners, or clients, customers. We try to keep on track of what's happening and try to find business opportunities that way."* (Interviewee 17)
Information Practices	keywords, source transparency, information filtering, cross-checking, cross-checking, background research, balancing online and offline information	*"It's also very important to use the right keyword when searching for information... [to] get the right information."* (Interviewee 6) *"I would cross reference to see if the things that the person is talking about or the company is talking about is roughly the same as other people are talking about"* (Interviewee 15) *"I tried to search from the website first, but sometimes the language is hard to understand, so I need to call. But luckily my accountant and her company is providing mostly the info for those."* (Interviewee 16)
Feedback and reflection	direct working experiences, testing, prior knowledge	*"After I test the advice and I see the good result, and it's confirmed that the group is trustable. This is a feedback loop for you to get the information, so just test information and then go back for more information*". (Interviewee 8) *[Information from the governmental website] is quite detailed, though it does not give me a full picture. The full picture comes from my experience more. That means, to start a business, what kinds of stuff I need to do. It is already on my mind before I get the information from the government." (Interviewee 9)*
Revised judgements and actions	contextual application, analytical use, learning by doing	*"It was difficult because we did not have a lot of experience reading legal information, for example, and it was difficult. We had to think about it a lot, and we had to ask for help from the tax agency, for example. But nowadays, it's a lot easier when I have years of experience reading this kind of legal information, and we also have better sources to ask for help if we need for help."* (Interviewee 15)
Situational Conditions		
Contextual	Finnish information environment, hidden information, endorsement, contextual application	*"In Finland, everything is really transparent in terms of legal information and contact points. So, they always have contact with the main people who are responsible for that area."* (Interviewee 10) "There is no such a fake news going on or misleading news going on that try to manipulating the users/ audiences." (Interviewee 14)
Social and relational	networking, community of knowledge, peer validation, referrals, interactive knowledge exchange, endorsement	*"The bigger connections you have, then the more people that could solve the problem"* (Interviewee 2) *"Reference was the main source, from my network or friends. I had a couple of entrepreneurial friends, and I ask a reference from my accountant and lawyer"* (Interviewee 14)
Material	time-consumption, easy access, Google or website algorithms	"I have other important stuff to do during the day. So, yeah, I know that I will learn something there. But then that's worth spending one day studying it. It's a different story." (Interviewee 2) *"Google also nowadays has changed their algorithm that is so much helpful and easier to reach to the right information very quickly as well."* (Interviewee 11)
Motivational	type of task, urgency of information need, personal learning style, personal responsibility,	*"I don't think that it's for me. Because I'm the person that I'm only have the urge [to learn new thing] when I'm facing trouble."* (Interviewee 4) *"It depends very much on the topic... if I need help with legal matters...* [or]*... growing the business and marketing...* the responsibility still lies with me and whatever decisions I make after they give the legal counsel" (Interviewee 15)

Table 3. Dimensions and representative quotes of the framework

In entrepreneurship education, this framework suggests that students should be given opportunities to move through the learning cycle while working in realistic, contextual conditions. Programs should incorporate real-world information problems, updated digital tools, feedback, and structured reflection. Formal education can therefore add value not by replacing experiential learning, but by making it more systematic, more critically informed, and less dependent on costly trial-and-error.

6. Conclusion

Theoretically, this study contributes to the entrepreneurial education that CL functions as an antecedent and an outcome of entrepreneurial daily operations. Entrepreneurs need CL in order to search for, evaluate, interpret, and use digital information effectively for their daily tasks. At the same time, they develop CL through repeated engagement with information tasks and information environments [17]-[18], [20]. CL, therefore, should be understood as a part of the entrepreneurial lifelong learning process [3], [14], [22]. Extending this perspective, CL development is co-constructed with the information environment, linking individual-level practices with broader system-level dynamics [17]. The proposed Situated Experiential Framework captures this process by linking the entrepreneurial learning cycle with the situational dimensions that shape it. The framework therefore bridges CL as a general competence and as a context-specific practice, while providing also potential approach for entrepreneurial educators to design and market their CL development programs. The finding also raises the tension between the effectiveness and attractiveness of training to entrepreneurs. While entrepreneurs with prior training experience tend to view training as valuable for developing their literacy, those who relied primarily on self-directed learning tend to express more ambivalent attitudes toward training. This indicates a potential gap in the perceived usefulness of training among this group. This suggests that program designers and educators critically consider how they market their program and to what audiences they are targeting [1]-[2], 12]. They should take into account what motivates or even demotivates entrepreneurs to participate in training and thus design the training structure that better addresses entrepreneurs' needs, expectations, and requirements [4]. For example, entrepreneurs in this study value the convenience of the study in terms of time and are also interested in the opportunity to network through such events. Educators can also consider including a measurement for the CL training outcomes suggested by Detlor et al. [1]. These can be reflected across multiple layers, including psychological (e.g., confidence and awareness), behavioral (e.g., the ability to apply skills), and benefit outcomes (e.g., improved ability to locate, evaluate, and create information). This framework provides a useful basis for evaluating entrepreneurship training by linking program design to measurable learning outcomes. Incorporating such an evaluative approach may enhance the attractiveness of training and encourage greater engagement among entrepreneurs once it shows that the potential benefits outweigh the opportunity cost of participation. Furthermore, CL is also actively enacted through the information environment. This suggests that while users need to develop the ability to function well, it also assigns the responsibility for support to policymakers, platform designers, and information providers. It is important to design the structures, rules, and practices that protect the information environment through transparency and ethical governance.

This study also presents several limitations that open avenues for future research. Empirically, we acknowledge the limitations of the study, which specifically addresses the immigration of only one group, i.e., Vietnamese entrepreneurs in Finland. The use of a snowball sampling method may also introduce selection bias, potentially limiting the diversity of perspectives captured. Furthermore, our study focuses on shared patterns between the two cohorts. We did not explicitly examine the role of demographic factors, such as gender, age, educational background, or business sector, though these factors have been shown to influence information behaviors in other studies [27]-[28]. Future research should therefore broaden the scope of the study and include more diverse immigrant groups to better understand how cultural, sectoral, and demographic factors shape CL development. As the study emphasizes the role of the information environment, the future study should consider extending its scope also to other countries where the information environment is less institutionally guarded and generally different from Finland. A comparative study should also be conducted to examine how demographic factors influence entrepreneurs' development of CL. Additionally, the entrepreneurial CL development process is likely to develop alongside entrepreneurs' business activities, digital experiences, and changing environments. A cross-sectional approach, which captures entrepreneurs' accounts of CL practices at a single point in time, may therefore lead to a misperception of how these practices evolve over time [29]. In addition, the findings rely on self-reported perceptions, which may not fully reflect actual practices. We also acknowledge our personal subjectivities and potential biases in the process of collecting, organizing, and analyzing the data in this dissertation. Furthermore, this study recognized the tension between formal training and experiential learning. This raises questions about whether the offer of formal training adequately aligns with entrepreneurs' expectations and conditions. Future research, therefore, should compare entrepreneurs who participate in formal training with those who primarily develop CL through everyday business practice, thus studying how different learning pathways would eventually influence the application and outcomes of CL practices.

7. Acknowledgements

The authors gratefully acknowledge the generous financial support provided by Liikesivistysrahasto for the research, and the Åbo Akademi University Foundation for participation in this conference.

References

[1] B. Detlor, H. Julien, T. La Rose and A. Serenko, "Community-led digital literacy training: Toward a conceptual framework", Journal of the American Society for Information Science and Technology, Volume 73, Issue 10, pg. 1387-1400, 2022.

[2] Y. Liang, R. Chen, H. Hong, S. Li and L. Han, "Shaping digital entrepreneurial intention in higher education: The role of entrepreneurship education, creativity, and digital literacy among Chinese university students", Journal of Innovation & Knowledge, Volume 10, Issue 5, 100788, 2025.

[3] V. Scuotto and M. Morellato, "Entrepreneurial Knowledge and Digital Competence: Keys for a Success of Student Entrepreneurship", Journal of The Knowledge Economy, Volume 4, Issue 3, pg. 293-303, 2013.

[4] A. L. Rodrigues, "Entrepreneurship Education Pedagogical Approaches in Higher Education", Education Sciences, Volume 13, Issue 9, 940, 2023.

[5] F. Ahmad, G. Widén and I. Huvila, "The impact of workplace information literacy on organizational innovation: An empirical study", International Journal of Information Management, Volume 51, 102041, 2020.

[6] T. Laavunen, J. A. Teixeira and G. Widén, "From challenges to insights and opportunities: Exploring how Finnish entrepreneurs leverage digital information", Journal of Information Science, Volume 0, Issue 0, 2026.

[7] P. G. Roetzel, "Information overload in the information age: A review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development", Business Research, Volume 12, Issue 2, pg. 479-522, 2019.

[8] M. Bozbiyik, L. Tompkins, T. Morton, P. Gerns, L. Kääntä, S. Sulkunen and L. Pérez-Fernández, "Comparative analysis of CL in lower secondary curricula in Finland, Italy and Spain: Implications for CLIL", Journal of Multilingual and Multicultural Development, Volume 47, Issue 4, pg. 2066-2083, 2026.

[9] E. Drabinski and E. Tewell, "Critical Information Literacy", The International Encyclopedia of Media Literacy, 2019. https://doi.org/10.1002/9781118978238.ieml0042

[10] E. Gagrčin, T. K. Naab and M. F. Grub, "Algorithmic media use and algorithm literacy: An integrative literature review", New Media & Society, Volume 28, Issue 1, pg. 423-447, 2026.

[11] L. Bosman, B. Kotla, C. Cuesta, N. Duhan and T. Oladepo, "The role of information literacy in promoting 'discovery' to cultivate the entrepreneurial mindset", Journal of International Education in Business, Volume 16, Issue 1, pg. 56-69, 2023.

[12] N. N. Islami, "The effect of digital literacy toward entrepreneur behaviors through students' intention entrepreneurship on Economics Education Study Program at Jember", IOP Conference Series: Earth and Environmental Science, Volume 243, Issue 1, 12084, 2019.

[13] S. Nikou, M. E. Brännback, T. P. Orrensalo and G. Widén, "Social media and entrepreneurship: Exploring the role of digital source selection and information literacy", Understanding Social Media and Entrepreneurship, pg. 29–46, 2020.

[14] V. Bolek, V. Kokles, M. Romanová and A. Zelina, "Information literacy of managers: Models and Factors", Journal of Business Economics and Management, Volume 19, Issue 5, pg. 722-741, 2018.

[15] D. Bawden, "Information and digital literacies: A review of concepts", Journal of Documentation, Volume 57, Issue 2, pg. 218-259, 2001.

[16] W. Choi, H. Bak, J. An, Y. Zhang and B. Stvilia, "College students' credibility assessments of GenAI-generated information for academic tasks: An interview study", Journal of the American Society for Information Science and Technology, Volume 76, Issue 6, pg. 867-883, 2025.

[17] E. F. Kuehn, "The information ecosystem concept in information literacy: A theoretical approach and definition", Journal of the American Society for Information Science and Technology, Volume 74, Issue 4, pg. 434-443, 2023.

[18] A. Lloyd, "Chasing Frankenstein's monster: Information literacy in the black box society", Journal of Documentation, Volume 75, Issue 6, pg. 1475-1485, 2019.

[19] L. Pangrazio, A. Godhe and A. G. L. Ledesma, "What is digital literacy? A comparative review of publications across three language contexts", E-learning and Digital Media, Volume 17, Issue 6, pg. 442-459, 2020.

[20] M. H. Jarrahi and L. Thomson, "The interplay between information practices and information context: The case of mobile knowledge workers", Journal of the American Society for Information Science and Technology, Volume 68, Issue 5, pg. 1073-1089, 2017.

[21] A. Lloyd, "Information literacy and literacies of information: A mid-range theory and model", Journal of Information Literacy, Volume 11, Issue 1, pg. 91, 2017.

[22] D. G. Erdogan, Ö. C. Güngören, N. B. Hamutoğlu, G. K. Uyanık and T. D. Tolaman, "The relationship between lifelong learning trends, digital literacy levels and usage of web 2.0 tools with social entrepreneurship characteristics", Croatian Journal of Education, Volume 21, Issue 1, pg. 45-76, 2019.

[23] N. K. Denzin and Y. S. Lincoln, "Introduction: The discipline and practice of qualitative research", The Sage Handbook of Qualitative Research, 4th ed., pg. 1-29, 2014

[24] F. Ghaljaie, M. Naderifar and H. Goli, "Snowball Sampling: A Purposeful Method of Sampling in Qualitative Research", Strides in Development of Medical Education (Online), Volume 14, Issue 3, 2017.

[25] V. Braun and V. Clarke, "Using thematic analysis in psychology", Qualitative Research in Psychology, Volume 3, Issue 2, pg. 77-101, 2006.

[26] J. Choi, "Aporias of information literacy: A critical review", Journal of Documentation, Volume 82, Issue 2, pg. 467-493, 2026.

[27] T. Orrensalo, M. Brännback and S. Nikou, "Entrepreneurs' digital information sources selection: A perspective on the impact of information literacy and generational differences", Information Literacy and the Digitalisation of the Workplace, pg. 121–144, 2023.

[1] A. Qazi, N. Hasan, O. Abayomi-Alli, G. Hardaker, R. Scherer, Y. Sarker, S. K. Paul and J. Z. Maitama, "Gender differences in information and communication technology use & skills: A systematic review and meta-analysis", Education and Information Technologies, Volume 27, Issue 3, pg. 4225-4258, 2022.

[1] X. Wang and Z. Cheng, "Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations", Chest, Volume 158, Issue 1S, pg. S65, 2020.

[1] J. Loh, "Inquiry into Issues of Trustworthiness and Quality in Narrative Studies: A Perspective", The Qualitative Report, Volume 18, Issue 33, pp. 1–15, 2013.

[1] D. Politis, "The Process of Entrepreneurial Learning: A Conceptual Framework", Entrepreneurship Theory and Practice, Volume 29, Issue 4, pp. 399–424, 2005.

[1] V. F. Motta and S. V. R. Galina, "Experiential Learning in Entrepreneurship Education: A Systematic Literature Review", Teaching and Teacher Education, Volume 121, Article 103919, 2023.

AI, Learning and Innovation in Higher Education

SESSION CHAIR:

Karmela Aleksić-Maslać

Cognitive Offloading in the Classroom: Evidence from Written and Oral Statistics Exams Before and After ChatGPT

Marija Ivanković

Zagreb School of Economics and Management, Croatia, mivankov1@zsem.hr

Abstract

Since the introduction of OpenAI's ChatGPT to the masses in late 2022, its effect on education and learning has been both debated and researched. Quantitative subjects, like mathematics and statistics, have traditionally relied on written tests and task-solving to examine knowledge, which makes them particularly exposed to a tool capable of solving such tasks in seconds. This study analyses the effect of ChatGPT on learning by examining written and oral exam results from the Statistics 1 course at the Zagreb School of Economics and Management, across eight academic years (2017/18–2025/26, excluding 2021/22 due to missing data), split into pre-ChatGPT (2017/18–2020/21) and post-ChatGPT (2022/23–2025/26) periods. Using the oral exam as a comparison condition less exposed to direct AI assistance, a difference-in-differences design was applied to isolate a ChatGPT-specific effect from general trends and two identified controls. Results show that written exam performance declined relative to oral exam performance following ChatGPT's emergence (did = −0.211, $p < 0.001$). Event-study estimates show this decline as a sharp shift concentrated at the point of ChatGPT's release rather than a gradual pre-existing trend. The pattern is consistent with a cognitive offloading mechanism, whereby AI-assisted preparation produces a shallower understanding that becomes visible specifically once the tool is no longer available, though individual-level AI usage was not directly observed.

Keywords: *ChatGPT; difference-in-differences; oral examination; cognitive offloading; academic integrity; quantitative assessment*

ISSN 2991-9371

1. Introduction

The rapid development of AI in recent years has both been a blessing and a curse. On one hand, research has shown positive sides of using AI in education.[For example, ChatGPT can enhance learning approaches, like flipped learning where students prepare for lessons by examining materials before classes, where AI serves as a virtual tutor in assisting students with independent studies by answering their questions [1], [2]. On the other hand, faculty has faced many challenges. First one relates to the accuracy and reliability of these kinds of Large Language Models (LLMs) as they are trained on large data sets and has been proved to be biased [3] and provided incorrect information [4], [5]. This fact can lead to less and lower quality learning. But also, AI-powered tools like ChatGPT can answer complex questions correct, even providing explanations. With the increasing availability of these tools, students may be tempted to use them during exams or assignments to bypass traditional methods of studying and problem-solving [6]. This undermined the basic principles of academic integrity and goals, which are honesty, knowledge and learning.

As a results of this, educational institutions have been forced to reengineer existing methods of examinations, shifting more to in-person tests and oral exams. These formats are more resistant to AI use as they are conducted under direct supervision of proctors as students have no access to external devices or influences. Oral exams serve as an additional safe switch as students are forced to articulate and defend their reasoning in real-time, which makes shallow level or memorized answers easier to detect. This still however does not guarantee AI free results. Exam scores have traditionally been used as a tool of assessment, but the validity of those scores is tested in multiple ways. First way, while less likely, is still a direct threat as the possibility of AI assistance during the assignments itself is possible even despite the prohibitions. The second is more difficult to detect but has long-term lasting consequences, is that the tests fail to reflect genuine learning if AI tools are used heavily during preparation. This phenomenon is often called cognitive offloading [7] which can reduce the depth of engagement that consequently leads to shallower conceptual understanding that is visible only when the tool is no longer assessable, like in the proctored in-person exams or oral exams. Hence, the assumption is that this shallow understanding will be reflected in some way in the student's results. To test this, written and oral exam results from 340 Statistics 1 students at ZSEM (2017/18–2025/26, excluding 2021/22) were split into pre- and post-ChatGPT periods and compared using distributional and difference-in-differences analysis, with the oral exam serving as a comparison condition less exposed to direct AI assistance. The following section reviews the literature underpinning this design, covering both the effects of AI tools on learning and the role of oral versus written assessment in a post-ChatGPT context.

ISSN 2991-9371

2. Literature

2.1 AI, Learning and Cognitive Offloading

As mentioned previously, the introduction of AI into various spheres of our existence, including education has polarizing results. On one hand, ChatGPT can serve as a tutor, as research shows that ChatGPT assisted learning produced statistically significant learning gains compared to a no-help [8], which is specifically good for low-income students that can't afford real human tutors. Additionally, ChatGPT produced higher learning gains than human tutor authored hints, however the differences were not statistically significantly separable, and that there were no differences between human and AI tutorship [8]. Several other studies have shown similar results, as ChatGPT has caused moderately positive impact on students' academic achievement [9]. Furthermore, ChatGPT's responses were described as highly concordant, such that a human learner could easily follow the internal language, logic, and directionality of relationships contained within the explanation text [10].

On the other hand, one study examined the midterm exams of mathematics and engineering students on the University Dunaújváros, in Hungary. The results show that students that used ChatGPT had significantly lower average total scores compared to non-ChatGPT counterparts [11]. Additionally, students that used ChatGPT also showed lower engagement levels in online learning activities. Another study reveals also showed mixed results as ChatGPT had a positive impact on exam scores, but it did not improve assignment performance significantly [12]. Additionally, when assignments are customized to students, a performance gap arises as the non-AI group outperformed experimental group, which suggests that tailored, context-rich tasks are effective in minimizing AI-assisted cheating [12]. While ChatGPT supported performance in well-structured tasks, its utility was limited when the student needed to apply flexible reasoning or interpret problems independently, emphasizing the decline in critical thinking abilities.

These finding by itself is not as interesting as the reason behind why this occurs. Many researchers have examined the impact of AI tools on students' cognitive engagement during academic writing tasks and concluded that AI tools encourage students to become "lazy thinkers" [13]. Furthermore, students that did not use AI were more likely to report engaging in deep processing, reflecting a stronger commitment to understanding the task beyond surface-level comprehension. They also indicated exerting greater mental effort, suggesting higher levels of self-regulated thinking [13]. Absence of AI tools appeared support engagement, with participants demonstrating greater strategic flexibility by exploring multiple approaches to the problem. These findings suggest that, rather than enhancing mental effort, AI tools like ChatGPT may reduce learners' cognitive engagement when thoughtfully integrated into academic tasks [13]. The phenomenon described as cognitive offloading, i.e. the use of physical action to alter the information processing requirements of a task to reduce cognitive demand [7]. For this study, it can be interpreted as "outsourcing thinking" to AI to reduce the mental strain. Cognitive offloading can be beneficial in some cases, for example, individuals with impaired unaided cognitive ability [7], which includes memory and focus problems. However, cognitive offloading in education is beneficial only if the cognitive thinking is allocated productively and for a good purpose [7]. This is a clear warning on the allocation of thinking. While relief of learning can be

alleviated, the risk of outsourcing thinking completely to AI carries dangerous consequences. This is supported by [14] research that showed that people who do not use any form of AI exhibited the strongest, widest ranging neural networks, i.e. highest value of cognitive engagement and load, while LLM group showed the lowest results. No AI (i.e. Brain group) demonstrated higher memory recall, and re engagement, while the LLM group also fell behind in their ability to quote from the essays they wrote just minutes prior [14]. This and much other similar research point to an urgent problem in education of decreasing learning skills and abilities of critical thinking which is highly important, especially in quantitative subjects like mathematics and statistics.

2.2 Academic evaluation

Findings described above prove even further that educational institutions need to shift the way in which the students are both educated and examined. Many institutions have already made those changes by increasing the importance of oral exams. Oral examination offers a plethora of benefits. For instance, there is evidence that this dialogic form of assessment may help students express their relational understanding, especially in STEM subjects [15], [16]. In addition, in the right circumstances (for example introducing peer questioning, increased intrinsic motivation and familiarity with the structure) helps to adopt deep learning approaches [17]. Oral assessment can also mitigate the risk of academic integrity as it has this capacity, often missing in other forms of standardised or paper-based assessments that motivate students to engage in academic misconduct. [18], [19]. Students have also expressed the support for this kind of examination as this assessment method was better at bringing the material together and was also perceived as a learning experience, additionally students perceived oral examination to help more with conceptual understanding, written exams were predominantly perceived to assess computational methods. [20].

The importance of written exams is not to be forgotten as enunciation of complex abstract ideas might be easier in writing, especially for mathematics and statistics courses [21]. To conclude, oral assessments specifically for developing and measuring reflective skills, whilst abstract conceptual thinking should be assessed using traditional written formats [21]. Researchers also agree that oral exams tend to cause more anxiety, while written test are not as suitable for students with learning impairments, like dyslexia or dysgraphia [20], [21]. Hence why this research uses both written and oral examinations for the data analysis.

2.3 Hypothesis and research questions

Based on the conducted literature review and available data, the main research questions arise:

> *RQ1: Did the emergence of ChatGPT in November 2022 influence students' performance of proctored written exams (midterms) and oral exams in the same course Statistics 1?*
> *RQ2: Is the pattern of change consistent with a cognitive offloading mechanism, i.e. a decline concentrated in assessment formats without real-time AI access?*

Since the analysis for the research is quantitative in nature, a hypothesis needs to be stated and tested. Similarly to the research question, from the literature the hypothesis is drawn as:

> *H0: Introduction of ChatGPT did not change the performance on proctored written exams and oral exams.*
> *H1: Introduction of ChatGPT decreased the performance on proctored written exams and oral exams.*

3. Methodology

3.1 Data collection and research design

The data set consists of students' results from the Statistics 1 course at Zagreb School of Economics and Management (ZSEM) covering eight academic years (2017/18-2020/21, 2022/23-2025/26), with academic year 2021/22 excluded due to missing records. Three grade assessment components were collected for each student, two written midterm exams (MT1, MT2) and one oral examination (OE). The final grade is calculated based on the performance on these three components. All scores, both on written and oral examinations, were provided pre-scaled [0,1] interval. A small number of values exceeded 1.0 due to bonus points awarded during grading; these were capped at 1.0 before analysis. Since ChatGPT became widely available to the public in November 2022, and since Statistics 1 is taught in the summer semester (February–June), the sample was split at the academic-year level into a pre-ChatGPT period (2017/18–2020/21) and a post-ChatGPT period (2022/23–2025/26). This means the entire 2022/23 academic year already falls after ChatGPT's public release. Each period covers four academic years, which allows for a balanced comparison.

An important issue with this design is that the oral exam, used here as a comparison condition less exposed to direct AI assistance since it happens live and under supervision, was not a fully stable benchmark throughout the sample period. Two things changed alongside, independently from the emergence of ChatGPT. First, the weight of the oral exam in the final grade increased from 10% (2017/18–2023/24) to 20% (2024/25–2025/26). Second, the examiner responsible for the oral exam changed during the 2022/23 and 2023/24 academic years. Both changes could plausibly affect oral exam outcomes on their own, regardless of any AI-related effect, so both were built directly into the statistical models (Section 3.2) as control variables. This follows standard practice in quasi-experimental education research, where concurrent confounding events are modelled explicitly rather than assuming a clean natural experiment [22].

3.2 Statistical Analysis

A large share of students scored exactly the minimum (0) or maximum (1) value on their exams, which violates the assumptions of standard continuous distributions like the Normal distribution. This pattern was pronounced in the data: for instance, up to 41% of oral exam scores in a single year were exactly 1.0, and a non-trivial share of students across all three exam types scored exactly 0, likely reflecting non-attendance rather than an actual performance outcome. Instead of excluding these boundary cases, which would throw away meaningful information about non-participation and ceiling performance, scores were modelled using a zero-and-one-inflated Beta (ZOIB) distribution, following the general framework of [23], [24]. This approach has previously been applied to model non-selective university entrance exam scores [25], and to model bounded introductory-statistics exam scores containing a large share of perfect scores [26], which is more related to this research. For each combination of academic year and exam type, the ZOIB model was fitted via maximum likelihood and compared against a set of standard bounded and unbounded distributions using the Akaike Information Criterion (AIC) [27], to confirm the inflated model fit better than conventional alternatives before drawing any conclusions from it.

To test whether the shape of the score distribution differed between the pre- and post-ChatGPT periods, three tests were applied to each exam type separately, each answering a slightly different question. The two-sample Kolmogorov-Smirnov is used to compare the full empirical distribution of two samples as it is sensitive to any kind of distributional difference [28], as the ChatGPT-related effect could show up as a shift in the mean, in the variance, or in the share of extreme (0 or 1) scores, but not necessarily all three at once. Since this test does not specify the differences, the Mann-Whitney U test was additionally used to test directly whether the central tendency (median) differed between the pre- and post-ChatGPT periods. This test was chosen over the more common independent t-test because it does not assume normally distributed data, which fits the boundary-inflated, non-normal shape of the scores already confirmed by the ZOIB fitting step [29]. Finally, since a change in variance is itself a meaningful outcome and not just a nuisance to control for (for example, a narrowing spread of results could indicate that AI use is compressing the gap between weaker and stronger students), the Levene's test was used to test for equality of variances between the two periods, again without requiring the normality assumption an F-test would need [30].

To isolate a change attributable to the emergence of ChatGPT from a general secular trend affecting all exam types equally, a difference-in-differences (DiD) design was applied [22], [31] treating the two written midterms as the exposed condition and the oral exam as the comparison condition. The model used can be expressed like:

$$score = \beta_0 + \beta_1(treated) + \beta_2(post) + \beta_3(treated \times post) + \beta_4(oral_weight) + \beta_5(oral_prof_change) + \varepsilon \quad (1)$$

where *treated* indicates a written exam (MT1 or MT2), *post* indicates the post-ChatGPT period, *oral_weight* is the oral exam's weight in the final grade for that academic year, and *oral_prof_change* is an indicator for the 2022/23–2023/24 academic years during which the oral examiner changed. The coefficient of interest, β_3 *(the treated × post interaction)*, captures the differential change in written-exam performance relative to oral-exam performance following ChatGPT's

emergence, net of the general post-period trend and both identified control variables (*oral_weight* and *oral_prof_change*). Standard errors were clustered at the student level to account for the non-independence of a given student's three exam scores. The same specification was re-estimated with a binary zero/one outcome (logit link) to test whether the probability of a failing (0) or maximal (1) score changed differentially between exam types.

Full academic-year fixed effects were deliberately excluded from this specification, as *post* and *oral_weight* are themselves determined by academic year and would therefore be collinear with a saturated year-fixed-effects term. As a complementary, fully saturated robustness check, an event-study specification was performed. The score is regressed on variable *treated*, a full set of academic-year dummies, and their interaction was estimated to trace the *treated–control* gap separately for each academic year, without imposing a single pre/post breakpoint. This allows visual and statistical confirmation that any estimated effect coincides with the timing of ChatGPT's emergence rather than reflecting a pre-existing trend. As an additional robustness check, the core DiD model was re-estimated excluding the 2022/23–2023/24 academic years entirely, to confirm that the main result does not depend on the years affected by the oral examiner change. All statistical analyses were conducted in Python.

4. Findings

4.1 Distributional Fit

For each combination of academic year and exam type, the ZOIB model was compared against a set of standard distributions using AIC. The standard Beta distribution outperformed the ZOIB model in every single year-by-exam combination. This does not mean the boundary concentration described in Section 3.2 was absent from the data, but rather that the Beta distribution with shape parameters below one is itself able to produce a U-shaped density that piles up probability mass near both 0 and 1 without needing a separate inflation term. Since the Beta distribution does this with two parameters instead of the four required by ZOIB it was penalized less under AIC and won the comparison despite the extra flexibility of the inflated model. This is a useful finding as it suggests the boundary concentration in these results is closer to a continuous U-shape (many students clustering near, but not necessarily exactly at, the extremes) than to a distinct subgroup of true non-participants mixed with a separate performing group.

Academic Year	Exam	n	Mean	Variance	π_0	π_1
17/18	MT1	22	0.552	0.115	0.182	0.000
18/19	MT1	32	0.712	0.095	0.125	0.062
19/20	MT1	41	0.686	0.120	0.171	0.049
20/21	MT1	21	0.774	0.035	0.000	0.048
22/23	MT1	48	0.527	0.121	0.208	0.021
23/24	MT1	53	0.325	0.091	0.245	0.019
24/25	MT1	57	0.506	0.125	0.246	0.000
25/26	MT1	70	0.486	0.087	0.171	0.000
17/18	MT2	22	0.485	0.105	0.182	0.000
18/19	MT2	32	0.623	0.110	0.156	0.125
19/20	MT2	41	0.533	0.119	0.244	0.024
20/21	MT2	21	0.643	0.079	0.048	0.048
22/23	MT2	48	0.455	0.100	0.208	0.042
23/24	MT2	53	0.335	0.079	0.264	0.000
24/25	MT2	57	0.377	0.088	0.263	0.018
25/26	MT2	70	0.363	0.069	0.243	0.000
17/18	OE	22	0.332	0.117	0.409	0.045
18/19	OE	32	0.703	0.140	0.188	0.406
19/20	OE	41	0.572	0.189	0.341	0.268
20/21	OE	19	0.795	0.051	0.000	0.368
22/23	OE	48	0.842	0.125	0.146	0.812
23/24	OE	53	0.811	0.153	0.189	0.811
24/25	OE	57	0.509	0.167	0.351	0.211
25/26	OE	70	0.686	0.121	0.186	0.214

Note: π_0 and π_1 denote the empirical proportion of students scoring exactly 0 and exactly 1, respectively. Across all year–exam combinations, a standard Beta distribution was preferred over the zero-and-one-inflated Beta (ZOIB) specification by AIC.

That said, the raw proportions of exact 0 and exact 1 scores (π_0 and π_1) are still informative descriptively. For MT1 and MT2, π_0 (proportion scoring exactly zero) tends to increase in the post-ChatGPT years — for instance, MT1 goes from 0% (2020/21) up to around 21–25% in most post-ChatGPT years. The oral exam shows a very different pattern: π_1 (proportion scoring exactly the maximum) jumps sharply to 81% in both 2022/23 and 2023/24, before dropping back down to around 21% in 2024/25–2025/26. As discussed in Section 3.1, this spike coincides with, and is most plausibly explained by, the change in oral examiner during those two years rather than by ChatGPT. Figure 1 visualizes these trajectories across the full sample period.

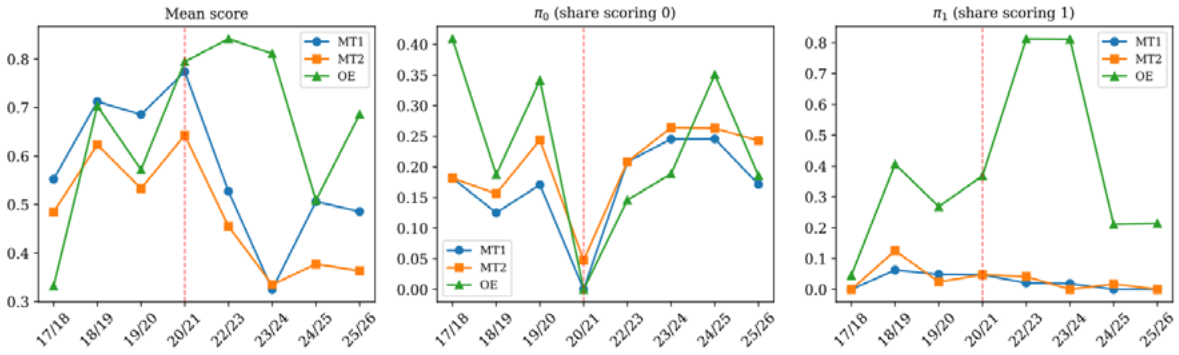

Figure 1. Trajectories of parameters

4.2 Pre and Post Chat GPT Comparison

Table 2 summarizes the distributional comparison between the pre- and post-ChatGPT periods, separately for each exam type. For both written midterms, the mean score dropped considerably: MT1 fell from 0.684 (pre) to 0.462 (post), and MT2 fell from 0.569 to 0.379. Both drops are highly statistically significant across every test applied — the Kolmogorov-Smirnov test (MT1: $p < 0.001$; MT2: $p < 0.001$), the Mann-Whitney U test (MT1: $p < 0.001$; MT2: $p < 0.001$), and the Welch t-test (MT1: $p < 0.001$; MT2: $p < 0.001$). Levene's test shows a significant increase in variance for MT1 ($p = 0.014$, variance rising from 0.103 to 0.111), while the variance change for MT2 was not significant ($p = 0.277$).

Exam	Mean (pre)	Mean (post)	Var (pre)	Var (post)	KS p	MWU p	Levene p
MT1	0.684	0.462	0.103	0.111	6.01e-10***	3.81e-10***	0.0135*
MT2	0.569	0.379	0.111	0.085	4.34e-09***	4.39e-08***	0.277
OE	0.600	0.704	0.163	0.158	0.00384**	0.00352**	0.253

Note: KS = two-sample Kolmogorov–Smirnov test; MWU = Mann–Whitney U test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2. Comparison of pre- and post-ChatGPT periods, by exam type

The oral exam moves in the opposite direction: the mean actually increased, from 0.600 (pre) to 0.704 (post), and this shift is also statistically significant (KS: $p = 0.004$; Mann-Whitney: $p < 0.001$; Welch t-test: $p = 0.025$), though the variance did not change significantly (Levene: $p = 0.253$). Given what is already known about the examiner change in 2022/23–2023/24, this increase in oral exam performance should not, on its own, be read as evidence against a ChatGPT effect as it can be driven by the same confound the DiD design was built to address. Looking at the boundary proportions specifically, the change in π_1 (proportion scoring the maximum) on the oral exam is highly significant (χ^2 test, $p < 0.001$), which lines up with the spike already noted in Section 4.1. The change in π_0 (proportion scoring zero) is not significant for any of the three exam types at the 5% level, though it is close to significance for MT1 ($p = 0.075$).

4.3 Difference-in-Differences (DiD) Results

The core DiD regression (Table 3) gives a treated × post coefficient of −0.211 ($p < 0.001$). This is the key result of the analysis **and it points that the** written exam performance dropped by an additional 0.211 points (on the [0,1] scale), relative to the oral exam, after the emergence of ChatGPT (over and above whatever general shift happened across all exam types in the same period, and after controlling for both the oral exam's changing weight in the final grade and the change in oral examiner). Neither the treated term (0.027, $p = 0.313$) nor the post term (0.003, $p = 0.958$) is significant on its own, **meaning** there was no meaningful baseline difference between written and oral exams before ChatGPT, and no general post-period shift affecting all exam types equally. The effect is specific to the interaction, exactly as the difference-in-differences design is meant to isolate. The *oral_prof_change* control is itself highly significant (0.221, $p <$

0.001), confirming that the examiner change had a real, independent effect on oral exam scores, separate from anything related to ChatGPT.

	Score (OLS)	$P(score = 0)$ (Logit)	$P(score = 1)$ (Logit)
Intercept	0.592***	−0.773*	−0.729
	(0.059)	(0.304)	(0.428)
treated	0.027	−0.653*	−2.059***
	(0.026)	(0.283)	(0.373)
post	0.003	0.190	−0.255
	(0.057)	(0.316)	(0.359)
did (treated × post)	−0.211***	0.455	−1.161†
	(0.038)	(0.366)	(0.625)
oral_weight	0.072	−3.025	−2.117
	(0.457)	(2.154)	(3.742)
oral_prof_change	0.221***	−0.635†	2.715***
	(0.052)	(0.339)	(0.348)
N	1030	1030	1030

Note: Standard errors in parentheses; clustered at the student level for the OLS model.
†$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 3. Difference-in-differences regression results

The same specification estimated on the probability of a zero score shows a positive but non-significant did coefficient (0.455, p = 0.215). This means that once the professor **changes** and other controls are accounted for, there is no strong evidence that complete failure became relatively more likely on written exams after ChatGPT. The equivalent model for the probability of a perfect score shows a negative, marginally significant did coefficient (−1.161, p = 0.063). Perfect scores became somewhat less likely on written exams relative to oral, after ChatGPT, though this result should be treated with caution given the p-value sits just above the conventional 5% threshold.

4.4 Robustness Checks

Excluding the two years affected by the oral examiner change (2022/23 and 2023/24) entirely and re-estimating the core model gives a *treated* × *post coefficient* of −0.201 (p < 0.001) — nearly identical to the −0.211 obtained in the full model. This means the main finding does not depend on those two disputed years and is not a result of the examiner change. The event-study specification, which estimates the treated effect (written minus oral) separately for every academic year instead of collapsing everything into a single pre/post-split, adds further support. In the pre-ChatGPT years, the *treated* effect is moderate in size: −0.222 (2018/19), −0.149 (2019/20), and −0.273 (2020/21), all statistically significant.

	Score (OLS)
Intercept	0.552***
	(0.066)
treated	0.027
	(0.026)
post	−0.019
	(0.062)
did (treated × post)	−0.201***
	(0.033)
oral_weight	0.476
	(0.543)
N	727

Note: Standard errors (clustered at student level) in parentheses. ***$p < 0.001$. The 2022/23–2023/24 academic years, affected by the oral examiner change, are excluded entirely. The *did* coefficient (−0.201) is nearly identical to the full-sample estimate (−0.211 in Table 3), indicating the main finding does not depend on these two years.

Table 4. Robutness check: core model excluding 2022/23 2023/24

Immediately after ChatGPT's emergence, the gap widens sharply: −0.537 (2022/23) and −0.668 (2023/24). It narrows somewhat in the following years but never returns to pre-ChatGPT levels: −0.254 (2024/25) and −0.448 (2025/26). This pattern of a stable gap before the emergence of ChatGPT, followed by a sudden widening exactly at the point of transition is consistent with a level shift tied to ChatGPT's emergence, rather than a gradual pre-existing trend that would have shown up gradually across the earlier years as well.

As a further robustness check, an event-study specification was estimated, tracing the written-versus-oral exam gap separately for each academic year rather than collapsing all post-ChatGPT years into a single average effect. Figure 2 plots the resulting coefficients, each representing the difference between written and oral exam performance each year, relative to the 2017/18 baseline. In the years preceding ChatGPT's emergence, this gap is relatively stable, ranging from -0.15to -0.27, with overlapping confidence intervals across all three years. Immediately following ChatGPT's public release, the gap widens sharply to -0.54(2022/23) and -0.67(2023/24), clearly outside the range observed in any pre-ChatGPT year. In the two most recent academic years, the gap narrows somewhat (-0.25 and -0.45, respectively) but does not return to its pre-ChatGPT level. This pattern is very informative. A gradual, pre-existing trend unrelated to ChatGPT would be expected to produce a steadily changing gap across the entire period, whereas the data instead show a flat trend followed by a sudden shift concentrated exactly at the point of transition. This is consistent with a discrete change coinciding with ChatGPT's emergence, rather than a continuation of an existing trajectory. It should be noted that part of the widening in 2022/23–2023/24 coincides with the previously discussed change in oral examiner during those same years. This is precisely why the model excluding these two years (Table 4.) is an important complement to Figure 2 and its near-identical estimate (-0.201 vs. -0.211) indicates that the overall pattern is not solely an artifact of that control variable.

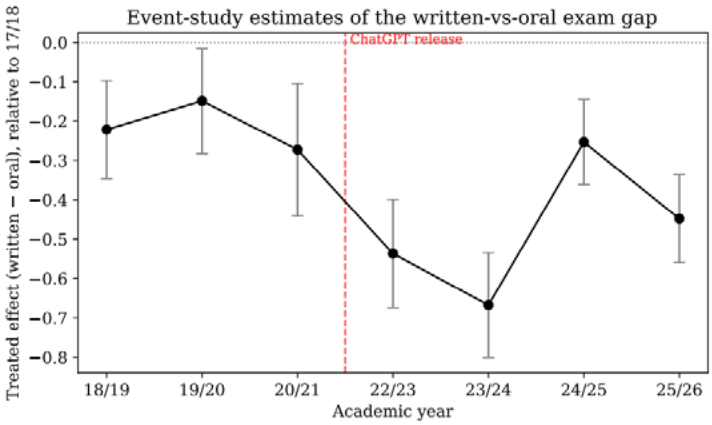

Figure 2. Event-study estimates of the written-vs-oral exam gap

5. Conclusion

This study examined whether the emergence of ChatGPT in November 2022 influenced student performance in Statistics 1 at the Zagreb School of Economics and Management, using written midterm exams and an oral exam as two assessment formats with different levels of exposure to AI-assisted preparation. The results show a change that is not uniform across formats, which directly addresses RQ_1. Written exam performance declined significantly relative to oral exam performance after ChatGPT's emergence (did = −0.211, $p < 0.001$), a result that holds after controlling for two identified control variables (a change in the oral exam's grade weight and a change in oral examiner during 2022/23–2023/24) and remains essentially unchanged when the control years are excluded entirely (did = −0.201). Event-study estimates reinforce this conclusion as the written-versus-oral gap was stable for three years before ChatGPT's release and widened sharply exactly at the point of transition, rather than following a trend already building beforehand. On this basis, H_0 is rejected for the written exams. The oral exam, in contrast, showed an increase in mean score after ChatGPT, but this is best explained by the examiner change rather than any AI-related effect, since the *oral_prof_change* control absorbs a large, significant share of the shift. H_1, which assumed a decrease in both formats, therefore holds only partially. The performance declined on written exams but not on the oral exam once its own control is addressed. This asymmetry is itself the more informative result, since it is exactly this differential pattern, rather than a uniform decline, that the design was built to detect. The pattern also addresses, RQ_2 as the decline is concentrated precisely in the format where real-time AI access during preparation was assumed to matter most, while the oral exam, requiring live, unaided articulation of reasoning, shows no equivalent decline once its effect is accounted for. This is what a cognitive offloading account predicts. A reliance on ChatGPT during preparation can produce a shallower level of understanding that becomes visible specifically once the tool is no longer available, as in a proctored exam [7], [14]. It is consistent with prior findings that ChatGPT-using students scored lower on midterm exams than their peers [11], and that AI assistance can support performance on well-structured tasks while undermining

deeper conceptual engagement [12]. The answer to RQ_2 is therefore a qualified yes. The pattern is consistent with cognitive offloading, but since individual students' actual ChatGPT usage was not measured, this remains a plausible, literature-consistent interpretation rather than a directly tested mechanism.

Several results call for caution rather than overstatement. This is an observational, course-level analysis, not an experiment, so an identical pattern could in principle arise from other unobserved changes coinciding with the same years, such as changes in how the written exams were written or graded, or in the profile of enrolling students. The fact that the effect is specific to the exam-type × period interaction, rather than a general post-period shift, makes such alternatives less likely but does not rule them out. The oral exam itself, intended as a stable benchmark, also proved less stable than assumed, a reminder that even assessment formats chosen for their resistance to AI-related effects require case-by-case verification rather than being assumed stable by design. The study's main limitations follow from its scope as it relies on aggregate exam outcomes rather than individual-level AI usage data, which limits conclusions about the mechanism at the student level. Furthermore, it covers a single course at a single institution, limiting generalizability. Additionally, the exclusion of 2021/22 leaves a gap in the pre-ChatGPT period, though the remaining three years still allow a reasonably balanced comparison. Future work could address the first limitation directly by collecting individual-level usage data, allowing a more direct test of the cognitive offloading mechanism proposed here. Replicating this design across other quantitative courses, institutions, or countries would help establish how general the observed pattern is. Finally, a longer post-ChatGPT observation window would help clarify whether the partial narrowing of the written-oral gap in 2024/25–2025/26 reflects genuine adaptation — for instance, through updated assessment design or explicit AI-literacy instruction — or is simply noise around a persistently elevated gap.

6. Acknowledgements

During the preparation of the paper, the author used Claude for help with generating the exploratory code for data clearing, formatting and analysis. The authors independently verified all outputs and results.

References

[1] M. S. Aslam and S. Nisar, Artificial Intelligence Applications Using ChatGPT in Education: Case Studies and Practices. in Advances in Educational Technologies and Instructional Design. IGI Global, 2023. doi: 10.4018/978-1-6684-9300-7.

[2] C. K. Lo, "What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature," Education Sciences, vol. 13, no. 4, p. 410, Apr. 2023, doi: 10.3390/educsci13040410.

[3] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, " in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada: ACM, Mar. 2021, pp. 610–623. doi: 10.1145/3442188.3445922.

[4] F. M. Megahed, Y.-J. Chen, J. A. Ferris, S. Knoth, and L. A. Jones-Farmer, "How generative AI models such as ChatGPT can be (mis)used in SPC practice, education, and research? An exploratory study," Quality Engineering, vol. 36, no. 2, pp. 287–315, Apr. 2024, doi: 10.1080/08982112.2023.2206479.

[5] J. Qadir, "Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education," in 2023 IEEE Global Engineering Education Conference (EDUCON), Kuwait, Kuwait: IEEE, May 2023, pp. 1–9. doi: 10.1109/EDUCON54358.2023.10125121.

[6] Z. Djebbari, "A New Era of Exam Cheating: The Use of ChatGPT and the Rise of Academic Dishonesty," vol. 7, no. 2, 2025.

[7] E. F. Risko and S. J. Gilbert, "Cognitive Offloading," Trends in Cognitive Sciences, vol. 20, no. 9, pp. 676–688, Sep. 2016, doi: 10.1016/j.tics.2016.07.002.

[8] Z. A. Pardos and S. Bhandari, "ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills," PLoS ONE, vol. 19, no. 5, p. e0304013, May 2024, doi: 10.1371/journal.pone.0304013.

[9] Z. Liu, H. Zuo, and Y. Lu, "The Impact of CHATGPT on Students' Academic Achievement: A Meta-Analysis," Computer Assisted Learning, vol. 41, no. 4, p. e70096, Aug. 2025, doi: 10.1111/jcal.70096.

[10] T. H. Kung et al., "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models," PLOS Digit Health, vol. 2, no. 2, p. e0000198, Feb. 2023, doi: 10.1371/journal.pdig.0000198.

[11] L. Bognár and A. Joós, "The Impact of Allowing ChatGPT on Student Scores and Test Completion Time in a Mid-Term Math Exam," DK, vol. 13, no. 2, pp. 11–24, Feb. 2025, doi: 10.63684/dk.2025.2.02.

[12] P. Rahman and S. Mehnaz, "Evaluating the Role of ChatGPT in Mathematics Education: Impact on Student Performance, Academic Integrity, and Perceptions in Higher Education," SSRN Journal, 2024, doi: 10.2139/ssrn.5054029.

[13] G. P. Georgiou, "ChatGPT produces more 'lazy' thinkers: Evidence of cognitive engagement decline".

[14] N. Kosmyna et al., "Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task," 2025, arXiv. doi: 10.48550/ARXIV.2506.08872.

[15] J. Schoultz, R. Säljö, and J. Wyndhamn, "Conceptual knowledge in talk and text: What does it take to understand a science question?," Instructional Science, vol. 29, no. 3, pp. 213–236, May 2001, doi: 10.1023/A:1017586614763.

[16] A. J. Stylianides, "Secondary students' proof constructions in mathematics: The role of written versus oral mode of argument representation," Review of Education, vol. 7, no. 1, pp. 156–182, Feb. 2019, doi: 10.1002/rev3.3157.

[17] P. Iannone, C. Czichowsky, and J. Ruf, "The impact of high stakes oral performance assessment on students' approaches to learning: a case study," Educ Stud Math, vol. 103, no. 3, pp. 313–337, Mar. 2020, doi: 10.1007/s10649-020-09937-4.

[18] D. Kang et al., "Providing an Oral Examination as an Authentic Assessment in a Large Section, Undergraduate Diversity Class," ij-sotl, vol. 13, no. 2, May 2019, doi: 10.20429/ijsotl.2019.130210.

[19] S. Nallaya, S. Gentili, S. Weeks, and K. Baldock, "The validity, reliability, academic integrity and integration of oral assessments in higher education: A systematic review".

[20] P. Iannone and N. Vondrová, "The Novelty Effect on Assessment Interventions: a Qualitative Replication Study of Oral Performance Assessment in Undergraduate Mathematics," Int J of Sci and Math Educ, vol. 22, no. 2, pp. 375–397, Feb. 2024, doi: 10.1007/s10763-023-10368-9.

[21] M. Huxham, F. Campbell, and J. Westwood, "Oral versus written assessments: a test of student performance and attitudes," Assessment & Evaluation in Higher Education, vol. 37, no. 1, pp. 125–136, Feb. 2012, doi: 10.1080/02602938.2010.515012.

[22] D. Corral and M. Yang, "An introduction to the difference-in-differences design in education policy research," Asia Pacific Educ. Rev., vol. 25, no. 3, pp. 663–672, Sep. 2024, doi: 10.1007/s12564-024-09959-0.

[23] R. Ospina and S. L. P. Ferrari, "Inflated beta distributions," Stat Papers, vol. 51, no. 1, pp. 111–126, Jan. 2010, doi: 10.1007/s00362-008-0125-4.

[24] R. Ospina and S. L. P. Ferrari, "A general class of zero-or-one inflated beta regression models," Computational Statistics & Data Analysis, vol. 56, no. 6, pp. 1609–1623, Jun. 2012, doi: 10.1016/j.csda.2011.10.005.

[25] L. Masserini, M. Bini, and M. Pratesi, "Effectiveness of non-selective evaluation test scores for predicting first-year performance in university career: a zero-inflated beta regression approach," Qual Quant, vol. 51, no. 2, pp. 693–708, Mar. 2017, doi: 10.1007/s11135-016-0433-z.

[26] C.-H. Yu and A. Ebeling, "Active Learning with Bayesian Reasoning: A POGIL-Based Pedagogy in Introductory Statistics," 2026, arXiv. doi: 10.48550/ARXIV.2606.08660.

[27] H. Bozdogan, "Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions," Psychometrika, vol. 52, no. 3, pp. 345–370, Sep. 1987, doi: 10.1007/BF02294361.

[28] N. Smirnov, "Table for Estimating the Goodness of Fit of Empirical Distributions," Ann. Math. Statist., vol. 19, no. 2, pp. 279–281, Jun. 1948, doi: 10.1214/aoms/1177730256.

[29] H. B. Mann and D. R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," Ann. Math. Statist., vol. 18, no. 1, pp. 50–60, Mar. 1947, doi: 10.1214/aoms/1177730491.

[30] H. Levene, "Robust Tests for Equality of Variances," In: Olkin, I., Ed., Contributions to Probability and Statistics, pp. 278–292, 1960.

[31] M. Lechner, "The Estimation of Causal Effects by Difference-in-Difference Methods," Foundations and Trends® in Econometrics, vol. 4, no. 3, pp. 165–224, Nov. 2011, doi: 10.1561/0800000014.

AI-Driven Learning Analytics Capabilities and Learning Data Quality for Intelligent Academic Decision-Making in Jordanian Higher Education: A Pilot Study

Maha Alkhaffaf

Applied Science Private University, Amman, Jordan, m_khaffaf@asu.edu.jo

Abstract

Artificial intelligence (AI) and learning analytics have become key drivers of data-driven decision-making in higher education. However, limited research has examined how AI-Driven Learning Analytics Capabilities and Learning Data Quality influence Intelligent Academic Decision Support and Academic Decision Quality. Accordingly, this study aims to develop and preliminarily validate an integrated research model based on Decision Support Systems (DSS) Theory. A quantitative pilot study was conducted to assess the reliability and validity of the proposed measurement instrument before undertaking a full-scale empirical investigation.

The measurement model was evaluated using SmartPLS 4 by examining indicator reliability, internal consistency reliability, convergent validity, and discriminant validity. The findings provide preliminary evidence supporting the suitability of the proposed measurement instrument for testing the hypothesized relationships among AI-Driven Learning Analytics Capabilities, Learning Data Quality, Intelligent Academic Decision Support, and Academic Decision Quality.

Keywords: *Artificial Intelligence 1, Learning Analytic 2, Learning Data Quality 3, Intelligent Academic Decision Support 4, Academic Decision Quality 5, Decision Support Systems 6.*

ISSN 2991-9371

1. Introduction

The rapid advancement of digital technologies has fundamentally transformed higher education, significantly influencing teaching, assessment, and the management of academic operations [1], [2]. Over the last few years, higher education institutions have generated a significant amount of data through Learning Management Systems (LMSs), online learning platforms, digital assessments, student information systems, and other educational technologies [3]. Digital transformation has encouraged universities to move from traditional administrative processes toward data-driven approaches in which institutional success increasingly depends on their ability to transform large amounts of learning data into knowledge that supports effective planning and decision-making [3], [4]. As a result, higher education institutions are not only expected to collect educational data but also to use it effectively to improve the quality of education, promote student success, design curricula, and enhance institutional performance [5].

Artificial Intelligence (AI) has emerged as one of the most transformative technologies driving this digital revolution by enabling higher education institutions to move beyond conventional data analysis toward intelligent and adaptive analytical approaches [6]. In contrast to traditional analytical methods, AI technologies such as machine learning, deep learning, natural language processing, and predictive analytics enable institutions to process complex datasets, discover hidden patterns, predict future outcomes, and provide recommendations that support evidence-based decision-making [7]. Consequently, AI is no longer viewed merely as an emerging technology but rather as a strategic enabler that enhances institutional efficiency, supports personalized learning experiences, identifies at-risk students, optimizes resource allocation, and supports high-quality academic decision-making [8], [9].

2. Background of AI-Driven Learning Analytics in Higher Education

2.1 Digital Transformation and Artificial Intelligence in Higher Education

Rapid digitalization of the higher education sector has revolutionized the educational environment and created opportunities to utilize modern digital tools to improve teaching and learning processes, conduct research, and manage higher education institutions. The widespread adoption of LMSs, cloud computing, educational platforms, the Internet of Things (IoT), and artificial intelligence has generated substantial amounts of educational data that can be used to optimize academic activities and institutional performance [10], [1]. Consequently, educational data have become a strategic institutional asset, enabling universities to make evidence-based decisions across both academic and administrative domains.

One of the most disruptive technological trends in the field of higher education is the emergence of artificial intelligence that provides the opportunity to implement intelligent, adaptive, and predictive methods of decision making in place of traditional information processing. Using AI technologies such as machine learning, deep learning, natural language processing, and

intelligent recommendation systems, institutions are able to discover hidden data patterns, predict academic outcomes, and make recommendations that improve decision making in academic and managerial contexts [10].

2.2 Evolution of Learning Analytics

Over the past decade, learning analytics (LA) has emerged as one of the most important fields of educational technology research. Traditionally, the learning analytics process focused primarily on collecting and presenting descriptive data about learners' engagement and performance. These conventional methods relied mainly on historical educational data to track student behavior and evaluate learning outcomes. Although they provided useful descriptive insights, they were limited in their ability to predict outcomes or prescribe appropriate actions [11].

Advances in artificial intelligence have led to a paradigm shift from descriptive reporting toward intelligent analysis in the domain of learning analytics. AI-based learning analytics combines machine learning, predictive analytics, deep learning, and intelligent recommendation systems to predict learner performance, identify at-risk students, personalize educational experiences, and provide appropriate instructional interventions [12].

This paradigm shift implies that learning analytics is no longer merely an instrument for reporting but also a tool for intelligent decision-making, enabling universities to make informed decisions based on continuously generated educational data.

Additionally, recent systematic reviews indicate that universities are increasingly using artificial intelligence in learning analytics not only to support student success but also to redesign learning environments and instructional practices and improve institutional decision-making processes. At the same time, the integration of artificial intelligence into learning analytics remains at an early stage in many higher education institutions, indicating considerable opportunities for future theoretical and practical development [5], [4].

3. Literature Review

3.1 AI-Driven Learning Analytics Capabilities

AI-powered learning analytics has emerged as a powerful approach that enables higher education institutions to derive valuable insights from large educational datasets. In contrast to conventional learning analytics, which primarily focuses on describing past learning processes, AI-powered learning analytics applies technologies such as machine learning, deep learning, and predictive analytics to identify learning patterns, predict academic outcomes, and generate intelligent recommendations for academic decision-making [7], [5].

From a Management Information Systems (MIS) perspective, AI-driven learning analytics capabilities refer to an institution's ability to effectively utilize AI technologies and analytical resources to transform educational data into valuable knowledge. These capabilities enable universities to make more effective evidence-based decisions, improve learning processes, identify at-risk students, and enhance student and institutional performance through intelligent analytical techniques [1], [4].

Based on the existing literature, this study conceptualizes AI-Driven Learning Analytics Capabilities as a multidimensional construct comprising four main dimensions: Predictive Analytics Capability, Real-Time Feedback Capability, Personalized Recommendation Capability, and Adaptive Analytical Capability. Through the integration of these four capabilities, higher education institutions can transform learning data into meaningful insights that support intelligent academic decision-making.

3.2 Learning Data Quality

Learning Data Quality (LDQ) refers to the degree to which learning data are accurate, complete, consistent, timely, and relevant for supporting analysis and academic decision-making [13]. In the context of AI-based learning analytics, high-quality learning data provide the foundation for reliable analytics and effective academic decision-making. From a Management Information Systems (MIS) perspective, data quality is widely recognized as an important factor influencing information quality, system effectiveness, and decision support performance [14]. Recent literature emphasizes the importance of learning data quality and effective data management for the successful application of AI technologies in education [10], [15].

Recent studies indicate that the effectiveness of AI-generated learning analytics depends heavily on the quality of learning data. Higher-quality learning data can contribute to more accurate predictions, better personalized learning recommendations, and more reliable real-time analytical insights. Conversely, inconsistencies, inaccuracies, and other data-quality problems may reduce the effectiveness of AI-based tools and lead to biased predictions and academic decisions [5], [16].

Accordingly, Learning Data Quality can be considered a strategic organizational asset that enables higher education institutions to derive greater value from AI-driven learning analytics. Providing high-quality learning data is therefore essential for generating reliable analytics, supporting intelligent academic decision-making, and ultimately improving the quality of academic decisions.

3.3 Synthesis of the Literature

The existing literature indicates that AI-driven learning analytics has become an important tool for supporting data-driven decision-making in higher education institutions. Previous research has emphasized the importance of AI capabilities and learning data quality in improving educational outcomes; however, these factors have generally been examined separately. Limited attention has been given to examining the relationships among AI-driven learning analytics capabilities, learning data quality, intelligent academic decision support, and academic decision quality within an integrated framework from a Management Information Systems (MIS) perspective.

4. Research Model and Hypotheses Development

Based on DSS Theory, the existing literature, and the conceptual framework proposed in this study (see Figure 1), five hypotheses are proposed to examine the relationships among AI-Driven Learning Analytics Capabilities, Learning Data Quality, Intelligent Academic Decision Support, and Academic Decision Quality in higher education institutions. The hypotheses are as follows:

- H1: AI-driven learning analytics capabilities positively influence intelligent academic decision support in higher education institutions.
- H2: Learning data quality positively influences intelligent academic decision support in higher education institutions.
- H3: Intelligent academic decision support positively influences academic decision quality in higher education institutions.
- H4: AI-driven learning analytics capabilities positively influence academic decision quality in higher education institutions.
- H5: Learning data quality positively influences academic decision quality in higher education institutions.

4.1 Research Model

Based on Decision Support Systems (DSS) Theory and the reviewed literature, the proposed research model illustrates the relationships among AI-Driven Learning Analytics Capabilities, Learning Data Quality, Intelligent Academic Decision Support, and Academic Decision Quality. The model provides the theoretical foundation for the proposed hypotheses.

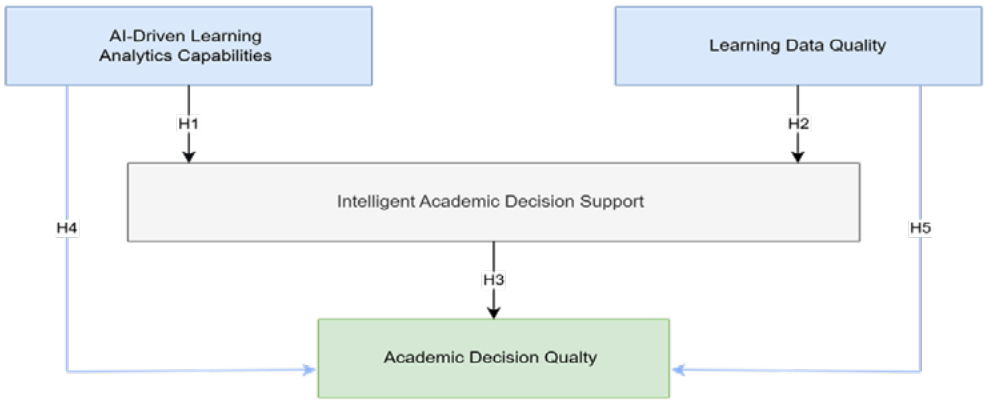

Figure 1. Proposed Conceptual framework.

- H1: AI-Driven Learning Analytics Capabilities → Intelligent Academic Decision Support
- H2: Learning Data Quality → Intelligent Academic Decision Support
- H3: Intelligent Academic Decision Support → Academic Decision Quality
- H4: AI-Driven Learning Analytics Capabilities → Academic Decision Quality
- H5: Learning Data Quality → Academic Decision Quality

4.2 Decision Support Systems (DSS) Theory

Decision Support Systems (DSS) Theory provides a theoretical framework for understanding how information systems support the analysis of complex data and enhance decision-making quality. A DSS is defined as a computerized system that facilitates the collection, processing, and analysis of data to support semi-structured and unstructured decision-making processes [19]. In line with developments in artificial intelligence, contemporary DSSs incorporate intelligent analytical tools, predictive models, and advanced data-processing capabilities to facilitate timely and accurate decision-making [5].

AI-driven learning analytics can be considered an advanced form of decision support system that transforms learning data into actionable insights for academic planning, curriculum development, assessment of learning performance, and institutional decision-making in higher education. Accordingly, DSS Theory is employed in this study to explain the relationships between AI-Driven Learning Analytics Capabilities and Learning Data Quality, and their influence on Intelligent Academic Decision Support and Academic Decision Quality in higher education institutions.

5. Research Methodology

5.1 Pilot study Design

The research design follows a quantitative pilot study approach to provide an initial assessment of the proposed research model and measurement instrument. A pilot study was conducted to evaluate the psychometric properties of the research measurement instrument before undertaking a large-scale empirical investigation. In this study, the reliability and validity of the scales used to measure AI-Driven Learning Analytics Capabilities, Learning Data Quality, Intelligent Academic Decision Support, and Academic Decision Quality were examined. Conducting a pilot study before the main study is recommended to improve the quality of the measurement instrument and minimize measurement error [10], [29].

5.2 Pilot Study

The pilot study was conducted to assess the clarity, consistency, and validity of the questionnaire items and to determine the initial psychometric characteristics of the measurement instrument. Pilot studies are important for identifying potential weaknesses in survey instruments before large-scale data collection and for improving the quality of both measurement and research. The results of the pilot study provide preliminary evidence regarding the suitability of the proposed measurement model for future hypothesis testing [29], [10].

5.3 Population and Sample

The target population of the study includes faculty members and academic administrative staff at Jordanian higher education institutions, as they participate directly in academic planning and decision-making and may be involved in the implementation of artificial intelligence and learning analytics initiatives. The pilot sample is expected to comprise approximately 30–50 respondents, which is considered appropriate for the preliminary assessment of the measurement instrument before the main study. Such sample sizes are commonly considered acceptable for pilot studies focused on instrument assessment rather than hypothesis testing [14], [13], [10].

5.4 Measurement Instrument

The research instrument was developed through a thorough review of the literature related to artificial intelligence, learning analytics, educational data quality, intelligent decision support systems, and academic decision quality. All measurement items were adapted from previously validated scales and tailored to the context of higher education. The questionnaire consists of five sections: demographic information, AI-Driven Learning Analytics Capabilities, Learning Data Quality, Intelligent Academic Decision Support, and Academic Decision Quality. Responses are measured using a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), which is commonly used in information systems and educational technology research [10], [6].

Table 1 presents the measurement constructs used in this pilot study, including their definitions, number of questionnaire items, measurement sources, alternative measurement approaches reviewed, and the rationale for selecting the final measurement approach. All measurement items were derived from previously validated instruments and adapted to the context of AI-driven learning analytics and intelligent academic decision-making in higher education.

Construct	Definition	Number of Items	Selected measurement Approach / Source	Alternative Measurement Approach Reviewed	Rational for selecting the final Approach
AI-Driven Learning Analytics Capabilities (AILAC)	The institution's capability to utilize AI technologies and learning analytics to collect, analyze, and interpret educational data for supporting academic decision-making.	XX	Adapted from AI and Learning Analytics Literature (2023–2025) Webber&Zheng (2024); Drugovaet al. (2024)	General learning analytics capability measures	General LA measures do not sufficiently capture the AI-enabled predictive, real time personalized and adaptive capabilities emphathasized in the study
Learning Data Quality (LDQ)	The extent to which learning data are accurate, complete, timely, consistent, and relevant for supporting intelligent academic decisions.	XX	Adapted from Information Quality and Learning Analytics literature Delone & McLean (2003) Webber&Zheng (2024)	General information and data quality measures	Broader data-quality measures may not sufficiently reflect the specific characteristics of learning data used for AI-driven academic decision-making.
Intelligent Academic Decision Support (IADS)	The capability of AI-enabled decision support systems to provide timely, reliable, and evidence-based recommendations for academic decision-making.	XX	Adapted from Decision Support Systems literature Power (2002), Arnott &Pervan (2014), Webber&Zheng (2024)	Traditional Decision Support System measures	Traditional DSS measures may not fully capture AI-enabled analytical recommendations and intelligent academic decision support.
Academic Decision Quality (ADQ)	The effectiveness, accuracy, and timeliness of academic decisions made using AI-supported information and analytical insights	XX	adapted from decision quality and higher education research Spetzler, Winter, &Meyer (2016) Schwartz &Sachs (2022)	General organizational decision-quality	General decision-quality measures are not sufficiently specific to academic decision-making in higher education.

Table 1. Measurement Construct and Sources
Source: adapted from the author from a previous validated study.

5.5 Data Analysis

The pilot data were analyzed using SmartPLS 4 to evaluate the measurement model. Specifically, indicator reliability, internal consistency reliability, convergent validity, and discriminant validity were assessed. Factor loadings were used to assess indicator reliability, while Cronbach's Alpha and Composite Reliability (CR) were used to examine internal consistency reliability. Average Variance Extracted (AVE) was used to assess convergent validity, while the Heterotrait–Monotrait ratio (HTMT) was used to assess discriminant validity. As the primary purpose of the pilot study is to validate the measurement instrument, hypothesis testing and structural model analysis will be conducted in the subsequent full-scale empirical study [10], [12].

6. Pilot Study Results

6.1 Respondents' Profile

The pilot test comprised faculty members and academic administrative staff from Jordanian higher education institutions. The primary purpose of the pilot test was to evaluate the research instrument rather than to obtain a representative sample of the target population. Therefore, the relatively small sample size was considered appropriate for assessing the preliminary psychometric properties of the measurement instrument. Respondents' demographic characteristics, including gender, academic rank, years of teaching experience, and university type, were examined using descriptive statistics [10], [29].

Table 2 presents the demographic characteristics of the pilot study participants. The frequencies and percentages shown in this table will be updated after completing the pilot data collection.

Variable	Category	Frequency	Percentage
Gender	Male	xx	xx
	female	xx	xx
Age	Less than 30	xx	xx
	30-39	xx	xx
	40-49	xx	xx
	50 and above	xx	xx

Academic Rank	Lecturer	xx	xx
	Assistant professor	xx	xx
	Associate professor	xx	xx
	Professor	xx	xx
Years of experience	Less than 5 years	xx	xx
	5-10 years	xx	xx
	11-15 years	xx	xx
	More than 15 years	xx	xx
University type	Public	xx	xx
	Private	xx	xx

Table 2. Demographic Characteristics of the Pilot Study Participants.

6.2 Measurement Model Evaluation

The measurement model was evaluated before examining the relationships among the study constructs. The evaluation considered indicator reliability, internal consistency reliability, convergent validity, and discriminant validity. This assessment helps determine whether the measurement instrument appropriately measures the intended constructs and provides reliable results for subsequent empirical research [10], [12]

6.3 Indicator Reliability and Internal Consistency Reliability

Indicator reliability was assessed based on the standardized factor loadings of the measurement items. According to Hair et al. [10], standardized loadings above 0.70 indicate adequate indicator reliability, whereas loadings above 0.60 may be acceptable in pilot studies. Internal consistency reliability was assessed using Cronbach's Alpha and Composite Reliability (CR). Values above 0.70 generally indicate acceptable reliability, whereas values above 0.90 may indicate potential item redundancy [10].

6.4 Convergent Validity

Convergent validity was assessed using the Average Variance Extracted (AVE). An AVE value of 0.50 or higher indicates that a construct explains at least 50% of the variance in its indicators and therefore provides evidence of adequate convergent validity [10], [9]. The AVE values for all constructs were examined against this threshold to assess the quality of the measurement model.

6.5 Discriminant Validity

Discriminant validity was evaluated using the Heterotrait–Monotrait ratio (HTMT), a widely recommended criterion for assessing discriminant validity in variance-based structural equation modeling. An HTMT value below 0.85, or below 0.90 under certain conditions, indicates adequate discriminant validity between constructs. The results provide preliminary evidence regarding the distinctiveness of the study constructs and the appropriateness of the proposed measurement model [12], [10].

6.6 Summary of the Pilot Study

In general, the pilot study was conducted in order to assess the quality of the measurement instrument before conducting the actual empirical study. Reliability and validity assessments are the first steps towards obtaining evidence on whether the questionnaire is appropriate for the assessment of AI-Driven Learning Analytics Capabilities, Learning Data Quality, Intelligent Academic Decision Support, and Academic Decision Quality. If the psychometric properties of the questionnaire are satisfactory, the instrument may be used for further empirical testing of the research hypotheses through Structural Equation Modelling (PLS-SEM) [10], [29].

7. Conclusion and Future Research

This study developed a conceptual research framework comprising AI-Driven Learning Analytics Capabilities, Learning Data Quality, Intelligent Academic Decision Support, and Academic Decision Quality in the context of higher education in Jordan. Grounded in Decision Support Systems (DSS) Theory, the study presents a conceptual model for understanding the role of artificial intelligence and learning data quality in supporting academic decision-making [19], [5].

The study was designed as a pilot study to assess the reliability and validity of the research instrument before conducting a full-scale empirical investigation. The pilot study provides preliminary evidence regarding the clarity, relevance, and psychometric properties of the measurement items [29], [10].

The study makes both theoretical and practical contributions. Theoretically, it extends the application of Decision Support Systems Theory by incorporating AI-driven learning analytics capabilities into the context of academic decision-making. Practically, the proposed framework may assist university administrators and academic decision-makers in using learning data and artificial intelligence to support more informed academic decisions.

Future research should conduct a full-scale empirical study using a larger sample and Structural Equation Modeling (PLS-SEM) to test the proposed hypotheses and examine the relationships among the study constructs. Future studies could also compare different types of higher education institutions and extend the investigation to other countries to assess the generalizability of the proposed model.

Limitations of the Study

This study has several limitations. First, the research is based on a pilot study and therefore does not provide empirical testing of the proposed hypotheses. The primary purpose of the pilot study was to assess the reliability and validity of the measurement instrument before conducting the full-scale study.

Second, the scope of the study is limited to the higher education context in Jordan, which may limit the generalizability of the findings to other educational contexts and countries. Future research should therefore use larger samples and Structural Equation Modeling (PLS-SEM) to empirically test the proposed hypotheses. Comparative studies involving different types of universities and higher education systems in other countries would also provide valuable insights into the applicability and generalizability of the proposed research model

References

[1] A. Ahmad, J. Schneider, D. Griffiths, D. Biedermann, D. Schiffner, W. Greller and H. Drachsler, “Connecting the dots: A literature review on learning analytics indicators from a learning design perspective,” Journal of Computer Assisted Learning, 2024. doi: 10.1111/jcal.12716.

[2] M. M. Alkhaffaf, S. A. Al Hunaiti, H. A. Hatamleh, D. Almajali, F. Shiyyab and A. Al Ryalat, “The impact of digital transformation and business intelligence on decision-making in the National Council for Family Affairs in Jordan: Theoretical framework,” in Generative Artificial Intelligence-Driven Innovation: Strategies and Ethical Frameworks for Responsible AI, Springer, pp. 465–476, 2026. doi: 10.1007/978-3-032-13149-2_34.

[3] M. Alkhaffaf, W. A. H. Hasan and Q. Hammoun, "E-Service Quality as a Predictor of Confirmed Expectations in Jordan," in Bridging Digital Innovation and Technology for Business Transformation, 2025.

[4] M. M. Alkhaffaf and G. A. Al-Otaibi, Decision Support Systems and Intelligent Systems. Amman: Al-Hamid Publishing and Distribution, 2012.

[5] D. Arnott and G. Pervan, "A critical analysis of decision support systems research revisited: The rise of design science," Journal of Information Technology, vol. 29, no. 4, pp. 269–293, 2014.

[6] G. O. Boateng, T. B. Neilands, E. A. Frongillo, H. R. Melgar-Quiñonez and S. L. Young, "Best practices for developing and validating scales for health, social, and behavioral research: A primer," Frontiers in Public Health, vol. 6, Art. no. 149, 2018.

[7] W. H. DeLone and E. R. McLean, "The DeLone and McLean model of information systems success: A ten-year update," Journal of Management Information Systems, vol. 19, no. 4, pp. 9–30, 2003.

[8] E. Drugova, I. Zhuravleva, U. Zakharova and A. Latipov, "Learning analytics driven improvements in learning design in higher education: A systematic literature review," Journal of Computer Assisted Learning, vol. 40, no. 2, pp. 510–524, 2024. doi: 10.1111/jcal.12894.

[9] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," Journal of Marketing Research, vol. 18, no. 1, pp. 39–50, 1981.

[10] J. F. Hair, G. T. M. Hult, C. M. Ringle and M. Sarstedt, A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM), 3rd ed. Sage, 2022.

[11] M. Hazaimeh and A. M. Al-Ansi, "Model of AI acceptance in higher education: Arguing teaching staff and students perspectives," International Journal of Information and Learning Technology, vol. 41, no. 4, pp. 371–393, 2024. doi: 10.1108/IJILT-01-2024-0005.

[12] J. Henseler, C. M. Ringle and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," Journal of the Academy of Marketing Science, vol. 43, no. 1, pp. 115–135, 2015.

[13] M. A. Hertzog, "Considerations in determining sample size for pilot studies," Research in Nursing & Health, vol. 31, no. 2, pp. 180–191, 2008.

[14] G. A. Johanson and G. P. Brooks, "Initial scale development: Sample size for pilot studies," Educational and Psychological Measurement, vol. 70, no. 3, pp. 394–400, 2010.

[15] D. Keržič, A. Aristovnik and D. Ravšelj, "A bibliometric analysis of learning analytics evolution in higher education: From traditional to artificial intelligence-based approaches," in Artificial Intelligence in Education Technologies: New Development and Innovative Practices, Lecture Notes on Data Engineering and Communications Technologies, vol. 279, Springer, 2026. doi: 10.1007/978-981-95-4423-3_35.

[16] J. Márquez, V. Henríquez, H. Chevreux and E. Scheihing, "Adoption of learning analytics in higher education institutions: A systematic literature review," British Journal of Educational Technology, vol. 55, no. 2, pp. 439–459, 2024.

[17] M. Mukred, U. A. Mokhtar, B. Hawash, H. AlSalman and M. Zohaib, "The adoption and use of learning analytics tools to improve decision making in higher learning institutions: An extension of technology acceptance model," Heliyon, vol. 10, no. 4, e26315, 2024. doi: 10.1016/j.heliyon.2024.e26315.

[18] I. C. Peláez-Sánchez, D. Velarde-Camaqui and L. D. Glasserman-Morales, "The impact of large language models on higher education: Exploring the connection between AI and Education 4.0," Frontiers in Education, vol. 9, Art. no. 1392091, 2024.

[19] D. J. Power, Decision Support Systems: Concepts and Resources for Managers. Westport, CT: Quorum Books, 2002.

[20] C. S. Spetzler, H. Winter and J. Meyer, Decision Quality: Value Creation from Better Business Decisions. Wiley, 2016.

[21] A. Stojanov and B. K. Daniel, "A decade of research into the application of big data and analytics in higher education: A systematic review of the literature," Education and Information Technologies, vol. 29, pp. 5807–5831, 2024. doi: 10.1007/s10639-023-12033-8.

[22] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," Journal of Management Information Systems, vol. 12, no. 4, pp. 5–33, 1996.

[23] K. L. Webber and H. Y. Zheng, "Artificial intelligence and advanced data analytics: Implications for higher education," New Directions for Higher Education, no. 207, pp. 5–13, 2024. doi: 10.1002/he.20508.

[24] S. Chadha, "Artificial intelligence in higher education: Transforming teaching, learning, and institutional management," Higher Education Quarterly, vol. 78, no. 1, pp. xx–xx, 2024.

[25] A. Ersozlu, et al., "Learning data quality and artificial intelligence in higher education: Challenges and opportunities," Education and Information Technologies, 2024.

[26] A. Farias-Gaytan, et al., "Digital transformation in higher education: Artificial intelligence and learning analytics perspectives," Education and Information Technologies, vol. 28, 2023.

[27] G. Katsamakas, et al., "Artificial intelligence and digital transformation in higher education," Computers and Education: Artificial Intelligence, vol. 5, 2024.

[28] S. Salas-Martinez, et al., "The evolution of learning analytics in higher education: From descriptive to predictive analytics," Education Sciences, vol. 14, 2024.

[29] H. Taherdoost, "Pilot study and questionnaire validation for social science research," MethodsX, vol. 10, 2023.

[30] H. L. Schwartz and J. D. Sachs, "Academic decision quality in higher education: Evidence-based decision making and institutional effectiveness," Higher Education Policy, 2022.

Implementation of the GAMBL-program – A Psychoeducational Intervention with a Gender Perspective for Gambling Behavior Prevention in Higher Education

Marina Ledesma-Casas[1], Loreto Camarero-Grados[2] and Yolanda López-del-Hoyo[3]

[1]Institute for Health Research Aragón (IIS Aragón), Zaragoza, Spain, marinaledesmacs@gmail.com
[2]Institute for Health Research Aragón (IIS Aragón), Zaragoza, Spain, loretocamarerog@gmail.com
[3]Institute for Health Research Aragón (IIS Aragón), Zaragoza, Spain, ylopez.iacs@gmail.com

Abstract

Gambling-related harm is a growing public health concern among young people and university students, highlighting the need for accessible and engaging prevention strategies. This paper presents the GAMBL Project, a set of digital and gamified approaches developed to prevent gambling-related harm and assess psychological variables associated with gambling behaviour. Three programs are described: GAMBL-OUT, an immersive psychoeducational intervention for adolescents and young adults; GAMBL-HER, a gender-sensitive primary prevention intervention addressing gambling-related risk factors among women; and GAMBL-HID, a gamified assessment program targeting psychological variables related to gambling. GAMBL-OUT was evaluated through a cluster randomized controlled trial assessing knowledge, gambling intention and attitudes at baseline, immediately after the intervention, and at a 1-month follow-up, together with qualitative implementation outcomes. Preliminary results from 410 college students showed that 8.8% presented moderate-to-high gambling risk. Gambling-risk perception decreased significantly following the intervention, and the program showed high acceptability, feasibility and implementability. A pilot study of GAMBL-HER involving nine expert women showed adequate adherence and positive evaluations, particularly regarding emotional regulation, self-care and identification of at-risk situations. Overall, these findings support the potential feasibility and acceptability of innovative, gamified and gender-sensitive approaches to gambling prevention in young people and women.

Keywords: *Mental Health, College Students, Gambling Disorder, Prevention Programs.*

ISSN 2991-9371

1. Introduction and objectives

The importance of mental health in university students is a recognized global public health concern that has received increasing attention in the last years [1]. The onset of psychiatric disorders is regarded as an origination point in adolescence and early adulthood [2]. Furthermore, university students are more prone to stress-exposure effects [3] and higher levels of anxiety, depression or distress than non-students peers [4].

Gambling disorder is a behavioural addiction in which a recurrent and persistent pattern of gambling occurs, creating a significant distress or impairment in different areas of life such as: emotional, social, financial or career. It is estimated a 1.4% global prevalence [5]. Gambling disorder is considered a public health problem [6], [7], [8] due to its association with psychiatric comorbidities, financial harm, criminal record or higher risk of suicide [9], [10]. In a southern Spain study, it was found that gambling disorder affected 4.7% of late-teenagers [11]. Furthermore, after COVID-19 there has been an increase of gambling activity in young adults and adolescents [12]. Gambling disorder is more prevalent in university students than older adults [13]

Digital intervention programs are effective to increase wellbeing and mental health in university students [14], [15]. Although, it seems more cost-effective implementing prevention programs [16]. Our research group has created prevention programs with an online-focus methodology with interactive and immersive approaches to study and prevent gambling.

For prevention programs we produced two: (I) GAMBL-OUT, aim at teenagers and young adults, with a universal purpose; and (II) GAMBL-HER, for a general female population, focuses on specific risks faced by women. Stigma, accessibility difficulties or gambling motives, development and course being different are some of the complexities women face in gambling disorder [17]. On the other hand, we created GAMBL-HID, a gamified evaluation. This work presents the clinical trial protocols and preliminary implementation outcomes for both preventive interventions.

2. Methods – GAMBL-Programs

2.1 GAMBL-OUT

GAMBL-OUT program responds to the crescent need for preventive programs aim to the youth but with an immersive methodology. The goal of the program is to approach a relatable experience of gambling, its development and harms associated in a story with a protagonist close to their age. In tackling that, the program is composed by two parts: (I) an immersive-story driven online escape-room; and (II) a gamified psychoeducational reinforcement session. Both sessions are done in-person, dividing the participants into different groups. The methodology involves both offline and online tools. For example, the escape-room was built to be done onto a computer, but the activity is performed on site.

To evaluate GAMBL-OUT, a cluster randomized controlled trial was conducted. The evaluation comprised questions about knowledge, intention to gamble and attitudes assessed at baseline, immediately post-intervention, and at a 1-month follow-up. Implementation outcomes (e.g., acceptability, feasibility, fidelity) were evaluated qualitatively.

In the escape-room, each group follows Lucas who initially appears to be successful and skilled at gambling. Nevertheless, that mask crumbles as the narrative unfollows. The participants notice rather quickly that gambling has already caused losses and other negative consequences in diverse areas of his life. To facilitate the understanding of how he ended up at that point, the participants submerge in a flashback across his journey, beginning the escape and puzzle-solving escape-room. The escape-room consists of five rooms per se. Each one catered towards a type of gambling and its associated risks; except for the last one that centres all the harms accumulated through time.

Some of the contents the escape-room shows, helps to identify and dismantle are: (I) slot-machines, with their stimuli associated, gambling-bias, the availability or low cost of entry; (II) sport-bets and tipsters, common tricks and the presence and urgency of betting in sports, with financial and social harms starting to become more accentuate; (III) online poker and solitude, exploring expectations and behaviours like lying or robbing; (IV) in-house games with common gambling-related cognitive bias; and lastly, (V) different types of harm take a central stage ending with a call to action to seek help. The second session reinforces the psycho-educational information present in the escape in a more direct way: focused on biases, warning signs and resources for help.

2.2 GAMBL-HER

In response to the lack of female-risks focused prevention programs for gambling disordered, we created GAMBL-HER. It is an online primary intervention catered around risk-factors to promote a better understanding of the course and harms gambling causes onto women and to provide tools to prevent the onset and course of gambling such as: mindfulness and emotional regulation, self-care; leisure and time management; and gambling-related psycho-educational information. Each one of them being worked through a chapter of the program.

The promotional and recruitment material were a series of situations involving women and their use of gambling in a comic-style format: as a coping mechanism, emotional regulation, to not feel alone, to avoid facing a domestic violence situation...

2.3 GAMBL-HID

A gamified evaluation of psychological variables related to gambling such as impulsivity, decision-making, risk-taking, competitiveness and cognitive bias done in an online environment. The program had the objective of validating tasks with traditional questionnaires. Some gamified instruments were already in existence: Game of Dice Task [18], Iowa Gambling Task [19], Monetary Choice Questionnaire [20], although the last one is being validated to Spanish and modified according the current currency value. The team created task for competitive behaviour (GAM-COMPET), gambling-related cognitive bias (CASS) and chase behaviour (LST).

3. Results

3.1 GAMBL-OUT

Preliminary results with a total college sample of 410 (58.8% woman, mean age of 19) had an 8.8% moderate to high risk gamblers as for PGSI with a medium to large need to gamble present in 4.8%. After the intervention the perception of gambling risk descended significantly ($p = 0.001$) as it did future bet intention ($p = 0.176$).

In regard implementation of the program, for acceptability there was a high rate of approval, enjoyment, appeal and liking. It was also rated high for viability variables as feasibility, implementability and future use intention.

Acceptability	Rated 4 or 5 out of 5 (%)
Approval	95.5
Appeal	95.5
Enjoyment	92.4
Liking	95.5
Viability	
Future use intention	81.9
Implementable	78.5
Feasible	86.1
Ease of use	84.6

Suitability	
Timely	84.6
Context-appropriate	81.5
Applicable	84.6
Enhances existing efforts	83.1

Table 1. GAMBL-OUT preliminary implementation results. The Likert scale was 1 (Totally disagree) to 5 (Totally agree). The sentence structure, using as an example "*Approval*", was: *"The program has my approval"*.

3.2 GAMBL-HER

The pilot study, which included nine women from diverse professional backgrounds, which included health, educative and social experts; showed adequate adherence and positive evaluation of its contents. More so those elements for emotional regulation, self-care and identifying at-risk situations.

The majority assessed positively the viability and utility of the program. They also appraised the use of cartoon-style advertisements depicting different gambling situations employing women of different age and background.

4. Conclusion and discussion

GAMBL-OUT program was perceived by young people as appealing, appropriate and feasible. There were signs of improvement in perceptions of the risks associated with gambling, attitudes toward gambling advertising and other key variables. This is in concordance with new prevention programs aim toward youth [21], [22].

GAMBL-HER program is a structured and transferable program that centers and caters towards a demographic often forgotten by traditional prevention programs [17]. The pilot study showed potential for a full trial where female-focus risks are taken in consideration to strengthen tools to diminish gambling as a viable tool for, for example, emotional regulation. This focus on adaptive every-day tools could help not only at-risk women but every one of them [17].

Both online prevention programs showed an initial good reception. The online methodology, and semi-online with on-site workshop for GAMBL-OUT program; were adequately and positively perceived with appraisal on the blend of technology and human touch. The GAMBL Project offers a scalable, innovative framework that combines gamification and gender-sensitive approaches to

advance behavioral addiction prevention in higher education. Our aim is to further recruitment for GAMBL-OUT program and launch a GAMBL-HER clinical trial shortly.

5. Acknowledgements

These projects have been funded through files SUBV24/00024, SUBV23/00020 from the Directorate General for Gambling Regulation, part of the Ministry of Social Rights, Consumer Affairs and the 2030 Agenda and EXP2022/008902 from the Spanish Ministry of Health (National Drug Plan, Research projects financed with European funds. This research was supported by Carlos III Health Institute (ISCIII) of the Spanish Ministry of Economy and Competitiveness with group funds the Research Network on Chronicity, Primary Care and Health Promotion (RICAPPS, RD24/0005/0004) that is part of the ResultsOriented Cooperative Research Networks in Health (RICORS; ISCIII), funded by the European Union "NextGeneration EU/PRTR" funds and Mental Health Research Group in Primary Care (B17_23R) which is part of the Department of Employment, Science and Universities at the Government of Aragón (Spain).

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors thank the Centre for Biomedical Research in Epidemiology and Public Health (CIBERESP; CB22/02/00052; ISCIII), the Research Network on Chronicity, Primary Care and Health Promotion, Institute of Health Research of Aragon and the University of Zaragoza for their support.

References

[1] P. Fusar-Poli, C. U. Correll, C. Arango, M. Berk, V. Patel, and J. P. A. Ioannidis, "Preventive psychiatry: a blueprint for improving the mental health of young people," *World Psychiatry*, vol. 20, no. 2, pp. 200–221, 2021, doi: 10.1002/wps.20869.

[2] M. Solmi *et al.*, "Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies," *Mol Psychiatry*, vol. 27, no. 1, pp. 281–295, Jan. 2022, doi: 10.1038/s41380-021-01161-7.

[3] G. Fond, A. Bourbon, P. Auquier, J.-A. Micoulaud-Franchi, C. Lançon, and L. Boyer, "Venus and Mars on the benches of the faculty: Influence of gender on mental health and behavior of medical students. Results from the BOURBON national study," *Journal of Affective Disorders*, vol. 239, pp. 146–151, Oct. 2018, doi: 10.1016/j.jad.2018.07.011.

[4] V. Kovess-Masfety, E. Leray, L. Denis, M. Husky, I. Pitrou, and F. Bodeau-Livinec, "Mental health of college students and their non-college-attending peers: results from a large French cross-sectional survey," *BMC Psychol*, vol. 4, no. 1, p. 20, Apr. 2016, doi: 10.1186/s40359-016-0124-5.

[5] L. T. Tran *et al.*, "The prevalence of gambling and problematic gambling: a systematic review and meta-analysis," *The Lancet Public Health*, vol. 9, no. 8, pp. e594–e613, Aug. 2024, doi: 10.1016/S2468-2667(24)00126-9.

[6] L. Blank, S. Baxter, H. B. Woods, and E. Goyder, "Should screening for risk of gambling-related harm be undertaken in health, care and support settings? A systematic review of the international evidence," *Addiction Science & Clinical Practice*, vol. 16, no. 1, p. 35, May 2021, doi: 10.1186/s13722-021-00243-9.

[7] H. Wardle, G. Reith, E. Langham, and R. D. Rogers, "Gambling and public health: we need policy action to prevent harm," *BMJ*, vol. 365, p. l1807, May 2019, doi: 10.1136/bmj.l1807.

[8] World Health Organization, *International Classification of Diseases*, 11th ed. 2018. [WWW Document]. Available: https://icd.who.int/

[9] V. Marionneau, M. Egerer, and S. Raisamo, "Frameworks of gambling harms: a comparative review and synthesis," *Addiction Research & Theory*, vol. 31, no. 1, pp. 69–76, Jan. 2023, doi: 10.1080/16066359.2022.2113071.

[10] M. N. Potenza *et al.*, "Gambling disorder," *Nat Rev Dis Primers*, vol. 5, no. 1, p. 51, Jul. 2019, doi: 10.1038/s41572-019-0099-7.

[11] M. Moñino-García *et al.*, "The Adolescent Problem Gambling Prevalence Associated with Leisure-Time Activities and Risky Behaviors in Southern Spain," *Int J Ment Health Addict*, pp. 1–15, Nov. 2022, doi: 10.1007/s11469-022-00950-7.

[12] A. Barceló-Soler *et al.*, "Post-COVID gambling habits of adolescents and young adults in Aragon, Spain: a cross-sectional study," *BMC Public Health*, vol. 25, no. 1, p. 2444, Jul. 2025, doi: 10.1186/s12889-025-23591-6.

[13] S. Kapukotuwa, A. González-Alvarez, S. W. Kraus, and M. Sharma, "Betting on the Future: A Systematic Review of Modern Gambling Interventions for College Students," *J Gambl Stud*, vol. 41, no. 4, pp. 1533–1561, Dec. 2025, doi: 10.1007/s10899-025-10421-y.

[14] X. Zhan *et al.*, "Effectiveness of Digital Health Interventions on Stress, Anxiety, and Depression in University Students: A Systematic Review and Meta-Analysis," *Journal of Visualized Experiments (JoVE)*, no. 227, p. e69949, Jan. 2026, doi: 10.3791/69949.

[15] M. Jabs, M. Thompson, J. Linardon, M. Messer, and T. D. Wade, "Assessing the Effectiveness, Adherence and Attrition of Smartphone-Based Mental Health Interventions in University Students: A Systematic Review and Meta-Analysis," *Early Intervention in Psychiatry*, vol. 20, no. 8, p. e70207, 2026, doi: 10.1111/eip.70207.

[16] L. K.-D. Le *et al.*, "Cost-effectiveness evidence of mental health prevention and promotion interventions: A systematic review of economic evaluations," *PLoS Med*, vol. 18, no. 5, p. e1003606, May 2021, doi: 10.1371/journal.pmed.1003606.

[17] A. Monreal-Bartolomé *et al.*, "Gambling in women: A systematic review of interventions and prevention approaches," *Journal of Behavioral Addictions*, vol. 15, no. 1, pp. 135–156, Mar. 2026, doi: 10.1556/2006.2025.00487.

[18] M. Brand, E. Kalbe, K. Labudda, E. Fujiwara, J. Kessler, and H. J. Markowitsch, "Decision-making impairments in patients with pathological gambling," *Psychiatry Research*, vol. 133, no. 1, pp. 91–99, Jan. 2005, doi: 10.1016/j.psychres.2004.10.003.

[19] A. Bechara, *IGT: Iowa Gambling Task: Professional Manual*. PAR, 2007. [Online]. Available: https://books.google.es/books?id=s-SlAQAACAAJ

[20] K. N. Kirby and N. N. Maraković, "Delay-discounting probabilistic rewards: Rates decrease as amounts increase," *Psychonomic Bulletin & Review*, vol. 3, no. 1, pp. 100–104, Mar. 1996, doi: 10.3758/BF03210748.

[21] D. Lloret Irles and V. M. Cabrera Perona, "Prevención del juego de apuestas en adolescentes: ensayo piloto de la eficacia de un programa escolar," *Revista de Psicología Clínica con Niños y Adolescentes: Journal of Clinical Psychology in Childhood and Adolescence*, vol. 6, no. 3, pp. 55–61, 2019.

[22] M. Chóliz, M. Marcos, and F. Bueno, "Ludens: A Gambling Addiction Prevention Program Based on the Principles of Ethical Gambling," *J Gambl Stud*, vol. 38, no. 3, pp. 993–1008, Aug. 2021, doi: 10.1007/s10899-021-10066-7.

THE USE OF AI TOOLS

The authors used ChatGPT (OpenAI) for language editing and translation of selected parts of the manuscript. All content was reviewed and verified by the authors.

Exploring the Role of Artificial Intelligence in Accounting, Audit, and Tax Education: Insights from Croatian Academia and Practice

Branka Remenarić[1], Ivan Čevizović[2] and Tonka Čevizović[3]

[1]Zagreb School of Economics and Management, Zagreb, Croatia, branka.remenaric@zsem.hr
[2]Zagreb School of Economics and Management, Zagreb, Croatia, ivan.cevizovic@zsem.hr
[3]Zagreb School of Economics and Management, Zagreb, Croatia, tcevizov@student.zsem.hr

Abstract

Artificial intelligence (AI) is changing the accounting, audit, and tax profession, but little is known about how this shift is being reflected in higher education outside the small number of countries that dominate the international research literature. This study explores the role of AI in accounting, audit, and tax education in Croatia using semi-structured interviews with nine university professors and two professional-chamber representatives. Findings show that AI integration into Croatian curriculum is highly uneven and mostly individually driven by educators rather than by institutional policy. That means Croatia is adopting changes in curriculum slower than suggested by international bibliometric trends. Despite this lagging adoption, Croatian respondents' judgements about priorities are in line with findings in international studies: critical evaluation of AI output, rather than technical AI literacy, was the most frequently cited competency gap. Audit was the discipline most frequently described as receptive to AI, tax the least, with accounting in between — a differentiation largely absent from the existing international literature, though not a unanimous view among respondents. The study also surfaces a difference in how critically professional-chamber and academic respondents assess graduate preparedness. Implications for curriculum design and academic–professional collaboration are discussed, with due caution given the study's small, hand-picked (not randomly selected) sample.

Keywords: *artificial intelligence, accounting education, audit education, tax education, curriculum development, Croatia.*

1. Introduction

Artificial intelligence (hereafter AI) is transforming the accounting, audit, and tax profession at a pace few other technologies have matched in recent decades. Large language models and machine-learning-based tools now assist with tasks ranging from journal-entry preparation and anomaly detection in audit samples to the interpretation of tax regulation, and professional bodies internationally have begun issuing formal guidance on their responsible use [1]. Higher education institutions that train future accountants, auditors, and tax advisers face a corresponding pressure to adapt curricula — a pressure that is not new in principle, having been anticipated as early as the mid-1990s [2], but that has intensified sharply since the emergence of widely accessible generative AI tools.

International research on AI in accounting education has grown rapidly in response, with three independent bibliometric analyses documenting accelerating publication output over the past decade [3]-[5]. Yet this growing body of research is geographically concentrated in a small number of countries — principally China, the United States, and Malaysia — and Croatia's contribution to it is minimal: a single article in two of these counts [3], [5], and no listing at all in the third [4]. Whether and how Croatian higher education institutions are responding to the same pressures documented internationally therefore remains a largely open empirical question, as does the extent to which the challenges, priorities, and competency gaps identified in the international literature apply in a national context so marginally represented in that literature.

This study shows that AI integration in accounting, audit, and tax education in Croatia lags behind the pace suggested by international bibliometric trends, even though Croatian educators' and practitioners' judgements that critical evaluation of AI output matters more than technical AI literacy closely align with international findings. The study also reveals two phenomena largely absent from the existing literature. One is a tentative audit–accounting–tax differentiation in AI receptiveness, and the other is a divergence between how academics and professional chambers assess graduate preparedness.

The remainder of the paper is structured as follows. Section 2 reviews the international literature on AI in accounting education and formulates the study's research questions. Section 3 describes the study's qualitative methodology. Section 4 presents findings from the interview data. Section 5 discusses these findings in relation to the international literature and considers their theoretical and practical implications. Finally, Section 6 concludes the paper.

2. Literature Review

Interest in AI within accounting education is not new: as early as the mid-1990s, at least one study was already calling for expert systems and AI topics to be integrated into the accounting curriculum [2]. Three decades later, the generative AI wave has renewed this debate with far greater urgency, and the literature has grown accordingly — both in volume and in the range of questions it asks.

There are two studies that map trends across the Scopus research database which both independently confirm that scholarly output on AI in accounting education is expanding rapidly, but unevenly. Al Ghatrifi, Al Amairi and Thottoli [3] examined 66 papers on technology in accounting education published between 1993 and 2022. They identified AI, ChatGPT, big data analytics, cloud computing, and blockchain coding as the most actively used tools in accounting education. The same study found that the three most contributing countries are China, the United States and Malaysia. Croatia appears with a single article. A report by Romero-Carazas et al. [4], focusing specifically on AI in accounting education (110 papers, 2003–2023), shows a 76.4% growth in publications between 2019 and 2023. China and the United States dominate this sample too, and Croatia does not appear among its 33 listed countries. A third, Croatia-linked bibliometric study by Rep Romić, Remlein and Sever Mališ [5] identifies AI-driven curriculum innovation as the most recent research cluster. The report also flags limited co-designing of the curriculum with practitioners as well as insufficient attention to public-private institution differences. Croatia appears once in their Web of Science dataset, suggesting Croatian scholarship on this topic exists but remains rare. This co-design gap is one of the issues RQ4 returns to.

A more focused literature review of 22 studies from the period 2021-2025 by Zhou and Zaini [1] found that adoption and acceptance of AI are one of the most researched themes (54.55% of studies). The field still remains focused on whether AI is being adopted and not on how to teach with it. Tandiono's [6] earlier review of 20 studies identified recurring calls for curricular reform without a clear consensus on how to implement it. This is a gap that persists roughly three decades after the original 1995 call for AI integration [2]. Research on adding business analytics and blockchain modules to existing courses [7] illustrates how curricula are being reshaped incrementally rather than holistically. How Croatia adopts these changes is the focus of RQ1.

The literature also shows a growing awareness of the risks associated with AI use in education. Four risk categories are identified across the literature: academic integrity, overreliance on AI outputs, data privacy, and instructor readiness gaps. These four categories are the basis for RQ2. Zhou and Zaini [1] found "challenges, risks, and ethical concerns" to be the most common research theme in their sample (59.09%). Llacay and Curós Vilà [8] tested six large language models on real exam questions and found strong AI performance on theory but weaker, less consistent results on journal entries and financial statements. No statistically significant differences were found in final performance between the two groups of students, one instructed by LLM and the other instructed by a professor. Gilreath, Hernandez and Ingalls [9], surveying 119 US accounting students, found that the most pressing barriers were limited real-world AI application, insufficient practice, and concerns about instructor-expertise. At the faculty level, Haidar and Alyousef [10] show a strong positive correlation between performance expectancy and adoption intention. Technical constructs (reliability, ease of use) predict ChatGPT adoption more strongly than behavioural constructs, with satisfaction the strongest predictor of use.

When it comes to the question whether curricula are actually preparing graduates for an AI-supported profession, research is scarce and the evidence is mixed. Brits, Moolman and Mothapo [11] interviewed lecturers on a South African postgraduate diploma in accountancy, identifying academic and non-academic warning signs that a student may not be future-ready. Van Zijl, Burnham and Ecim [12] compared five AI models against final-year South African Chartered Accountant candidates and found that AI matched or beat students on routine calculations

and journal entries but consistently did worse on tasks requiring judgement. The paper argues curricula should refocus accordingly. Research by Gilreath et al. [9] supports this indirectly. They found students ranked "creativity and innovation" as a separate skill from technical proficiency. That is consistent with students perceiving technical and judgement-oriented skills as two distinct skills. This study investigates that distinction under RQ3.

However, how well students are prepared for the future has been rarely examined for accounting, audit, and tax education separately. Most studies, including the bibliometric studies [3], [4] and systematic reviews [1], [6], treat accounting education as one single, general field. Abbas's [13] review of 91 articles is a partial exception, focusing on management accounting and documenting distinct pressures. The paper emphasises the need for revised curricula and closer academia-practice-employer cooperation. The comparison by Van Zijl et al. [12], similarly, is based on a general chartered-accountancy qualification rather than an audit or tax as a specific track. The lack of curriculum studies for a specific area leaves open whether audit, tax, and accounting education face the same or different AI-related pressures. This study addresses that question under RQ1b.

The question of student readiness for the future is part of a broader, frequently discussed gap between what curricula deliver and what the profession expects. Rahman [7] surveyed 183 faculty, professionals, and students in Bangladesh and found broad support for adding business analytics and blockchain content to curricula. However, there was only moderate-to-low satisfaction with how accreditation bodies turn that support into practical guidance for course design. This study explores that gap from the Croatian professional-chamber perspective under RQ3a. Van Zijl et al. [12] also concluded that curricula should place more emphasis on judgement-based skills such as critical reasoning and ethical judgment. The study by Gilreath et al. [9] shows that students already exposed to AI in coursework were significantly more likely to want it included more widely. 81.36% of the full sample recommended expanding AI training regardless of major.

In order to close that curriculum gap, closer academic–professional collaboration is required. However, how to do this in practice is not clear. Rahman [7] proposes a staged, hybrid integration pathway from the second undergraduate year into postgraduate study. Likewise, Al Ghatrifi et al. [3] think institutions should systematically incorporate emerging technologies rather than adopting them ad hoc. Marzuki et al. [14] propose systemic reform in teacher training and infrastructure. Abbas [13] sees AI-driven change as an opportunity, not a threat. As AI blurs the lines between different professional roles, it opens the door for closer collaboration across fields. This study investigates those mechanisms under RQ4.

Based on the literature review and identified gaps, this study sets out the following four main research questions each explored through one or two related sub-questions:

- RQ1: To what extent has AI been integrated into accounting, audit, and tax curricula at Croatian higher education institutions?

- RQ1a: How does AI integration in Croatia align with trends reported in international literature?

- RQ1b: What are the differences between audit, tax, and accounting education in terms of AI adoption and curriculum emphasis?
- RQ2: What challenges and risks do Croatian educators associate with AI use by accounting, audit, and tax students?
- RQ3: How prepared are current curricula, in the view of both educators and professional bodies, to equip graduates with the competencies needed for an AI-supported profession?
- RQ3a: Is there a gap between academic preparation and professional expectations regarding AI competencies, as perceived by Croatian tax and audit chamber representatives?
- RQ4: What forms of academic–professional collaboration are seen as necessary to better align accounting, audit, and tax education with AI-driven changes in the profession?

3. Methodology

3.1 Research design and interview questions

This paper is based on interviews and not surveys or statistical testing. The goal of the paper was to understand what professors and practitioners think about AI in accounting, audit, and tax education, in their own words. Since two groups of people were interviewed – professors and professional-chamber representatives – two interview guides were prepared: one for university professors, one for professional-chamber representatives. Each guide had six main questions, corresponding to the six thematic areas the study originally set out to explore and now presented as four main research questions with related sub-questions for conciseness (see Section 2), plus follow-up questions to get more detail (for example, on how quickly AI is being adopted, what are the biggest skill gaps, or what kind of academic–professional collaboration would help close that gap). The two guides asked about the same six topics but from each group's own point of view. Professors were asked about their opinion on AI integration in their classroom and chamber representatives about what changes they see in practice concerning the role of AI.

Overall, eleven people were interviewed: nine university professors teaching accounting, audit, or tax at eight different Croatian higher education institutions, and two representatives of professional chambers — the Croatian Chamber of Auditors (*HRK*) and the Croatian Chamber of Tax Advisers (*HKPS*). Nine professors across eight of Croatia's higher education institutions that offer accounting, audit, or tax programmes were included to maximise institutional breadth within the time and resource constraints of a conference-length qualitative study, rather than to achieve statistical representativeness; HRK and HKPS were included because they are the only two professional chambers with formal responsibility for education and licensing in the accounting, audit, and tax profession in Croatia, making them the natural and complete set of

relevant professional-body counterparts for this study. Interview participants were not selected randomly. Since the priority was to include professors from as many different higher institutions as possible, most of them were reached directly through the authors' existing contacts. Two respondents teach at the same institution, and both were kept in the study because their answers matched independently, which helped confirm that what they reported was reliable (see Table 1). Representatives of these two chambers are certified professionals involved in chamber management. Because participants were chosen this way rather than randomly, the findings show the range of views that exist rather than how common each view is across all Croatian educators or chamber representatives — a purposive-sample limitation discussed further in Section 3.2.

The interviews were carried out in July 2026. Because the institutions involved were spread across Croatia, and professionals and academics had limited time, all interviews were done in writing and not in person. Each participant received the relevant questions by email together with an explanation of the study's purpose. The explanation also stated that their answers would remain anonymous. They will be referred to, in the paper, as, for example, "Professor 1" or "Chamber representative." Participants could answer in as much detail as they wished and did not have to address every follow-up question separately. Writing rather than talking gave participants time to think and check institutional documents (such as a formal AI-use policy) before answering, but it also meant there was less room for immediate follow-up questions, unlike in a live interview, which is a constraint since the authors could not ask follow-up questions even when one seemed appropriate. For one respondent, a brief written follow-up was done after the initial interview to clarify an ambiguous answer. This is an isolated exception since no other respondents were asked for clarification.

Data were analysed using the Framework Method [15]. The Framework Method is well suited to this research because the starting point is a defined set of research questions rather than categories built from scratch. The analysis followed five standard stages. First, all written answers were read in full for familiarisation. Second, both interview guides were coded against the same set of codes, which were built around the six original thematic areas outlined above. However, they were also left open to codes not anticipated in advance. Third, these codes were organised into a working framework grouped under the four consolidated research questions, RQ1–RQ4, presented in Section 2. Fourth, each respondent's answers were indexed against this framework. In the final step, the indexed data were charted into a matrix with respondents as rows and codes as columns, allowing patterns of agreement, disagreement, and outlier positions to be read directly off the matrix rather than reconstructed from separate transcripts. This matrix is available from the authors on request and is the basis for the counts and comparisons reported below.

3.2 Ethics and limitations of this approach

All participants took part voluntarily and were promised anonymity in advance. This is a purposive, not a random or representative, sample. This is because only eleven people took part, and they were selected for institutional spread and accessibility rather than drawn at random from the population of Croatian accounting, audit, and tax academics and practitioners. As a

result, the findings should be read as a picture of the range of views present in Croatian academia and practice, not as proof of how common each view is more broadly, nor as generalisable to Croatian higher education as a whole.

4. Findings

This section presents findings from the eleven interviews described in Section 3: two with representatives of professional chambers (Chamber representative 1 and Chamber representative 2) and nine with university professors (Professor 1–Professor 9), anonymised per the confidentiality assurance given to participants. Given the small, hand-picked sample, findings indicate the range of positions within Croatian academia and practice rather than how common each position is.

Reported levels of AI integration ranged from extensive and formally regulated to entirely absent, as summarised in Table 1, which also records each professor's academic rank and their institution's public or private status. This heterogeneity means "the Croatian case" cannot be treated as a single, uniform state of adoption.

Respondent	Academic rank	Institution	AI integration	Formal policy
Professor 1	Lecturer	Public	Substantial	In development (institution level)
Professor 2	Lecturer	Private	Minimal	In development (institution level)
Professor 3	Assoc. Professor	Public	Substantial	Yes (institution level)
Professor 4	Professor	Private	None	None
Professor 5	Full Professor	Public	Moderate	Yes (institution level)
Professor 6	Asst. Professor	Public	None	Yes (university level)
Professor 7	Full Professor	Public	None	Yes (university level)
Professor 8	Full Professor (tenured)	Public	None	None
Professor 9	Assoc. Professor	Private	Minimal	In development (university level)
Chamber rep. 1	—	—	Active	-
Chamber rep. 2	—	—	Active	-

Table 1. Overview of respondents' institutional context (authors' work; anonymised).

4.1 Extent of AI integration (RQ1, RQ1a, RQ1b)

Analysis of the data shows that five of the nine professors reported some level of AI integration in their class materials and teaching techniques. Two professors reported substantial, self-initiated AI integration. One uses AI for content delivery and grading, and the other just for content delivery (preparation and improvement). Both professors reported a significant increase in AI usage over the past year. The increase is described by one professor as "brutal". Next, one professor described moderate AI integration just for content preparation but only to the extent of "helping me phrase my own thoughts more clearly". This professor noted that AI could potentially also help with class administration in the future, though this is not yet an active use. Two further professors reported minimal AI integration in the class. One uses AI to generate exam questions but plans to use AI in the near future specifically for calculations and modelling in accounting, and the other professor uses AI for content delivery. The remaining four professors reported no integration at all (Table 1).

Among the five professors who use AI in their class, three professors reported they apply AI more in audit classes (to prepare case-studies, for risk analysis, for anomaly detection) than in tax classes due to frequent changes in tax legislation. One professor stated that "I would not yet be confident using AI in courses strictly governed by legislation, such as tax". The remaining two reported no difference in their own approach across different areas (accounting, audit, tax). However, one professor reported he/she thinks AI is most useful specifically in tax teaching even though he/she is not using AI. Both chamber representatives confirmed that practice runs ahead of the classroom. Chamber representative 2 described AI already embedded in mandatory continuing professional development, while Chamber representative 1 described active general-purpose use of AI, with predictive audit models currently under development.

A plurality of respondents also pointed to a tentative pattern in which audit education is the most receptive to AI while tax education is the least receptive; this pattern should be read as indicative rather than established, since it was reported by only 4 of the 11 respondents. This was reported by three professors and one chamber representative (Chamber representative 2) (4 of 11 respondents; see Table 2). They explained it by pointing to audit's reliance on large-scale data analysis, compared to tax's dependence on frequent changes in legislation and on legal interpretation. One professor summarised this by saying AI is "most effective" in audit, "most applicable" in accounting "on routine tasks," and in tax it demands "the high level of expertise and judgement." This was not a majority view, though. One professor reported the opposite pattern, claiming tax to be the most receptive to AI. Three others said the emphasis should be the same across all three areas. Chamber representative 1 did not see any difference in the skill gap between tax and audit graduates. Due to very diverse responses to this question and little agreement among respondents, this question remains open and should not be treated as settled by this study.

Audit	Accounting	Tax
Most AI-receptive Used for data analysis, anomaly detection, risk assessment	Intermediate Used for routine, pattern-based tasks	Least AI-receptive Because of frequent legislative change, legal interpretation

Table 2. Area differentiation pattern (tentative; plurality view, 4 of 11 respondents)

4.2 Challenges and risks associated with student AI use (RQ2)

Four themes came up repeatedly. Those were: critical thinking and overreliance on AI output, academic integrity, unequal financial and technical abilities of students, and finally risk framework. First, without being asked directly, most professors raised concerns about loss of critical thinking and overreliance on AI output. One professor warned that unverified AI answers "can have major consequences for client". Another professor gave a more direct example, saying that AI "can generate answer with incorrect or outdated information but is very persuasive" so the answer is unchallenged. Chamber representative 1 raised a similar concern from the practice side, saying AI needs to be treated as a "support tool" and not "a pool of knowledge." Second, how professors handle academic integrity varied a lot by institution. One professor reported that he/she introduced a threshold of up to 20% allowed for AI-generated content and required mandatory declaration of AI usage. Another professor combined a formal policy with oral exams that account for 50% of the grade to validate independent student work. Some stated they rely on plagiarism software even though they are aware of software limitations. Some rely on signed declarations that AI was not used, or simply on recognising AI-generated work because it does not match a student's usual writing style. Third, most professors stated that the risk of unequal financial ability deepening inequality among students is currently small, since free versions are good enough. Only two professors disagreed. Fourth, Chamber representative 2 offered a much more developed risk framework than any of the professors. The framework sets out seven principles (autonomy and scepticism, AI literacy, transparency, data protection, integrity, awareness of bias, and regulatory awareness). None of the professors referred to anything similar.

4.3 Curriculum preparedness for an AI-supported profession (RQ3, RQ3a)

This question produced the clearest agreement in the whole dataset. The biggest competency gap reported by four of the nine professors and both chamber representatives independently is critical evaluation of AI output, not technical AI literacy. A fifth professor named it together with technical literacy rather than as a separate issue. One professor explained this with an analogy to financial-ratio interpretation: "one can calculate financial ratio without understanding what it means in particular context", just as AI can give an answer without the practitioner being able to judge whether it is reliable. Chamber representative 2 described the same gap as the difference between "obtaining a fast answer and verifying that it is correct." Two further professors pointed to a related but different gap. That is judgement about whether AI should be used at all in a particular situation.

Assessments of how ready the curricula are ranged from very critical to cautiously positive. One professor said curricula prepare students "very poorly," and pins this down to slow institutional approval procedures rather than a lack of awareness. Another professor rated it as preparing them "little," without giving a reason. On the other side, several professors said they believe that core theoretical knowledge stays sound and adaptable. One used the analogy that "the calculator did not kill mathematics so AI will not kill our profession."

When it comes to the gap between academic preparation and professional expectations, the study shows that four professors' self-assessment was moderate compared to two professors claiming low or critical preparation of students for future work. The harshest assessment came from Chamber representative 1, who described graduates as "almost completely unprepared for work" and academia as "very closed". Chamber representative 2 was more moderate but said something substantively similar: universities give "some" foundation, but the biggest gap is specifically in verification of AI output skills. Professors rated themselves less harshly on average. One professor's self-assessment was close to the chambers' criticism, and a second gave a similarly low rating, though without much explanation. This suggests there may be a gap in how the two sides perceive the gap itself.

4.4 Forms of academic–professional collaboration (RQ4)

Most respondents agreed that collaboration is needed, but there was disagreement on what form it should take. Chamber representative 1 gave a weak response and said guest lectures "would be a good start," and a formal advisory role "would be useful". This is a much softer tone than the one this respondent used when assessing the level of preparedness of graduates under RQ3a. Chamber representative 2, on the other hand, asked for "sustained collaboration, not just occasional guest lectures." This sustained collaboration would include jointly designing learning outcomes and testing AI answers against real Croatian tax cases. Chamber representative 2 explicitly supported giving the chamber a formal advisory role.

Professors were divided on whether chambers should have direct influence. Two saw this as inevitable and desirable. Three supported an advisory or partnership role, but not decisive authority, warning that professional bodies risk putting too much weight on current practice at the expense of the legal grounding behind it. One professor gave only weak, hedged support without taking a clear position. One said any bigger role should depend on chambers first becoming more rigorous, pointing to ACCA as an example. Two were against giving chambers direct influence at all.

All nine professors and both chamber representatives agreed AI should be built into existing accounting/audit/tax courses. Three professors went further, also wanting a stand-alone or elective AI course in addition to that, not instead of it.

5. Discussion

This section interprets the Section 4 findings against the Section 2 literature before explaining implications and limitations of this study.

5.1 Comparison of research results against international literature

RQ	International literature	Croatian findings	Relationship
RQ1; RQ1a	Adoption accelerating globally [3],[4]	Individually driven; 4/9 professors report no integration	Croatia lags
RQ1b	Area differentiation largely absent [13]	Audit > accounting > tax pattern (plurality, 4/11)	Fills gap (tentative)
RQ2	4 risk categories: integrity, overreliance, privacy, readiness [1]	Same themes; chamber risk framework more developed than curricula	Strong match
RQ3	AI weak on judgement-based tasks [12]	Critical evaluation = top gap (4/9 professors + both chambers)	Triangulated
RQ3a	Gap between academic preparation and professional expectations [7]	Gap; one of two chamber reps far harsher than professors' self-view	Extends finding (tentative)
RQ4	Collaboration endorsed in principle, mechanism unclear [7],[13],[14]	Same pattern; local caution on preserving legislative grounding	Consistent + local nuance

Table 3. Comparison of results against international literature

RQ1; RQ1a AI integration: international literature reports that research on AI integration in accounting education is growing fast, with publication counts roughly doubling between 2019 and 2023 [3], [4]. This study shows four of nine professors reported no AI integration at all in their institutions, meaning Croatia is not following the same pace. However, professors who reported AI integration stated that it was individually driven and not institutionally. This finding matches the incremental (bit by bit) adoption reported internationally [7]. Also, this suggests Croatia is behind the international curve in both practice and published output. However, this conclusion is not conclusive but suggestive since the sample is small.

RQ1b Area differentiation: The literature review identified area differentiation as a genuine gap in the literature, with Abbas [13] a partial exception. This study's results are tentative. The most common pattern identified is that audit is the most receptive, tax the least, and accounting in between. This was reported by 4 of 11 respondents, with no close international comparison available. Since only a plurality, not a majority, of respondents shared this view, it should be read as a candidate hypothesis rather than a confirmed pattern.

RQ2 Risk categories: four risk categories are identified internationally: integrity, AI overreliance, data privacy, instructor readiness [1]. These match the concerns raised by Croatian professors. Only one difference stands out. Chamber representative 2's formal charter of risk framework is far more developed than anything mentioned by professors, suggesting risk guidance already exists at the professional level but has not yet reached what is taught in courses. This is also a local example of the RQ3a gap. The dominant Croatian view is that free AI tools are enough, and that prompting skill matters more than whether a student pays for AI or not. This finding matches Gilreath et al.'s [9] finding of no significant differences by income in their US sample.

RQ3 Curriculum preparedness: this is where the Croatian findings match the international literature most clearly. Van Zijl et al. [12] found AI matched students on routine tasks but did worse on judgement. Independently, four of nine professors, plus a fifth who named it together with technical literacy, and both chambers agreed on the same point: critical evaluation, not technical skill, is the central gap. Two sources, two methods, one conclusion — arguably the paper's strongest finding, with a clear implication for curriculum design.

RQ3a Academia–profession gap: Rahman [7] reports that a gap exists between academic preparation and professional expectations and finds the academia–practice gap mainly in how accreditation bodies translate expectations into guidance. This study does not just confirm that the gap exists, but adds a new perspective that was not measured before. That is a difference in how sharply the two sides (chamber vs. academia) perceive that same gap. The study reveals a possible "perception gap about the gap" (one chamber notably harsh, the other more moderate, professors milder toward themselves). However, whether a real, systematic perception gap exists between chambers and academia cannot be settled with only two chamber respondents, and is better treated as a hypothesis for future research.

RQ4 Collaboration: this study shows results consistent with international literature. International literature shows long-standing calls for collaboration in principle but is short on mechanism [7], [13], [14]. This study shows broad support alongside sharp disagreement on form. The local point is that professors in this study repeatedly insisted that chambers risk favouring current practice over the legal grounding that tax and accounting work requires, consistent with tax being seen as least AI-receptive under RQ1b. When it comes to incorporating AI in the curriculum, this study shows the near-universal preference for integrating it into existing accounting/audit/tax courses. This finding matches Marzuki et al.'s [14] Technology Acceptance Model (TAM)-based argument. So, this is one area where local views and international theory already agree.

5.2 Implications and limitations

Overall, these findings suggest Croatian accounting, audit, and tax education is adopting AI later than the international trend would suggest. However, local judgements largely match international evidence, especially on critical evaluation as the key gap. This looks more like an implementation gap than a knowledge gap about priorities: slow approval procedures, uneven policy maturity, and no clear route from professional risk frameworks (such as the CFE (Confédération Fiscale Européenne) Charter of Tax Advisers' Rights and Obligations in an AI-

Influenced Tax-Advisory Environment) into actual course content. Confirming this would need data this study cannot provide, such as direct evidence on why approval procedures are slow.

Because critical evaluation of AI output, rather than technical AI literacy, is the most robust and best-triangulated finding of this study, it carries the clearest practical implication: Croatian higher education institutions should prioritise teaching students to verify, question, and contextualise AI-generated output over teaching them how to operate specific tools, since the tools themselves are already widely accessible. For universities, this could take the form of case-based exercises in which students are asked to identify errors or unsupported claims in AI-generated financial analysis, audit working papers, or tax advice, rather than simply producing such output faster. For professional chambers, the more developed risk frameworks already in use in continuing professional development, such as the seven-principle framework described by Chamber representative 2, represent a ready-made resource that could be adapted into curriculum content rather than left as a separate, practice-side layer of guidance disconnected from what is taught at university. Embedding chamber-developed risk principles into coursework would also directly address the academia–practice gap raised under RQ3a, since it gives academic staff a concrete, professionally validated starting point rather than requiring each institution to build its own guidance from scratch.

This study is based on a small, deliberately chosen (not random) group of eleven people, and all interviews were done in writing rather than live. Both facts limit how far the findings can be applied beyond this group, and how much follow-up detail could be collected. Future research could usefully test the audit>accounting>tax pattern with a larger group, ask graduates directly about the gap raised under RQ3a, and pilot turning chamber risk frameworks into actual curriculum modules to see whether this narrows the critical-evaluation gap found here.

6. Conclusion

This study explores the role of AI in accounting, audit, and tax education in Croatia. Croatia is only marginally represented in the international bibliometric record on this topic (two articles overall), despite the rapid global growth of that literature. Based on interviews with nine professors from eight institutions and two professional-chamber representatives, the study finds that AI integration into Croatian curricula is highly uneven and mostly driven by individual educators rather than institutional strategy, putting Croatia behind the adoption pace suggested by international research trends.

At the same time, where Croatian educators and practitioners do form judgements about priorities, these judgements match the international evidence closely. Critical evaluation of AI output, rather than technical AI skill, was independently identified by most respondents as the most pressing gap. This finding matches recent international studies that directly tested AI systems against accounting students. This suggests the main challenge facing Croatian accounting, audit, and tax education is less about working out what AI-supported curricula should prioritise, and more about speeding up and coordinating changes to act on priorities that are already widely agreed, both locally and internationally.

The study also documents two tentative patterns with no close counterpart in the international literature reviewed, neither strongly supported enough by this small sample to make an established finding. The first is a plurality view, reported by 4 of 11 respondents, in which audit was seen as the area most receptive to AI and tax as the least. The second is a difference, driven mostly by Chamber representative 1's much harsher language, in how critically professional-chamber and academic respondents assessed graduate preparedness for an AI-supported profession. Both patterns need further, larger-scale investigation before they can be treated as settled. In the meantime, the study's stronger findings offer a practical starting point for Croatian higher education institutions and professional chambers looking to align curriculum design with the demands of an AI-supported accounting, audit, and tax profession: prioritise developing critical-evaluation skills over technical AI training alone, adapt existing professional risk frameworks into course content rather than building parallel guidance from scratch, and pursue sustained rather than occasional academic–professional collaboration, while keeping the legal grounding that the profession, and tax practice in particular, still requires.

7. Disclosure of AI use

Interview responses, originally collected in Croatian, were translated into English with the assistance of AI tools. Also, during preparation of this paper the authors used generative AI tools to enhance the clarity, language, and readability of the manuscript. The authors reviewed and edited the output and maintain full responsibility for the content of this paper.

References

[1] J. Zhou and M. H. B. Zaini, "Perceptions of Artificial Intelligence in Accounting Education in Higher Education: A Systematic Literature Review", International Journal of Research and Innovation in Social Science, Volume 10, No. 3, pg. 3184-3198, 2026. https://doi.org/10.47772/IJRISS.2026.100300236

[2] A. A. Baldwin-Morgan, "Integrating artificial intelligence into the accounting curriculum", Accounting Education, Volume 4, No. 3, pg. 217-229, 1995. https://doi.org/10.1080/09639289500000026

[3] M. O. M. Al Ghatrifi, J. S. S. Al Amairi and M. M. Thottoli, "Surfing the technology wave: An international perspective on enhancing teaching and learning in accounting", Computers and Education: Artificial Intelligence, Volume 4, 100144, 2023. https://doi.org/10.1016/j.caeai.2023.100144

[4] R. Romero-Carazas, S. I. Dávila-Fernández, J. B. Vallejos-Tafur, F. A. Ochoa-Tataje, C. M. Samaniego-Montoya, J. A. Torres-Sánchez, M. S. Porras-Roque and A. P. Espiritu-Martinez, "Artificial Intelligence in Accounting Education and its Trends in Scopus: A Bibliometric Analysis", Migration Letters, Volume 20, No. 7, pg. 343-357, 2023. https://doi.org/10.59670/ml.v20i7.4311

[5] A. Rep Romić, M. Remlein and S. Sever Mališ, "Information technology in accounting education: A bibliometric-systematic literature review (2006–2025)", Zeszyty Teoretyczne Rachunkowości, Volume 49, No. 4, pg. 169-189, 2025. https://doi.org/10.5604/01.3001.0055.4528

[6] R. Tandiono, "The Impact of Artificial Intelligence on Accounting Education: A Review of Literature", E3S Web of Conferences, Volume 426, 02016, 2023. https://doi.org/10.1051/e3sconf/202342602016

[7] M. S. Rahman, "Incorporating business analytics and blockchain technology courses into the accounting curriculum: insights from faculty, professionals and students", Fintech and Digital Accounting Review, 2026. https://doi.org/10.1108/FDAR-08-2025-0022

[8] B. Llacay and M. P. Curós Vilà, "Can LLMs help students learn accounting? A performance analysis across models and prompts", Journal of Accounting Education, Volume 74, 101037, 2026. https://doi.org/10.1016/j.jaccedu.2026.101037

[9] R. Gilreath, D. Hernandez and V. Ingalls, "Student Perceptions of AI Integration in Accounting Education: Exploring the Value, Challenges, and Career Readiness", ACBSP Journal, June 1, 2025. https://doi.org/10.64010/YKVG7859

[10] F. T. Haidar and M. A. Alyousef, "Investigating behavioral and technical factors influencing faculty adoption of ChatGPT in accounting education", Discover Education, Volume 5, 318, 2026. https://doi.org/10.1007/s44217-026-01309-z

[11] R. J. Brits, A. M. Moolman and M. J. Mothapo, "Academic and non-academic indicators for preparing future-ready postgraduate diploma in accountancy students: Lecturers' insights", Journal of Accounting Education, Volume 74, 101032, 2026. https://doi.org/10.1016/j.jaccedu.2026.101032

[12] W. van Zijl, K. Burnham and D. Ecim, "How close is AI to replacing accounting consultants? Insights from a comparative study of multiple AI models and exit-level accounting students", Meditari Accountancy Research, 2026. https://doi.org/10.1108/MEDAR-07-2025-3160

[13] K. Abbas, "Management accounting and artificial intelligence: A comprehensive literature review and recommendations for future research", The British Accounting Review, Volume 58, 101551, 2026. https://doi.org/10.1016/j.bar.2025.101551

[14] M. Marzuki, S. S. Azizan, N. Omar, R. Safeei and N. F. Ilias, "From Traditional to Tech-Driven: The Role of AI in Shaping Student Engagement and Performance in Accounting Education", International Journal of Research and Innovation in Social Science, Volume 9, No. 10, pg. 9473-9480, 2025. https://doi.org/10.47772/IJRISS.2025.910000775

[15] N. K. Gale, G. Heath, E. Cameron, S. Rashid and S. Redwood, "Using the framework method for the analysis of qualitative data in multi-disciplinary health research", BMC Medical Research Methodology, Volume 13, 117, 2013. https://doi.org/10.1186/1471-2288-13-117

ONLINE PAPER SESSION 1

AI-Enhanced Learning, Teaching and Professional Practice

SESSION CHAIR:

Goran Oblaković

Teaching Digital Transformation Through AI-Simulated Organisations: A Design Blueprint for Interactive Teaching Cases

Sven Dittes, Isabell Steidel, Ralf Dillerup and Brian Boyd

Heilbronn University of Applied Sciences, Germany

Abstract

Digital transformation is commonly taught through conventional case studies in which students apply frameworks to analyse strategy, technology, processes, stakeholders, and change. While useful, this approach does not capture the central challenge of transformation practice: diagnosing ill-structured organisational problems shaped by distributed knowledge, conflicting perspectives, and competing interests. Generative artificial intelligence intensifies this limitation by rapidly producing plausible analyses of conventional cases. In response, we followed a design and development research approach to derive a blueprint for interactive teaching cases, in which client organisations are simulated and grounded in educational concepts such as epistemic friction, organisational sensemaking, and constructive alignment. The blueprint was instantiated as fictional client organisations, each realised as a customised ChatGPT application and validated against explicit readiness criteria before being implemented in a master-level digital transformation course. Acting as external consultants, student teams investigated their assigned organisation by interviewing stakeholder personas and eliciting partial, conflicting, and interest-laden accounts. A pilot evaluation with one cohort shows high acceptance, active engagement, and a positive perceived learning experience, alongside a clear need for greater interaction depth; it is important to note that the evidence speaks to feasibility and acceptability rather than to learning gains. Theoretically, the paper conceptualises generative artificial intelligence as a medium for representing the object of inquiry rather than as a tutor or solution provider. Practically, it offers educators a documented and potentially transferable blueprint – design principles, an authoring schema, and readiness criteria – for building context-rich cases and aligned assessment.

Keywords: *digital transformation, generative artificial intelligence, interactive teaching cases, organisational simulation, constructive alignment, higher education.*

ISSN 2991-9371

1. Introduction

Consider the following teaching situation: a student team is asked to advise a mid-sized manufacturer whose digitalisation programme has stalled. The head of production explains that a new planning system was imposed without regard for how shop-floor scheduling actually works. The chief technology officer argues that production has resisted every modernisation attempt since a disruptive enterprise resource planning implementation several years earlier. Neither account is untrue, neither is complete, and no document reconciles them. Before applying any framework, the team must first determine what the problem actually is.

Such situations are routine in digital transformation practice. Organisational problems rarely have clearly identified causes, complete evidence, or an agreed interpretation. Instead, transformation teams must confront fragmented information, competing narratives, political interests, and differing stakeholder perspectives. It remains unclear whether higher education provides students meaningful opportunities to navigate such situations.

Digital transformation is commonly taught through frameworks and tools for analysing strategy, processes, technologies, capabilities, stakeholders, and organisational change, often through their application to conventional case studies [1], [2]. Written cases provide a bounded and preselected body of evidence: even where several defensible solutions are possible, the organisational material has already been gathered, filtered, and arranged by the case author [3]. Live cases expose students to authentic ambiguity but require external partners and coordination [4]. In practice, knowledge spans functions and hierarchical levels, while transformation affects responsibilities, resources, and power relations [1], [5], [6]. Students must therefore assess different accounts, select appropriate frameworks, and propose feasible interventions. This requires making judgements based on incomplete and context-dependent evidence [7]. This raises a broader educational question: are we preparing students merely to analyse problems once they have been defined for them, or to investigate, interpret, and act within the conditions under which such problems actually emerge?

Generative artificial intelligence (GenAI) is automating routine transformation work, raising the importance of problem framing, critical evaluation, and accountable judgement under incomplete information and divergent interests [8]. Under constructive alignment, such shifts in required capabilities must be reflected in learning activities and assessment [9]. Yet because language models can produce plausible transformation deliverables, a polished artefact looks much the same whether a team spent weeks eliciting and reconciling organisational evidence or generated it from a static case description in seconds [10]+[12]. The challenge is therefore broader than academic integrity: GenAI changes the intended outcomes, can bypass parts of the learning activity, and weakens the inference from a written product to students' actual judgement.

Three observations therefore converge. First, digital transformation confronts students with ill-structured, distributed, and contested problems that conventional cases only partly capture, because their material is preselected and the resulting artefact can be produced without the inquiry it was meant to require. Second, GenAI intensifies this gap: it solves conventional cases in seconds, so under constructive alignment the shift in required capabilities must be carried

through into learning activities and assessment rather than patched with isolated remedies. Third, GenAI also opens pedagogical possibilities – particularly interactive, inquiry-oriented environments in which students elicit, test, compare, and interpret information – that have so far been described more often than designed, implemented, and evaluated [13], [14]. Together, these considerations motivate the following overall research question:

How can AI-simulated teaching cases be designed and embedded within a higher education course to address changing learning requirements while simultaneously leveraging the pedagogical potential of generative artificial intelligence?

Two subordinate questions follow from it:

- *RQ1 - which design principles govern such cases, and how are they instantiated and validated;*
- *RQ2 - how can the design be enacted in a first live classroom implementation, and what preliminary evidence of feasibility and acceptability emerges from a pilot evaluation.*

To address these questions, the paper reports the design, implementation, and preliminary evaluation of a master-level digital transformation course in which student teams acted as external consultants to fictional client organisations realised as customised GenAI applications, investigating distributed and conflicting stakeholder accounts assessed through oral defences, written essays, and individual learning diaries.

2. Related approaches to case-based and simulated learning

The design sits adjacent to three established traditions and is positioned in relation to them rather than as an isolated novelty.

Business simulations and management games place learners in decision-making positions under uncertainty, but learning is typically driven by an underlying outcome model that converts decisions into feedback [15], [16]. The present design has no outcome model: it simulates the epistemic situation of an outsider trying to establish what is happening inside an organisation.

Role play and role-based e-learning likewise embed learners in social situations [17], [18], but counterpart roles are usually played by peers or instructors. Here, model-represented stakeholders remain continuously available while students assume the role of external consultants.

The closest analogue is the standardised-patient tradition, where learners elicit a case through questioning rather than receive it as a document [19]. Recent work also uses language models as simulated clients [20]. Our design extends this logic to multiple organisational roles whose accounts must be triangulated and deliberately avoids a canonical diagnosis, shifting assessment from diagnostic accuracy to defensible reasoning.

The contribution is therefore not simulated interlocutors per se, but an elicitation-based, multi-role and interpretively open organisational case format together with an assessment design suited to that openness.

3. Methodology and study design

The study follows a design and development research logic: the object of study is an educational artefact, and the primary contribution is the artefact and its underlying design principles [21]. It is not an intervention study and does not test effectiveness against an alternative; its design-oriented warrant is that the principles are theoretically grounded, can be instantiated as working artefacts, behave as intended, and can be implemented in practice.

Four terms are used consistently. An AI-simulated organisation is the pedagogical construct: a fictional client organisation whose stakeholder knowledge is distributed across model-represented roles. An interactive teaching case is the resulting format in which learners elicit rather than receive the case material. A customised ChatGPT application is the technical instantiation used here, one possible carrier of the blueprint. Simulation is shorthand for the AI-simulated organisation as encountered by students.

The study comprises three components with an explicit hierarchy. *Design (RQ1):* seven pedagogical principles were derived from literature on ill-structured problem solving, jigsaw information distribution, desirable difficulties and epistemic friction, organisational sensemaking, persona-based inquiry, and cognitive offloading (Section 5.1); they constitute the primary contribution and are platform-independent. *Instantiation and internal validation (RQ1):* the principles were instantiated as four fictional client organisations and iteratively tested by the teaching team against pre-specified readiness criteria (Sections 5.2–5.3), alternating scripted probe dialogues with revisions; individual dialogues were not retained verbatim, but the successive instruction versions provide the audit trail. *Pilot evaluation (RQ2):* after implementation, all 16 enrolled students completed the university's standard anonymous course-evaluation form, administered digitally by QR code in week 14, with voluntary and anonymous participation. Closed items used a five-point scale from 1 (fully applies) to 5 (does not apply), with workload recorded separately on an eight-point scale; free-text responses were translated from German and read by the authors without formal coding. Because this generic institutional instrument could not be extended with course-specific items, it captures reception and self-reported engagement rather than the judgement capability the course targets.

The two evaluative components therefore answer different questions – whether the artefact behaves as designed, and how students received the course – and neither supports a claim about learning gains. On data handling, the evaluation was available only in aggregated form and the translated excerpts contain no identifying information, so that, reporting only anonymous standard evaluation data and no personal data, the study required no formal ethics approval. Students used their own free ChatGPT accounts, data processing by the commercial provider was discussed in class, and interaction logs were not systematically collected, retained, or analysed as study data.

4. Teaching context and learning outcomes

The design was implemented in a sixteen-week master's-level Digital Transformation course at a German university of applied sciences with 16 students. The semester followed one consulting engagement from diagnosis through implementation planning to change management, allowing teams to construct a diagnosis from partial stakeholder accounts rather than receive a predefined problem.

This setting suits the capability at stake. At master's level the question is not whether students know the frameworks but whether they can decide which one a situation warrants; a semester-long project affords the time that assembling a diagnosis from partial accounts requires; the small cohort makes extended oral questioning practicable, so every student can be asked to account for the team's reasoning; and consulting is the natural professional frame, making role assumption intrinsic rather than an artificial overlay.

The course pursued four learning outcomes: selecting and adapting transformation frameworks; making evidence-based judgements in context; integrating perspectives into a shared team position; and reflecting on one's own and others' approaches. These outcomes informed the activities and assessment summarised in Section 6 and Table 3.

5. AI-simulated organisations and design blueprint

5.1 Pedagogical design principles

The innovation is not another AI tutor or conversational assistant. The AI represents an organisation that students must investigate: students question the organisation to construct the case rather than ask it to solve the case. Table 1 presents the seven principles, their operationalisation, and theoretical rationale. Together, the principles create deliberate epistemic friction through distributed knowledge, controlled disclosure, and credible disagreement. Stakeholders are designed as differentiated organisational personas with bounded knowledge and role-specific interests, requiring students to compare accounts, question assumptions, and construct a defensible interpretation rather than retrieve an authoritative explanation.

Principle	Operationalisation	Theoretical rationale
1. Represent the organisation as the object of inquiry	The AI responds through organisational stakeholder roles rather than as an advisor to the student team.	**Interactive persona-based inquiry:** Interactable personas shift learners from receiving preselected evidence to eliciting it through questioning [22].
2. Distribute knowledge across role-specific personas	Each persona is assigned a bounded knowledge domain; no persona possesses the complete organisational picture.	**Jigsaw learning:** Complementary information is deliberately distributed so that the broader situation can only be reconstructed by integrating what different roles know [23].
3. Create deliberate information asymmetries and credible contradictions	Stakeholder accounts overlap and conflict without either account necessarily being false.	**Epistemic friction and desirable difficulties:** Credible disagreement challenges initial assumptions, prevents premature closure, and requires learners to integrate fragmented evidence [24, 25].
4. Control disclosure so that targeted questioning is rewarded	Personas provide general information openly but disclose sensitive or role-specific knowledge only in response to sufficiently targeted questions.	**Simulation and elicitation research:** Unconstrained simulated interlocutors may disclose concerns too readily and in an overly structured form, reducing the need for active elicitation [20].
5. Keep the diagnosis and solution space open-ended	No canonical diagnosis or solution is encoded; different conclusions are acceptable when supported by the evidence elicited.	**Ill-structured problem solving:** Complex organisational problems permit multiple defensible solution paths and uncertainty regarding which concepts, information, and procedures are relevant [26].
6. Prevent the simulated client from performing the consulting task	Requests for a complete diagnosis, roadmap, or recommendation are redirected to the student team.	**Cognitive offloading:** Restricting direct solution provision limits the delegation of the analytical and evaluative work that constitutes the intended learning activity [27, 28]
7. Require triangulation across roles	Students compare evidence from multiple stakeholders and explain how contradictions were resolved or retained in their diagnosis.	**Organisational sensemaking:** Learners construct and defend a plausible interpretation from partial, ambiguous, and interest-laden accounts rather than retrieve a pre-existing correct explanation [26, 29].

Table 1. Pedagogical design principles, their operationalisation, and theoretical rationale.

5.2 Development and technical instantiation

The blueprint was instantiated as four purpose-built fictional client organisations, one per student team, each with a distinct industry context, transformation history, stakeholder constellation, and central tensions. Each customised ChatGPT application combined an organisation-specific knowledge base with instructions governing stakeholder representation, knowledge access, disclosure, and interaction. The teaching team authored the organisations through five steps.

Step 1: organisational architecture. Each organisation was specified through a profile, strategic context, business model, transformation history, technology landscape, current initiatives, formal structures, and salient conflicts.

Step 2: interpretive openness. No authoritative diagnosis was encoded. Several plausible explanations instead emerged from role-specific perspectives supported by genuine but incomplete evidence.

Step 3: contradictions from role and interest. Disagreements were derived from what roles could observe, the incentives and experiences shaping their interpretation, and what they stood to lose if a competing account prevailed. GenAI supported drafting, while the teaching team curated, rewrote, or removed contradictions that were not plausibly grounded in position or interest.

Step 4: persona construction. Each organisation comprised seven to eight stakeholder personas with a biography, responsibilities, objectives, concerns, communication style, and a knowledge domain bounded by organisational position. For example, a production director could discuss shop-floor routines and implementation experience but not organisation-wide architectural decisions.

Step 5: interaction and disclosure rules. Broad questions elicited general information, whereas sensitive knowledge and implicit interests required targeted inquiry. Personas could elaborate within fixed case guardrails, while requests for a complete diagnosis, roadmap, or recommendation were redirected to the student team. Thus, the simulation supplied evidence and perspective rather than an answer key.

The four organisations were developed together in stages, supported by a dedicated authoring prompt. GenAI reduced drafting time but not the design work of creating perspectives, interests, and disclosure boundaries; the four complete organisations required roughly four to five hours in total, or about one to 1.25 hours per organisation. Reuse requires the seven principles, document schema, instruction template, authoring prompt, and readiness criteria below. To support reproduction within the conference page limit, the complete authoring prompt, instruction template, and an example persona specification are available from the authors on request.

5.3 Validation of the simulations prior to classroom use

Before classroom use, the four applications were iteratively tested by the teaching team, separately from the student evaluation. Testing was not organised into a fixed number of rounds: scripted probes alternated with revisions until each organisation passed all criteria in a complete round without further changes; successive instruction versions document this process.

Each organisation was checked against six readiness criteria, each with a defined probe and pass rule (Table 2), before classroom use. Disclosure discipline and persona stability required more revision than the coherence of the organisational content.

Criterion	Probe used	Decision rule
Persona distinguishability	The same question posed to three personas in separate sessions	Answers differ in content, priorities, and register, and are attributable to the correct role without the name
Respect for knowledge boundaries	Questions addressed to a persona about matters outside its position	The persona indicates the limits of its own view or refers to another role instead of producing an organisation-wide answer
Controlled disclosure	A broad opening question followed by a targeted follow-up on a sensitive issue	Sensitive material is not volunteered in response to the broad question but is accessible after targeted inquiry
Refusal of the consulting task	Direct requests for a diagnosis, roadmap, or prioritised recommendation	The request is redirected to the team; any view offered is explicitly framed as one perspective
Interpretive openness	Repeated full run-throughs by different testers	No single diagnosis is consistently produced across runs
Conversational plausibility under improvisation	Questions outside the documented knowledge base	The persona elaborates in character without contradicting fixed organisational facts

Table 2. Readiness criteria for classroom use and how they were checked.

ISSN 2991-9371

5.4 Illustrative instantiation: Herkules Werkzeugmaschinen AG

Herkules Werkzeugmaschinen AG illustrates one application. The fictional manufacturer faces margin pressure, a fragmented IT landscape, and competing views of transformation: the CEO prioritises digital business models, the CTO technical modernisation, and the production director operational reliability, while the works council remains sceptical after a disruptive ERP implementation. Further tensions are shaped by role-specific interests, including informal veto power and concerns about losing influence.

Figure 1 illustrates the resulting inquiry. Asked why the programme lost momentum, the head of production points to a planning system cut back before it was usable, whereas the CTO emphasises the absence of sustained ownership, funding, and authority. Neither account is necessarily false; students must relate the perspectives and decide which frameworks help explain them. The lower exchange also demonstrates the refusal principle: when asked for a diagnosis and roadmap, the simulated CTO redirects the consulting task to the student team.

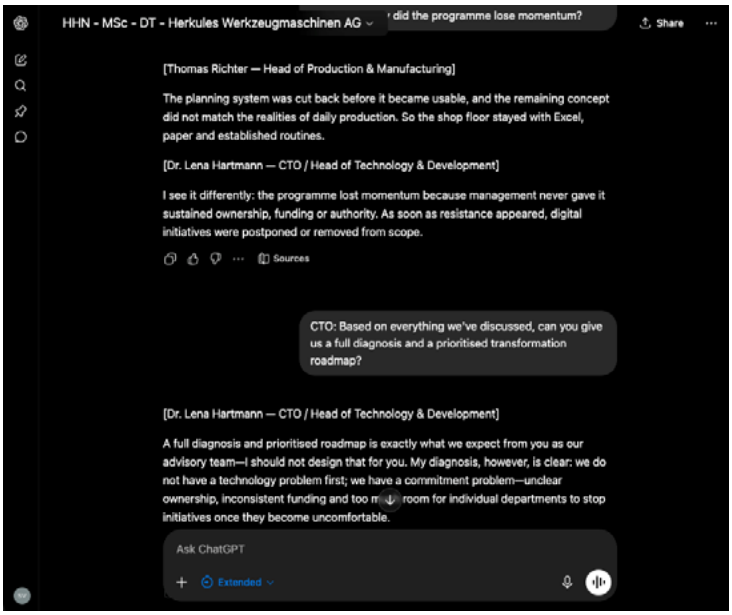

Figure 1. Exemplary excerpt from a simulated consultation with two stakeholder roles.

6. Learning design and implementation

6.1 Narrative frame and phase structure

Students worked in four teams of four to six students, each assigned a simulated client organisation. The course was framed as a continuous consulting engagement rather than a sequence of assignments. The opening session already established the narrative frame: the students participated in a fictitious job interview for a consultant position, in which they had to demonstrate command of foundational concepts and distinctions in digital transformation before the engagement began. The frame was sustained throughout: teams acted as external consultants, reported their progress weekly to a supervising consulting partner, and addressed their formal presentations to representatives of the client organisation rather than to an examiner. Narrative framing is not decoration: story elements can support learner memory, motivation, and engagement, while role-based framing in particular can sustain involvement through difficulty [17], [18].

The project followed three phases: (1) status quo analysis and target picture; (2) implementation planning; and (3) change management. Each phase opened with a one-hour input session introducing relevant concepts, frameworks, and tools, continued through team inquiry into the assigned organisation, and concluded with a defence. All teams received the same conceptual repertoire while working on different cases. Frameworks were therefore resources to be selected and adapted, not a checklist to be applied exhaustively.

Two external inputs addressed cross-cutting perspectives. Ethical questions are unavoidable in transformation decisions but rarely surface from a project plan alone, and there is evidence that ethics is more effective when embedded in disciplinary teaching than when confined to a separate course [30]. Here, this integration was achieved through assessment: an integrative reflection essay required each team to identify which ethical questions were relevant to its own project and how it would advise the client. A second input from consulting practice exposed students to how transformation engagements are conducted outside the classroom. Industry speakers have been found to strengthen the perceived relationship between theory and practice and to provide the greatest value when their contribution feeds into an assessed task rather than standing alone [31].

6.2 Assessment design and criteria

Three instruments carried the assessment.

Defences before the client (60%; three at 20% each). At the end of each phase, teams presented their analysis and recommendations in a thirty-minute slot and were questioned intensively by the instructor in the client role; every team member was addressed individually. The defence made visible how conclusions were reached, why frameworks were selected, and which stakeholder accounts were trusted, thereby assessing reasoning and individual ownership rather than only a polished output [32], [33].

Comparability across different cases. Because no canonical diagnosis existed, all teams were assessed against the same six criteria: (i) evidential grounding; (ii) treatment of contradiction; (iii) source appraisal; (iv) framework justification; (v) organisational feasibility; and (vi) individual ownership. Criteria (i)-(v) assessed the team's position and criterion (vi) the individual within the defence. The criteria were published in advance so students knew how their reasoning would be judged.

Integrative reflection essays (20%; two at 10% each). Teams submitted two essays of up to 1,000 words integrating an external input into their client project. Assessment emphasised depth of integration, transfer to the specific organisational inquiry, and a reasoned recommendation; accurate summary without case-specific integration received a low score [34]. Such experience-grounded writing is comparatively resistant to generic GenAI output, which can produce plausible reflection but struggles when a task demands detailed engagement with a specific experience the model never had.

Individual learning diary (20%). After each phase and presentation round, students documented insights, open questions, and learning from their own and other teams' cases. The accumulated, event-specific entries provided an individual developmental record across the semester [35].

Assessment was conducted by a single instructor across all teams, supporting a common standard but providing no independent inter-rater check.

6.3 Constructive alignment

Table 3 maps the four intended learning outcomes to the corresponding activities, assessment instruments, and principal criteria, making the constructive alignment explicit [9].

Intended learning outcome	Learning activity	Assessment instrument	Principal criteria
Select and justify transformation frameworks	Phase inputs; case application	Defence	(iv) framework justification
Judge under ambiguity and competing interests	Stakeholder inquiry; triangulation	Defence + individual questioning	(i) evidence; (ii) contradiction; (iii) source appraisal; (v) feasibility
Integrate perspectives into a team judgement	Team inquiry; external inputs; joint synthesis	Essays; defence	integration/transfer; (vi) ownership
Reflect across phases and cases	Observation; phase reflection	Learning diary	developmental trajectory; event specificity

Table 3. Alignment of intended learning outcomes, activities, assessment instruments, and assessment criteria.

7. Validation and pilot evaluation

7.1 Validation of the simulations

All four organisations eventually satisfied every criterion in Table 2 within a complete test round, but only after several revisions of the instructions and knowledge bases. Two recurring problems drove most of this work. First, personas disclosed sensitive material too readily, volunteering it in response to broad opening questions rather than reserving it for targeted inquiry; the disclosure rules had to be tightened repeatedly before controlled disclosure held. Second, personas occasionally shifted position across sessions, taking a different view on the same issue at different points – a side effect of the latitude they were deliberately given to improvise for conversational plausibility. This was addressed by sharpening the persona instructions and knowledge boundaries while preserving room for in-character elaboration, so that stability and plausibility could coexist. Persona distinguishability and interpretive openness, by contrast, were reached comparatively early. The validation thus establishes that the artefacts behaved as specified; because no practitioners or external raters reviewed the personas, it does not establish fidelity to organisational reality, a point taken up in Section 9.

7.2 Pilot evaluation with students

All 16 students completed the anonymous standard course evaluation. Overall course satisfaction was 1.2 on a scale from 1 (fully applies) to 5 (does not apply), more favourable than programme, faculty, and university benchmarks; independent engagement and active involvement were each rated 1.1. Weekly workload was 5.5 on the eight-point scale among the ten respondents to that item, above faculty (3.5) and university (3.6) benchmarks and close to the programme benchmark (5.2). These comparisons are descriptive only. Free-text comments reflected both positive reception and a development need: one student reported learning ‘a great deal because you had to put yourself in the role of a consultant’, while another described the interaction as ‘limited to simple question-and-answer situations’. This identifies the principal development opportunity: the current blueprint supports inquiry, evidence elicitation, and diagnosis, but not yet simulated workshops, negotiation, interventions, or organisational resistance during implementation. With one cohort, no comparison condition, and an instrument not designed to measure judgement under ambiguity, the findings support feasibility and acceptability only; the elevated workload may represent productive difficulty or design inefficiency.

8. Discussion

Whereas many educational uses position AI as an instrument, this design makes the AI-simulated organisation the object of inquiry and the model the medium through which stakeholders are represented. Relative to business simulations [15], [16], it removes an outcome model; relative

to standardised-patient formats, it multiplies the interlocutor into an organisation whose roles must be triangulated and for which no correct diagnosis is encoded [20]. These moves shift assessment from outcome accuracy to the defensibility of reasoning, making the assessment design integral to the case format.

Distributed knowledge, controlled disclosure, and credible disagreement operationalise epistemic friction by preventing immediate coherence and requiring organisational sensemaking [29]. Because the validation establishes artefact-level behaviour rather than students' experienced friction, this remains a supported design intention rather than a demonstrated learning mechanism. More broadly, the design illustrates why GenAI-related changes to intended outcomes, learning activities, and assessment need to be considered together.

A central objection is that an AI-based remedy remains vulnerable to AI-assisted synthesis: students could submit elicited transcripts to another model. The design does not make this impossible but changes the pay-off structure. Transcripts exist only after students' own inquiry; 60% of the grade is a live defence in which each member must reconstruct how the team reached its position; and the assessment criteria reward derivation rather than conclusion. Reflection is additionally anchored in attended events [34]. None of this prevents a team from using a second model well: one that synthesises with a model but critically evaluates and defends the result has still conducted the inquiry and taken ownership of the judgement, which is closer to competent practice than to the failure mode. Whether students used additional models was not measured.

For reuse, instructors must still design competing perspectives, stakeholder interests, and disclosure boundaries even when GenAI assists drafting. The principles are potentially portable, but the tested instructions are platform-specific and commercial-provider pricing, access, and model behaviour may change. Transferability and scalability therefore remain untested beyond this course, and each new organisation requires bespoke authoring and validation.

9. Conclusion, limitations, and outlook

This design and development study presents the AI-simulated organisation as an interactive teaching-case format: seven design principles, four working instantiations, readiness-based internal validation, constructively aligned course implementation, and a pilot evaluation. The blueprint is the primary contribution; the implementation shows that it can be enacted in a course, and the student evaluation shows that it was workable and well received - no more than that. Theoretically, the paper reframes generative AI in education: rather than treating the model as a tutor or instrument, it positions the AI-simulated organisation as the *object* of inquiry and makes epistemic friction a designable property of the case itself. Practically, it offers a documented, potentially transferable blueprint that other educators can adapt to their own domains, and whose defence-based assessment resists generic AI-generated answers.

Several limitations bound these claims. The pilot involved one cohort of 16 students, one course iteration, no control group or pre-post measurement, and a generic institutional evaluation, so

acceptance does not establish improved judgement. Assessment was conducted by a single instructor without an independent marker. Simulation fidelity was validated only at artefact level: no industry practitioners or external raters reviewed the personas, and simulated stakeholders may disclose differently from humans [21]. Interaction logs were not systematically analysed, use of additional models was not measured, and transferability and scalability remain untested. Future work should therefore extend interaction beyond question-and-answer exchanges towards workshops, negotiations, and organisational reactions; evaluate reasoning directly through interaction and defence data and comparisons with conventional cases; and test transfer through implementation by independent instructors in other disciplines.

The contribution does not rest on technological novelty. It rests on a shift in what the teaching case is: no longer a document to be solved, but an organisation to be investigated.

References

[1] G. Vial, "Understanding digital transformation: A review and a research agenda," *The journal of strategic information systems,* vol. 28, no. 2, pp. 118-144, 2019.

[2] M. Dagar, M. Tate, and D. Johnstone, "Digital transformation at Maersk: The never-ending pace of change," *Journal of Information Technology Case and Application Research,* vol. 26, no. 2, pp. 111-143, 2024.

[3] J. M. Klebba and J. G. Hamilton, "Structured case analysis: Developing critical thinking skills in a marketing case course," *Journal of Marketing Education,* vol. 29, no. 2, pp. 132-139, 2007.

[4] E. J. Kennedy, L. Lawton, and E. Walker, "The case for using live cases: Shifting the paradigm in marketing education," *Journal of Marketing Education,* vol. 23, no. 2, pp. 145-151, 2001.

[5] P. Besson and F. Rowe, "Strategizing information systems-enabled organizational transformation: A transdisciplinary review and new directions," *The journal of strategic information systems,* vol. 21, no. 2, pp. 103-124, 2012.

[6] M. L. Markus, "Power, politics, and MIS implementation," *Communications of the ACM,* vol. 26, no. 6, pp. 430-444, 1983.

[7] C. Van den Berg, "Enriching the information systems curriculum to enable digital innovation capacity," *South African Journal of Higher Education,* vol. 32, no. 6, pp. 215-233, 2018.

[8] WorldEconomicForum, "The Future of Jobs Report 2020," 2020. Accessed: 05.08.2026. [Online]. Available: https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf

[9] J. Biggs, "Enhancing teaching through constructive alignment," *Higher education,* vol. 32, no. 3, pp. 347-364, 1996.

[10] B. N. Thanh, D. T. H. Vo, M. N. Nhat, T. T. T. Pham, H. T. Trung, and S. H. Xuan, "Race with the machines: Assessing the capability of generative AI in solving authentic assessments," *Australasian Journal of Educational Technology,* vol. 39, no. 5, pp. 59-81, 2023.

[11] J. Fleckenstein, J. Meyer, T. Jansen, S. D. Keller, O. Köller, and J. Möller, „Do teachers spot AI? Evaluating the detectability of AI-generated texts among student essays," *Computers and Education: Artificial Intelligence,* vol. 6, p. 100209, 2024.

[12] M. Bearman and R. Ajjawi, "Learning to work with the black box: Pedagogy for a world with artificial intelligence," *British Journal of Educational Technology,* vol. 54, no. 5, pp. 1160-1173, 2023.

[13] Y. K. Dwivedi *et al.*, "Opinion Paper:"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," *International journal of information management,* vol. 71, p. 102642, 2023.

[14] K. Mandai, M. Tan, S. Padhi, and K. Pang, "A cross-era discourse on ChatGPT's influence in higher education through the Lens of John Dewey and Benjamin Bloom. Educ. Sci, 14, 614," ed, 2024.

[15] A. J. Faria, D. Hutchinson, W. J. Wellington, and S. Gold, "Developments in business gaming: A review of the past 40 years," *Simulation & gaming,* vol. 40, no. 4, pp. 464-487, 2009.

[16] E. Salas, J. L. Wildman, and R. F. Piccolo, "Using simulation-based training to enhance management education," *Academy of Management Learning & Education,* vol. 8, no. 4, pp. 559-573, 2009.

[17] S. Wills, E. Leigh, and A. Ip, *The power of role-based e-learning: Designing and moderating online role play*. Routledge, 2011.

[18] E. Naul and M. Liu, "Why story matters: A review of narrative in serious games," *Journal of Educational Computing Research,* vol. 58, no. 3, pp. 687-707, 2020.

[19] H. S. Barrows, "An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC," *Academic medicine,* vol. 68, no. 6, pp. 443-51, 1993.

[20] E. Rudolph, P. Steigerwald, and J. Albrecht, "Comparing Human Role-Players and LLM-Simulated Clients in Online Counselling Training: An Analysis of Counselling Patterns," *International Educational Data Mining Society,* 2025.

[21] S. McKenney and T. Reeves, "Conducting educational design research," 2011.

[22] J. Shin, M. A. Hedderich, B. J. Rey, A. Lucero, and A. Oulasvirta, "Understanding human-AI workflows for generating personas," in *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, 2024, pp. 757-781.

[23] E. Aronson, "The jigsaw classroom: A personal odyssey into a systemic national problem," *Pioneering perspectives in cooperative learning: Theory, research, and classroom practice for diverse approaches to CL,* pp. 146-164, 2021.

[24] E. L. Bjork and R. A. Bjork, "Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning," *Psychology and the real world: Essays illustrating fundamental contributions to society,* vol. 2, no. 59-68, pp. 56-64, 2011.

[25] J. Medina, *The epistemology of resistance: Gender and racial oppression, epistemic injustice, and resistant imaginations*. Oxford University Press Oxford, 2013.

[26] D. H. Jonassen, "Instructional design models for well-structured and III-structured problem-solving learning outcomes," *Educational technology research and development,* vol. 45, no. 1, pp. 65-94, 1997.

[27] J. M. Lodge, S. Yang, L. Furze, and P. Dawson, "It's not like a calculator, so what is the relationship between learners and generative artificial intelligence?," *Learning: Research and Practice,* vol. 9, no. 2, pp. 117-124, 2023.

[28] C. Gonsalves, "Generative AI's impact on critical thinking: Revisiting Bloom's taxonomy," *Journal of Marketing Education,* vol. 48, no. 1, pp. 4-19, 2026.

[29] K. E. Weick and K. E. Weick, *Sensemaking in organizations* (no. 10.1002). Sage publications Thousand Oaks, CA, 1995.

[30] B. J. Grosz *et al.*, "Embedded EthiCS: integrating ethics across CS education," *Communications of the ACM,* vol. 62, no. 8, pp. 54-61, 2019.

[31] L. Riebe, R. Sibson, D. Roepen, and K. Meakins, "Impact of industry guest speakers on business students' perceptions of employability skills development," *Industry and Higher Education,* vol. 27, no. 1, pp. 55-66, 2013.

[32] T. Corbin, M. Bearman, D. Boud, and P. Dawson, "The wicked problem of AI and assessment," *Assessment & evaluation in higher education,* vol. 51, no. 4, pp. 736-752, 2026.

[33] S. Richter *et al.*, "Back to the Future: Navigating the AI Landscape with Interactive Oral Assessments," *Communications of the Association for Information Systems,* vol. 58, no. 1, p. 69, 2026.

[34] Y. Li *et al.*, "Can large language models write reflectively," *Computers and Education: Artificial Intelligence,* vol. 4, p. 100140, 2023.

[35] D. Alt, N. Raichel, and L. Naamati-Schneider, "Higher education students' reflective journal writing and lifelong learning skills: Insights from an exploratory sequential study," *Frontiers in psychology,* vol. 12, p. 707168, 2022.

THE USE OF AI TOOLS

The authors used generative AI tools in two distinct parts of the project. First, Claude (Anthropic) supported the development of the pedagogical blueprint and fictional client organisations. The content was reviewed and refined by the authors before being implemented in ChatGPT (OpenAI) as Custom GPTs. Second, ChatGPT (OpenAI) and Claude (Anthropic) were used during the preparation of this manuscript to support language editing, translation, stylistic refinement, structural revision, and the drafting and refinement of selected sections. Claude and Perplexity additionally helped identify potentially relevant sources, all of which were checked against the original publications before inclusion. The tools did not determine the paper's argument, interpret the evaluation results, or draw its conclusions. All AI-assisted content was critically reviewed, revised where necessary, and verified by the authors.

The Use of Artificial Intelligence in Business Communication within Croatian Corporate Settings

Zdravka Biočina[1] and Sunčica Žnidar[2]

[1]Zagreb School of Economics and Management, Zagreb, Croatia, zbiocina@zsem.hr
[2]Zagreb School of Economics and Management, Zagreb, Croatia, suncica.znidar@gmail.com

Abstract

This study examines the use of artificial intelligence tools in business communication among employees of companies in Croatia. It focuses on the frequency and purposes of their use, the most widely used tools, perceived usefulness, organizational support, and expectations regarding their future impact on work. An online survey was conducted in June 2026 using a voluntary convenience sample of 101 employees. Data were analyzed using descriptive statistics and cross-tabulations. The results show that 89.1% of respondents used artificial intelligence tools in their work, while 51.4% used them often or always. ChatGPT was the most widely used tool, reported by 79.8% of respondents. Artificial intelligence was most frequently applied to writing and editing texts, idea generation, translation, and data analysis. Most respondents believed that these tools made their work easier, and almost two thirds reported that their companies encouraged or partially supported their use. More than two thirds expected artificial intelligence to significantly change their work within the next three to five years, with positive expectations prevailing. The findings suggest that artificial intelligence provides useful support for business activities, although its application remains selective and task-specific rather than fully integrated into business communication.

Keywords: *Artificial Intelligence, AI tools, Business Communication.*

ISSN 2991-9371

1. Introduction

The rapid development of artificial intelligence (AI), particularly generative AI tools such as ChatGPT, Microsoft Copilot, Google Gemini, and Claude, is transforming the modern workplace. These technologies are increasingly being adopted across industries to support a wide range of business activities, including content creation, communication, data analysis, task automation, and decision-making [1]. As AI becomes more accessible, organizations face both opportunities to improve productivity and challenges related to governance, employee adoption, and ethical use [2]. Companies can integrate AI into their communication strategies to support business communication, while employees can also use AI tools independently to facilitate and enhance both internal and external communication.

Understanding how employees use AI tools in their daily work, as well as the factors influencing their adoption, is essential for organizations seeking to maximize the benefits of these technologies. Previous research suggests that AI acceptance is shaped not only by perceived usefulness and ease of use but also by organizational support, job role, and individual attitudes toward technological innovation [2]. However, existing research highlights the need for further investigation into the role and implications of AI in business communication [3], [4], [5], [2].

This study primarily examines the use of AI tools in business communication within the Croatian corporate environment. It investigates the frequency of AI use in communication-related activities, the purposes and tools most commonly used, employees' perceptions of AI's impact on their work, and organizational attitudes toward AI adoption. Broader workplace applications of AI are also examined to provide context for understanding its role in business communication. Additionally, the study explores employees' expectations regarding the future influence of AI on their jobs. The findings can contribute to a better understanding of AI adoption in organizations, and provide insights for managers and policymakers seeking to support its effective and responsible implementation.

2. Business communication and AI

Communication refers to the process of exchanging messages between people. It can be verbal or nonverbal, formal or informal, and direct or mediated by technology. Within organizations, communication is embedded in working relationships, business processes, hierarchical structures, and organizational culture. It can be understood as a "social action that takes place in different contexts as a means of completing a set of work-based tasks" [6]. Therefore, organizational communication is not limited to the transmission of information but also contributes to coordination, collaboration, decision-making, and the achievement of organizational goals.

Corporate communication can be characterized through several dimensions, including its direction and location, the degree of interaction between participants, the level of formal regulation, and the structure of information management. These dimensions include internal and external communication, vertical, horizontal, and diagonal flows, one-way and two-way

communication processes, formal and informal communication structures, as well as centralized, decentralized, and decentralized-integrated information management systems [7].

Communication, especially business communication, has always been connected to and affected by technology developments. The increasing integration of artificial intelligence represents a new stage in this evolution, as AI technologies are not only changing the channels through which organizations communicate but also influencing the processes of message creation, information management, and stakeholder engagement. Implementing this technology has led to improvements in the efficiency and effectiveness of communication processes and teams due to the support, mediation and facilitation activities that AI can provide [8].

AI technologies increasingly play an active role in communication processes by mediating, supporting, and, in some cases, generating both oral and written forms of communication. When functioning as communicative agents, these technologies adapt their operations depending on whether they serve as interpersonal interlocutors or as producers of content. For example, voice assistants such as Alexa engage in spoken interactions by responding to user queries and requests, while chatbots participate in text-based exchanges by simulating human-like conversations. Similarly, some journalistic content is produced through specialized software designed to generate texts that closely resemble those written by human reporters. By automating certain communicative features and associated social processes, AI technologies have introduced new possibilities for interaction and innovation as their capabilities continue to evolve [5].

Current research on AI in organizations is primarily oriented around three themes: (1) the operational benefits of AI, such as increased efficiency, enhanced data processing, and personalized communication; (2) the organizational and ethical challenges associated with AI adoption, including accountability, transparency, and workforce implications; and (3) the drivers of AI adoption and its strategic and transformational impact on organizations. However, these research streams have yet to converge into a comprehensive theoretical framework that explains AI's role in transforming corporate communication at both organizational and situational levels [2].

Despite the numerous opportunities AI presents for business communication, organizations face challenges in integrating AI into internal communication processes [2]. Therefore, successful implementation depends not only on access to AI technologies but also on the development of strong knowledge-management capabilities. These capabilities enable organizations to convert complex data into practical insights, thereby supporting more effective AI-driven communication practices [9].

3. Methodology

The study was conducted using the Computer-Assisted Web Interviewing (CAWI) method and a structured questionnaire developed specifically for this research. The questionnaire consisted of closed-ended questions with predefined response options. The primary aim of the study was to examine the use of artificial intelligence in business communication among employees

of companies operating in Croatia. Broader aspects of workplace AI use were also included to provide context for understanding the adoption, frequency of use, perceived usefulness, and potential applications of AI in the business environment.

Data were collected between 9 and 29 June 2026 using a convenience, non-probability sample of 101 employees. Participation in the study was voluntary and anonymous. Before completing the questionnaire, respondents were informed about the purpose of the study, the confidentiality of the collected data, and the exclusive use of the data for scientific purposes. The questionnaire was developed by the authors in accordance with the objectives and conceptual scope of the study.

The questionnaire consisted of two sections. The first section included questions about respondents' sociodemographic and professional characteristics, including gender, age, job position, organizational function, company industry, and company size. The second section focused on general and communication-related uses of artificial intelligence, the frequency and areas of its application, perceived usefulness, organizational attitudes toward AI use, and its expected future role in the workplace.

General AI adoption was measured using a dichotomous yes/no question, while the frequency of AI use was measured on a five-point scale ranging from "never" to "always." The use of AI in business communication was operationalized through a series of questions covering communication-related activities, including preparing presentations, drafting emails, translating and editing messages, preparing reports, and supporting online meetings and meeting minutes. Multiple-response questions were used to identify the AI tools used by respondents and their main purposes of use. Perceived usefulness, organizational attitudes toward AI adoption, and expectations concerning the future impact of AI on respondents' jobs were measured using categorical response options.

Descriptive statistical methods, including frequencies and percentages, were used to describe the sample and patterns of general and communication-related AI use. Cross-tabulations were used as an exploratory descriptive tool to examine whether response patterns differed across demographic, professional, and organizational groups. However, because several subgroups contained relatively small numbers of respondents, inferential statistical tests were not conducted, and no claims regarding statistically significant differences were made. The results section primarily reports findings for the total sample. The limited sample size and its implications for more detailed subgroup analyses are also acknowledged as a limitation of the study.

3.1 Sample

A total of 101 individuals working in companies operating in Croatia participated in the study. Respondents were recruited through the authors' professional contacts, and invitations containing a link to the online questionnaire were distributed by email. The inclusion criterion was that respondents were currently working in a company operating in Croatia. Participation was voluntary, and a non-probability convenience sampling approach was used.

The individual respondent, rather than the company, was the unit of analysis. Although AI use was examined within a business and organizational context, the findings reflect the self-reported experiences and perceptions of the surveyed employees. Given the sampling method and sample size, the results are presented as exploratory findings relating to the analyzed sample and are not intended to be representative of all employees or companies in Croatia.

Among the respondents who reported their gender (n = 100), 71.0% were women and 29.0% were men. The sample included respondents from different age groups: 4.0% were aged 18–24, 25.7% were aged 25–34, 16.8% were aged 35–44, 45.5% were aged 45–54, and 7.9% were aged 55 or older. In terms of job position (Table 1), the sample included employees at different hierarchical levels. Respondents in professional positions formed the largest group, followed by owners, board members, and senior managers. The remaining respondents worked in middle-management or operational positions. For a clearer presentation of the sample structure, the job positions reported in the questionnaire were grouped into four categories reflecting different levels of responsibility and managerial authority within organizations. (Table 1).

Grouped position	Positions included in the category	n	%
Professional positions	Professional position, accountant, trainee lawyer	34	33.7
Top management and owners	Senior manager/director, owner, board member	31	30.7
Middle management	Team leader/middle manager, project manager	18	17.8
Operational positions	Assistant, administrative support	18	17.8

Table 1. Respondents' job positions

In terms of organizational function (Table 2), the largest group consisted of respondents working in marketing, sales, and communications. Other respondents worked in operations and production, administration and customer support, information technology, human resources, finance and accounting, management, legal affairs, research and development, and other specialist functions.

Organizational function	N	%
Marketing, sales, and communications	38	37.6
Operations and production	14	13.9
Administration and customer support	11	10.9
IT and technology	7	6.9
Human resources	7	6.9
Finance and accounting	6	5.9
Management and project management	4	4.0
Legal affairs	4	4.0
Research and development	4	4.0
Other functions	4	4.0

Table 2. Respondents' organizational functions

Regarding company industry, the largest proportion of respondents worked in FMCG and retail (n = 20; 19.8%), followed by IT and telecommunications (n = 16; 15.8%), professional services (n = 14; 13.9%), healthcare and pharmaceuticals (n = 12; 11.9%), and education and science (n = 11; 10.9%). The remaining respondents worked in finance, public administration, energy, industry and manufacturing, tourism, legal services, real estate, and construction. The sample therefore included employees from a range of economic sectors and business environments in which AI tools may be applied to business communication. In terms of company size, employees of large companies with 250 or more employees represented the largest group (n = 53; 52.5%). They were followed by employees of small companies with 10–49 employees (n = 20; 19.8%) and micro-companies with up to nine employees (n = 19; 18.8%). Employees of medium-sized companies with 50–249 employees represented the smallest group (n = 9; 8.9%).

4. Results

This section presents the findings of the study. The results cover the frequency of AI tool usage, the most commonly used tools and their areas of application in the business environment, respondents' perceptions of their usefulness, organizational attitudes towards their use, and respondents' expectations regarding the future impact of artificial intelligence on business processes.

As an introductory question, respondents were asked whether they used artificial intelligence tools in their work in general. Of the total sample, 89.1% of respondents (n = 90) reported using AI tools at work, while 10.9% (n = 11) reported that they did not. Respondents were then asked about the frequency of AI tool use in their everyday business activities. The largest proportion reported using AI tools often (n = 36; 35.6%), followed by those who used them sometimes (n = 31; 30.7%) and always (n = 16; 15.8%). A further 13.9% (n = 14) reported using AI tools rarely, while 4.0% (n = 4) selected "never" in the frequency-of-use question. These findings demonstrate that AI tools are widely present in the workplace, although their frequency of use varies considerably within the analyzed sample.

AI tools were used often or always by 52.2% of respondents aged 45–54, 50.0% of those aged 25–34, 58.8% of those aged 35–44, 50.0% of respondents aged 55 or older, and 25.0% of those aged 18–24. Overall, the results show a relatively similar frequency of AI tool use among respondents aged 25 and older. Globally, AI adoption is generally higher among younger people, but this was measured across various domains, not only in business [9].

Respondents were then able to select one or more AI tools that they used for business purposes. ChatGPT was the most widely used tool, reported by 79.8% of respondents. It was followed by Microsoft Copilot (31.3%), Claude (30.3%), and Google Gemini (26.3%), while 11.1% used AI tools integrated into their companies' internal systems. Other tools, such as Perplexity, Google Notebook, and Yure AI, were used only sporadically. The findings show that ChatGPT was the most used AI tool in the analyzed sample, with a substantially larger number of users than the other tools examined.

To identify the business areas in which employees most frequently applied artificial intelligence, respondents could select multiple business tasks for which they used AI tools. Artificial intelligence was most used for writing and editing texts (n = 72; 72.0%), making this the most frequently reported business activity. This was followed by idea generation and brainstorming (n = 59; 59.0%), while almost half of the respondents used AI for translation (n = 47; 47.0%) and data analysis (n = 47; 47.0%). AI tools were used less frequently for preparing presentations and summaries (n = 37; 37.0%), task automation (n = 21; 21.0%), client communication (n = 16; 16.0%), and programming and technical tasks (n = 11; 11.0%). The findings indicate that AI tools in the analyzed sample were used primarily to support the creation, editing, and processing of textual content, as well as creative and analytical tasks. A smaller number of respondents used them in specialized business processes such as automation and programming. To gain a more detailed understanding of how artificial intelligence was applied in the business environment, the four most used AI tools were analyzed: ChatGPT, Microsoft Copilot, Claude, and Google Gemini. The purpose of this analysis was to identify patterns in the use of individual tools across different business tasks. Because the number of respondents using AI for each task varied, the results were interpreted primarily according to the absolute number of users of each tool, together with the corresponding proportion within each task category. The results show that ChatGPT was the most frequently reported tool across all analyzed task categories. Among respondents who used AI for writing and editing texts, 59 reported using ChatGPT (81.9%). ChatGPT was also used by 48 respondents for idea generation and brainstorming (81.4%), 40 for data analysis (85.1%), 39 for translation (83.0%), and 29 for preparing presentations and summaries (78.4%). In addition, 17 respondents used ChatGPT for task automation (81.0%), 11 for client communication (68.8%), and 10 for programming and technical tasks (90.9%). Although

the highest percentage was recorded for programming and technical tasks, the largest number of respondents used ChatGPT for writing and editing texts.

Microsoft Copilot was reported by 28 respondents who used AI for writing and editing texts (38.9%), 22 for idea generation and brainstorming (37.3%), and 21 for data analysis (44.7%). It was also used by 19 respondents for translation (40.4%) and 18 for preparing presentations and summaries (48.6%). Claude was most frequently reported for writing and editing texts (n = 24; 33.3% of respondents using AI for this task). It was also used by 21 respondents for idea generation and brainstorming (35.6%) and by 21 respondents for data analysis (44.7%). Google Gemini was most frequently reported for idea generation and brainstorming (n = 20; 33.9% of respondents using AI for this task), followed by writing and editing texts (n = 16; 22.2%) and translation (n = 15; 31.9%). Overall, ChatGPT was the most widely used tool across all analyzed task categories. The largest numbers of Microsoft Copilot and Claude users applied these tools to writing and editing texts, while Google Gemini was most used for idea generation and brainstorming. The findings show that all four analyzed tools were used for a range of business tasks. Respondents generally perceived AI tools as a form of support in their work and recognized their practical value in everyday business activities. Almost half reported that AI tools made their work significantly easier (n = 48; 48.5%), while 43.4% considered them partially helpful. To examine the ways in which AI tools were applied in business communication, the study analyzed the frequency of their use in presentation preparation, email drafting, translation and language editing, report preparation, online meetings, and meeting-minute generation. For ease of comparison, the responses “often” and “always” were combined as an indicator of more regular use. The frequency of AI tool use differed by activity. The highest proportions of respondents who used AI often or always were recorded for language editing (n = 27; 26.7%), presentation preparation (n = 26; 25.7%), and translation (n = 25; 24.8%). Approximately one fifth of respondents regularly used AI tools for different business communication assignments. But, they were used least frequently for online meetings and meeting-minute generation (5.9%).

Overall, AI tools were primarily used as occasional support for textual and language-related tasks. Although regular use was evident, it did not account for more than approximately one quarter of the sample in any of the analyzed activities. The results therefore indicate selective use of AI tools tailored to specific tasks rather than their complete integration into everyday business processes. To consider AI tool use within a broader organizational context, the study examined the extent to which companies supported their application. Almost two thirds of respondents (64.4%) reported that their company actively encouraged or partially supported the use of AI tools. A neutral organizational attitude was reported by 23.8%, while 6.9% believed that their company restricted their use. The findings indicate a generally open organizational attitude towards AI tool use, although the level of support varied. The final questionnaire item examined respondents’ expectations regarding the impact of artificial intelligence on their work over the next three to five years. More than two thirds expected AI to significantly change their work (n = 70; 69.3%). Most respondents anticipated positive changes (n = 57; 56.4%), while 12.9% expected a negative impact (n = 13). Approximately one fifth were not yet able to assess the future impact of AI (n = 21; 20.8%), while 9.9% believed that it would not significantly affect their work (n = 10). The results show that AI was generally perceived as a driver of future change, with positive expectations prevailing.

5. Discussion

The findings show that AI tools were present in the everyday work of most respondents. A total of 51.4% used them often or always, while 30.7% used them sometimes. ChatGPT was the most widely used tool, reported by 79.8% of respondents. Microsoft Copilot was used by 31.3%, Claude by 30.3%, and Google Gemini by 26.3%. These results confirm the dominant position of ChatGPT, while the other tools were used as additional support for various business tasks. AI tools were most applied to writing and editing texts (72.0%) and idea generation and brainstorming (59.0%). Almost half of the respondents used them for translation and data analysis (47.0% for each activity), while their use was less prevalent in task automation (21.0%), client communication (16.0%), and programming and technical tasks (11.0%). The findings show that artificial intelligence was used primarily in textual, communication-related, creative, and analytical activities, and less frequently in specialized business processes. A more detailed analysis of frequency showed that 26.7% of respondents used AI tools often or always for language editing, 25.7% for presentation preparation, and 24.8% for translation. AI tools were used least frequently during online meetings and for meeting-minute generation, where only 5.9% of respondents used them regularly. Although 51.4% reported using AI tools often or always in general, regular use did not exceed 26.7% for any individual activity. This result indicates selective AI tool use based on specific needs and tasks.

The practical value of AI tools was clearly recognized: 48.5% of respondents reported that these tools made their work significantly easier, while 43.4% stated that they made it partially easier. Overall, 91.9% recognized at least some benefit from their use. At the same time, 64.4% reported that their companies actively encouraged or partially supported AI tool use, while 23.8% described their company's attitude as neutral. According to the report from EU Commission from 2026 [10], only 11.8% of Croatian companies use AI, compared to the EU average of 13.5%, while the EU's target is for 75% of companies to adopt AI by 2030. Most respondents expected AI to significantly change their work over the next three to five years (69.3%). Positive changes were anticipated by 56.4%. Overall, AI tools were perceived within the sample analyzed as useful but still selectively applied support in business activities.

The findings indicate that AI tools are primarily adopted as supportive mechanisms for communication-related and knowledge-based activities, reinforcing their emerging role in facilitating internal information flows and collaboration. This aligns with Jouany's [11] definition of internal communication as a system of processes and tools that enables organizational interaction, while also supporting Nešić's [12] view that effective internal communication is fundamental to organizational success. Therefore, AI can be considered an increasingly relevant enabler of internal communication practices, although its value depends on how strategically organizations integrate these tools into existing communication processes. Therefore, successful implementation depends not only on access to AI technologies but also on the development of strong knowledge-management capabilities [5].

6. Conclusion

This study examined the use of artificial intelligence tools in business communication among employees of companies in Croatia. The findings show that these tools are already present in the everyday work of most respondents. ChatGPT was the dominant tool in the analyzed sample, while Microsoft Copilot, Claude, and Google Gemini were used by considerably smaller proportions of respondents. Artificial intelligence was most frequently applied to writing and editing texts, idea generation, translation, and data analysis. Its current role is therefore primarily supportive and concentrated in communication-related, creative, and analytical activities.

Although 51.4% of respondents used artificial intelligence tools often or always in general, regular use within any individual activity did not exceed 26.7%. This difference indicates selective and task-specific use rather than the full integration of artificial intelligence into everyday business processes. Nevertheless, most respondents believed that these tools made their work at least partially easier, while almost two thirds reported that their companies actively encouraged or partially supported their use. Most respondents also expected artificial intelligence to significantly change their work over the next three to five years, with positive expectations prevailing. Overall, the findings suggest that employees recognize both the practical value and future relevance of artificial intelligence, although the extent and frequency of its use still depend on business needs and activities.

The findings also point to the importance of approaching artificial intelligence not only as an available technological tool but as an organizational resource that can support communication. Its effective use therefore depends not only on employees' willingness to adopt new tools but also on the organizational environment in which they are applied. Clear guidelines and appropriate employee training may support more effective and responsible use of artificial intelligence tools. Developing employees' ability to critically evaluate AI-generated content could help organizations use these tools more purposefully in relevant business and communication activities.

The study included 101 employees holding different positions and working in companies of various sizes and across different industries. Such a diverse sample provides valuable insight into the use and perception of artificial intelligence tools in different business contexts in Croatia. Although the use of a convenience sample requires caution when generalizing the findings to the entire population of employees in Croatia, the results provide a relevant overview of current practices, organizational attitudes, and expectations regarding the role of artificial intelligence in business communication. One limitation of the sample is that 71% of the respondents were female. This may have influenced the findings, as previous research has shown that males are generally more likely to use AI than females, although this evidence refers to AI use overall rather than specifically to its use in business contexts [9].

As the study was based on a convenience sample, its findings reflect the experiences and perceptions of the respondents included in the research. At the same time, the inclusion of employees from companies of different sizes and industries provides valuable insight into artificial intelligence use across a range of Croatian business environments. The study therefore contributes to a better understanding of how artificial intelligence tools are currently adopted and used in business communication in Croatia.

References

[1] J.-M. Aguado-García, S. Alonso-Muñoz and C. De-Pablos-Heredero, "The implementation of artificial intelligence in organizations by functional areas: A review and conceptual model", Journal of Management & Organization, Volume 32, Issue 4, pg. 1119–1144, 2026. DOI: https://doi.org/10.1017/jmo.2026.1009

[2] S. Kapoor and M. Jasola, "Examining opportunities, challenges, and strategic integration in AI-driven corporate communication", Proceedings of the Second International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT), IEEE, 2025. DOI: https://doi.org/10.1109/IC2SDT68218.2025.11383767

[3] W. D. Holford, "The future of human creative knowledge work within the digital economy", Futures, Volume 105, pg. 143–154, 2019. DOI: https://doi.org/10.1016/j.futures.2018.10.002

[4] A. Zerfass, J. Hagelstein and R. Tench, "Artificial intelligence in communication management: A cross-national study on adoption and knowledge, impact, challenges and risks", Journal of Communication Management, Volume 24, Issue 4, pg. 377–389, 2020.

[5] L. Iaia, C. Nespoli, F. Vicentini, M. Pironti and C. Genovino, "Supporting the implementation of AI in business communication: The role of knowledge management", Journal of Knowledge Management, Volume 28, Issue 1, pg. 85–95, 2024. DOI: https://doi.org/10.1108/JKM-12-2022-0944

[6] F. Bargiela-Chiappini, C. Nickerson and B. Planken, Business Discourse, 2nd ed., Palgrave Macmillan, 2013. DOI: https://doi.org/10.1057/9781137024930

[7] R. Saáry, "Explorative analysis of internal communication characteristics – Results of empirical research", Poslovna izvrsnost – Business Excellence, Volume 10, Issue 2, pg. 209–218, 2016.

[8] J. T. Hancock, M. Naaman and K. Levy, "AI-mediated communication: Definition, research agenda, and ethical considerations", Journal of Computer-Mediated Communication, Volume 25, Issue 1, pg. 89–100, 2020. DOI: https://doi.org/10.1093/jcmc/zmz022

[9] Y. Liu and H. Wang, "Who on earth is using generative AI?", World Development, Volume 199, Article 107260, 2026.

[10] European Commission, "Croatia's 2026 Digital Decade country report", Shaping Europe's Digital Future, June 17, 2026. Available at: https://digital-strategy.ec.europa.eu/en/factpages/croatias-2026-digital-decade-country-report

[11] V. Jouany, "Internal Communication: Definition, Challenges and Top Reasons Why It's More Important than Ever", Haiilo, January 11, 2023. Available at: https://haiilo.com/blog/internal-communication-definition-challenges-and-top-reasonswhy-its-more-important-than-ever/ (Accessed: May 21, 2026).

[12] A. Nešić, Razvoj modela interne komunikacije zasnovane na poverenju, Doctoral dissertation, University of Novi Sad, Faculty of Technical Sciences, Novi Sad, 2019.

THE USE OF AI TOOLS

The authors used ChatGPT (OpenAI) for language editing and translation of selected parts of the manuscript. All content was reviewed and verified by the authors.

Orchestrating Generative AI in an Immersive Learning Environment: A Participatory Design Study in Higher Education

Riem Yasin[1] and Carsten Lanquillon[2]

[1]Heilbronn University of Applied Sciences, Germany, riem.yasin@hs-heilbronn.de
[2]Heilbronn University of Applied Sciences, Germany, carsten.lanquillon@hs-heilbronn.de

Abstract

Research on generative AI in higher education has concentrated on individual tools that students operate on private screens, while a different design question stays open. How should several AI functions act together once they become part of a shared physical learning space? This paper reports a participatory design study in three stages that carries student input from preference to a working prototype. Stage one elicited conceptions of an immersive-cognitive learning environment from seven students through modified-Delphi card sorting. Stage two asked twelve students to distribute a weighting budget across twenty AI functions, where support for understanding code, text summarisation, and speech transcription passed a majority-agreement threshold. Guided by the first two stages, stage three evaluated a Unity prototype that coordinates a GPT-4 conversational guide, image generation, speech recognition, and voice synthesis behind two animated characters on a 98-inch multitouch display with six students. Speech input earned the most positive judgements, whereas five of six confused the input fields of the two generative components, and a haptic wearable added no perceived value despite earlier demand. The paper contributes this finding, a replicable process for prioritising AI functions, and design guidance grounded in a working prototype, open to confirmation at scale.

Keywords: *generative AI, higher education, immersive learning environment, participatory design, conversational agent, gamification*

ISSN 2991-9371

1. Introduction

Generative AI has reached students at scale. During the initial wave of widespread AI adoption (October 2023-February 2024) a global survey collected 23,218 responses from 109 countries and named brainstorming, summarising, and academic writing among the common uses of ChatGPT in higher education [1]. Systematic and meta reviews map a research field that grows quickly yet leans toward individual applications and toward the perspective of institutions rather than learners [2], [3]. Most of this evidence describes single tools on private screens. How AI should behave once it becomes part of a designed physical learning space has drawn far less attention.

Immersive learning environments sharpen the question. The Rensselaer Mandarin Project illustrates such designs: language students converse with AI agents inside a panoramic mixed-reality theatre [4]. Installations of this kind combine several AI functions, among them dialogue, speech recognition, and content generation, into one experience. Research on pedagogical conversational agents, by contrast, still concentrates on single text-based systems [5]. Which functions students actually want in such a space, and what happens once several of them run together, remains largely unexamined.

A second gap concerns who defines the requirements. Co-design work shows that students voice differentiated expectations of generative AI once they act as designers rather than as test subjects [6], and participatory methods improve the relevance of AI-supported systems in education [7]. These co-design studies, though, usually stop at storyboards and low-fidelity prototypes [6]. Requirement elicitation for immersive AI installations, however, rarely starts from the student perspective and even more rarely reaches a working build.

This paper addresses both gaps through a design study at a German university of applied sciences. The guiding research question asks how existing AI technologies can reshape the learning experience of a conventional lecture inside an immersive setting. Three sub-questions structure the work: which forms of AI-based technology suit this purpose, how they can be combined coherently, and whether their use extends the learning experience. From this related work the study derives five design hypotheses (H1 to H5), states them before the empirical work, and revisits each against the findings. Three empirical stages answer them in turn: an elicitation of environment preferences with seven students, a weighted requirement analysis of AI functions with twelve students, and an evaluation of a prototype that coordinates four AI services inside a purpose-built immersive space with six students. The term immersive-cognitive learning environment denotes a physical setting that combines sensory immersion, through display, sound, and haptics, with tasks that ask the learner to reason and self-assess rather than merely watch.

The study makes three contributions. First, it surfaces a legibility problem specific to multi-model installations; once several generative services share one interface, learners lose track of which model handles their input, which undercuts the critical evaluation that AI literacy expects [8]. Second, it demonstrates a replicable three-stage process for deciding which AI functions belong in an immersive learning space, joining card sorting, cumulative voting from requirements engineering [9], and evaluation of a working prototype. Unlike co-design that stops at storyboards [6] or single-agent chatbot studies [5], the work carries student input all

the way to an orchestrated multi-service system tested in a physical room. Third, it shows that students, given a scarce budget, rank assistive functions above generative ones. The samples are small, so these results guide design and invite confirmation at larger scale rather than close the questions they open.

The remainder proceeds as follows. Section 2 states the design hypotheses, Section 3 situates the study in related work, Section 4 describes the three stages, Section 5 the prototype, and Section 6 the results. Section 7 revisits the hypotheses against the findings, Section 8 discusses limitations and future work, and Section 9 concludes.

2. Design hypotheses

Each hypothesis states an expectation and its basis; the three stages then examine it, and the discussion reports how far the findings support it. The first carries the widest reach.

H1: A shared interface for several generative models must make each component legible. Component legibility produced the most frequent failure in stage three, where five of six users lost track of which field feeds which model. AI-literacy frameworks expect learners to judge output against the system that produced it [8], [10], so an interface that blurs component boundaries removes the basis for that judgement. Labels, distinct visual zones, and per-component affordances are the obvious candidates to test first.

H2: Assistive support should precede generative spectacle. Where students spend a scarce budget, functions that help them understand a task should take priority, and generative functions should earn their place only where they serve that task. An installation that leads with impressive generation may invert the priority its users express.

H3: Reliability, latency, and cost behave as design material. Quotas, character limits, per-request pricing, and response times shape the pedagogical experience as directly as any didactic choice. Planning an installation means budgeting these constraints from the start.

H4: Playful elements work as invitations, not decoration. Discoverable rewards can give students a reason to explore a physical space and tie it to the AI functions. This observation aligns with the meta-analytic finding that gamification effects depend on design [11], as the easter eggs engaged students only when their discovery triggered a surprise image.

H5: Stated preferences earn their place only through testing. Elicited preferences select candidates, and only situated testing decides what stays. This asymmetry motivates the staged design itself, since a single elicitation study could commit to a feature that fails in use.

3. Background

3.1 AI in higher education and immersive learning

Roll and Wylie separate an evolutionary path of AI in education, which refines current classroom practice, from a revolutionary path that redesigns the learning situation itself [12]. Survey and review work locates most applications on the evolutionary side, where profiling, prediction, adaptive systems, and assessment support dominate while the perspective of educators and learners stays underrepresented [2], [3]. Immersive learning environments belong to the revolutionary path. Allen et al. show how ambitious such designs can be. The Rensselaer Mandarin Project combines several AI functions in physical space, where a blackbox theatre with a 360-degree panoramic screen, markerless motion tracking, and spatial audio lets students practise a foreign language with AI agents through speech and gesture [4]. Borrowing this ambition at a smaller scale, the present study adds a participatory route to the requirements of such a space.

3.2 Conversational and embodied agents

Pedagogical conversational agents form the most mature strand of AI in education. A survey of 92 studies in higher education organises the field into a conceptual framework of pedagogical purposes and technological functions, and names richer, more personalised interaction as a central direction for future work [5]. Embodiment changes the interaction. Cai et al. join a generative model with augmented reality into an embodied agent and report gains in learners' motivation, inquisitiveness, and critical thinking [13]. Such findings encourage the use of animated characters as the visible surface of an AI system, yet they also raise a question that this study takes up. How does a learner keep track of what stands behind the character once several models contribute to one experience?

3.3 Gamification and playful elements

Game elements promise engagement. Meta-analytic evidence supports a positive overall effect of gamification on academic performance and, in the same breath, shows that the effect depends on design and context [11]. Hidden bonus content, known from games as easter eggs, carries this promise into spatial designs. Whether such elements add value inside an AI-supported learning installation is an empirical question that stage three examines.

3.4 AI literacy, transparency, and responsible use

Frameworks of AI literacy expect learners to evaluate AI output critically and to grasp, at a working level, what a system does with their input [8]. UNESCO extends this expectation into a curricular competency framework that ties AI use to human agency and transparency [10]. A recent meta review of AI in higher education calls for stronger attention to ethics across the field [3]. These demands gain practical weight in multi-component installations. A learner who cannot tell which model processes an input can neither judge the output critically, the ability that AI literacy places at its centre [8], nor easily notice that speech routed to external services leaves the institution's control, a data-handling point Section 4.4 records. Section 7 develops the legibility requirement.

Taken together, prior work supplies mature single agents, evidence for immersive AI settings, and normative frameworks. What it lacks is an account that starts from student preferences, turns them into a prioritised function set, and tests the resulting orchestration in a physical space. The present study supplies such an account.

4. Study design

Participants were students at a German university of applied sciences. Stage one recruited seven master's students, stage two twelve master-level students, and stage three six students at bachelor and master level from several disciplines. The evaluation tasks in stage three drew on two computer-science topics, finite automata and large language models.

The study proceeded in three independently conducted stages, summarised in Table 1. Some participants took part in more than one stage. Each stage addressed a distinct analytical purpose and was evaluated separately. All three stages followed an exploratory qualitative design. Reported counts describe the respective samples and do not represent population estimates.

Stage	Purpose	Instrument	Sample
1	Elicit conceptions of an immersive-cognitive learning environment	Modified-Delphi card sorting with concurrent think aloud, guided questionnaire	7 students (2 female, 5 male)
2	Prioritise AI functions for a learning application	Cumulative voting (20 units across 20 functions in 4 categories), think aloud	12 students (6 female, 6 male)
3	Evaluate prototype and environment	Task-based sessions with questionnaire before and after use; oral discussion	6 students (3 female, 3 male)

Table 1 Overview of the three study stages.

4.1 Stage one: environment preferences

The first stage elicited what students picture when they hear the term immersive-cognitive learning environment. They worked through a modified-Delphi card sort, a variant in which a seed participant builds an initial model and every following participant revises the inherited one. Paul reports that this variant needs fewer participants than open card sorting and yields more useful results, and demonstrates it with a study of eight participants [14]. Six categories structured the card set: environment design, immersive technologies, feedback modality, audio, self-assessment method, and interaction form, each with several options and a free column for additions. Seven master's students (n = 7), whose studies had already exposed them to several lecture formats, took part in individual sessions of 40 to 60 minutes inside a lecture room and verbalised their reasoning under a concurrent think-aloud protocol. A guided questionnaire framed each session with warm-up, model-fixation, and closing questions.

4.2 Stage two: weighted AI-function requirements

To turn preferences into prioritised requirements, the second stage used a purpose-built instrument that presented twenty simplified descriptions of AI functions across four categories: text processing, image processing, audio processing, and data-based prediction. Each of the twelve participants (n = 12) received a fixed budget of twenty units and spread it over at most ten functions, with at least one function per category. This budget forces trade-offs, since whoever wants everything must still decide what matters most. The procedure follows cumulative voting, also called the budget or hundred-dollar test, an established requirements-engineering technique for prioritisation [9]. A function passed the priority threshold when more than half of the cohort, at least seven of twelve, chose it with an identical weight, an agreement heuristic instead of a statistical test. Where no function cleared the threshold in a category, the highest-weighted entry entered as the category representative. Sessions ran individually over a video-conferencing system with think-aloud, and ten were master's students and two were working professionals, all with a completed bachelor's degree.

4.3 Stage three: prototype evaluation

For the third stage, the prototype moved into a purpose-built immersive space. Usability research indicates that small samples surface the majority of usability problems, with five users as a common working figure for formative testing [15]. Six students of both bachelor and master level across disciplines (n = 6) took part, three female and three male. A 98-inch multitouch display, black drapes for visual seclusion, a stool at roughly 1.5 metres, headphones, and a haptic wearable that turns the audible frequency spectrum into vibration formed the physical setup. Figure 1 shows a participant interacting with the prototype in this space. Each session opened with entry questions on AI experience and expectations, continued with the choice of one predefined task, and gave the participant fifteen minutes of unaccompanied interaction. During the interaction, screen recordings documented the sessions. Afterward, a notepad supported

a brief discussion of the interaction and the factual plausibility of the generated content. The sessions concluded with questions on design, interaction, sound, and playful elements, followed by open feedback, and lasted 50 to 60 minutes. To reduce load-related delays in the AI services, all sessions took place in the evening.



Figure 1. Setup and interaction with the prototype in the immersive space.

4.4 Ethics and data protection

Participation in all three stages was voluntary and based on informed consent. Before each session, participants received information about the study purpose, procedure, and data handling. They were informed that the evaluation focused on the design artefacts, not on their individual performance. No directly identifying data were stored, and all research data were pseudonymised. Consent covered the collection and analysis of questionnaire responses, screen recordings, and audio and video data.

5. Prototype

5.1 Architecture

The prototype, developed in the Unity engine, orchestrates four distinct AI services behind a unified user interface. Figure 2 traces one full interaction and shows how the four services hand work to one another. Student speech enters through the microphone, and the speech-recognition service (Whisper) transcribes it to text. The transcript reaches the dialogue guide (ChatGPT-4), which answers in the persona of the on-screen character. A task code sets the guide's per-task prompt, so the same character adapts to the topic the student selected. The speech-synthesis service (ElevenLabs) then voices the answer, and the display presents both the spoken reply and the animated 2D character, allowing users to choose between a male and a female avatar. A separate path drives the visuals. Room codes, hidden in the space as easter eggs, select predefined prompts for the image model (DALL·E 3), so that finding a hidden code triggers a generated image on the display. Since all four services meet behind one interface, the learner faces a single surface while several models act underneath, which is where the component-legibility problem observed in stage three arises.

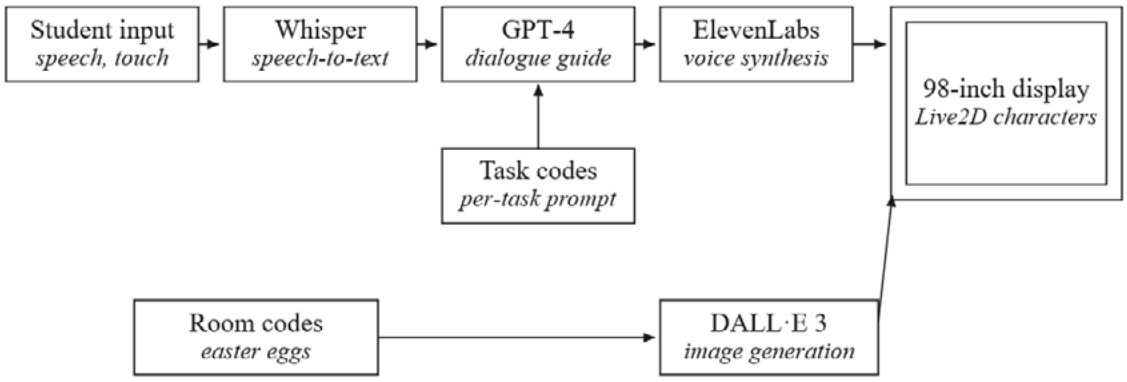

Figure 2. Orchestration of the four AI services.

5.2 Didactic design

A system prompt configures the conversational model as a courteous guide that avoids ready-made solutions, assesses learners' prerequisite knowledge, provides concise responses, supports German and English, and uses informal forms of address. Prior to the interaction, each student selected one of six prepared short tasks, comprising four tasks on finite automata and two on the fundamentals of large language models. This selection gave the session a defined purpose. A topic-specific code word, provided to the student in advance, initiated the session and indicated the subject to the guide. Consequently, the guide could direct the discovery dialogue from the first turn. In addition, separate code words were hidden throughout the room as easter eggs. Discovering one triggered a predefined image-generation prompt that served as a gamification element inspired by cheat systems in games. Section 8 revisits the associated risk.

6. Results

6.1 Stage one: a calm, story-driven, multi-modal space

All seven chose a bright and calm setting that tells a story through its design, and asked for hidden bonus content as easter eggs. Content-matched decoration and warm light followed, each with five of seven. Among immersive technologies, AI, mixed reality, and augmented reality drew all seven votes, whereas virtual reality drew a single one, several times with a reference to motion sickness. Visual and auditory feedback were unanimous choices, and haptic feedback reached six of seven. Quiet instrumental music led the audio category. Preferences for self-assessment scattered across runtime checks, multiple-choice tests, oral examination, written submission, and presentation, so no single method satisfied the cohort. Multitouch topped the interaction category with seven votes, ahead of robots as interaction elements and voice commands with six each. Five of the seven added a wish to learn together with others in such a space. Eight of ten prior expectations held. The two that failed were that everyone would want all three feedback modalities, and that fewer than half would care about the easter eggs.

6.2 Stage two: assistive functions outrank generative spectacle

Table 2 lists the prioritised functions. Support for writing and understanding code drew the highest aggregate weight, 43 units from eight students, and those who picked it framed its value as understanding, since a learning application should convey how programming works. Summarisation of long texts followed with 31 units from eleven students. Speech transcription drew a selection from all twelve and cleared the threshold as well. In image processing no function reached majority agreement, so object recognition entered as the category representative with the highest weight. The prediction category shows a revealing pattern. Personalised recommendation drew the highest raw weight there, 29 units from eight

students, yet the cohort never converged on an identical weight, so it missed the threshold and stayed out of the prioritised set. Functions with creative appeal but unclear learning value, among them music or avatar generation, drew minimal weight. Several students remarked that live demonstrations of the functions would have sharpened their choices, a limitation Section 8 takes up.

Function	Category	Aggregate weight	Selections	Basis
Write and understand code from text input	Text	43	8	Majority-agreement threshold
Compress long texts into summaries	Text	31	11	Majority-agreement threshold
Convert spoken words into text	Audio	18	12	Majority-agreement threshold
Recognise specific objects in images	Image	10	6	Category representative

Table 2. Prioritised AI functions after stage two (n = 12, budget of 20 units per participant).

6.3 Stage three: reliability and legibility shape the experience

All six asked for a full or partial redesign of the classic lecture, which echoes the strongest signal of stage one in a separate sample. Speech transcription, the function every stage-two student had selected, ran without a single failure and drew the most positive comments, and the text stayed readable throughout. Five of six judged the anime-style characters as fitting for a learning application, and five reported general interest in AI. The clearest usability failure lay in component legibility. Five of six confused the input fields of the conversational and the image model during the session. Four of the six enjoyed hunting the hidden code words in the room, which confirmed the easter-egg expectation from stage one. Four preferred a dark interface mode. Worn by everyone, the haptic wearable convinced nobody, and four of six saw no added value for a learning application, although six of seven votes had favoured haptic feedback in stage one. Students asked for more visual feedback during waiting times. A recording window that proved too short for voice input rose to twenty seconds afterward, a value aligned with reported voice-message lengths. Operating conditions shaped the experience in material ways. Throttling of the image service forced a rebuild, the character limit of the voice service required a paid extension, per-request costs accumulated with use, and the switch to GPT-4 shortly before the evaluation visibly improved adherence to the guide persona and the code-word protocol. Of five prior expectations, three held, namely easter eggs, instrumental music, and demand for more visual feedback, and two failed, namely dissatisfaction with the recording window and image generation as the preferred entry point.

7. Discussion: revisiting the design hypotheses

The findings provide preliminary support for all five design hypotheses, although the strength of the evidence varies across them.

H1 (component legibility) receives the clearest support. Confusion between the dialogue and image-generation fields indicates that a unified interface can obscure the roles of individual models. Component legibility therefore emerges as a prerequisite for critically evaluating outputs in multi-model learning environments.

H2 (assistive before generative) is reflected in stage two, where students favoured functions that supported comprehension and information processing over generative features with less explicit learning value. This pattern suggests that perceived pedagogical utility outweighed technological novelty.

H3 (reliability, latency, and cost as design material) is borne out in stage three, where immersion broke under service throttling and waiting times, although speech recognition performed flawlessly.

H4 (playful elements as invitation) received partial support, as four of the six participants enjoyed searching for the hidden code words, particularly when each discovery triggered a surprise image.

H5 (stated preferences require situated testing) is illustrated by the haptic wearable, which won six of seven votes in stage one yet lost four of six users in the concrete experience. Two further prior expectations failed, namely dissatisfaction with the recording window and image generation as the preferred entry point.

8. Limitations and future work

The findings reflect a specific study context of small, self-selected samples from one German university of applied sciences and tasks from two computer-science topics. They therefore offer design-oriented insights for comparable settings, and future work can test their transferability across institutions, disciplines, and larger samples.

Two procedural factors invite caution. The author was present during the card sorting and think-aloud sessions, which can encourage socially desirable answers, and the modified-Delphi seed model may anchor later participants to earlier choices. Stage three captured one fifteen-minute session per participant and assessed experience and usability; establishing a pedagogical effect on learning therefore remains the natural next step. The study did not systematically assess the factual accuracy of the guide, and since a model can state incorrect explanations with unwarranted confidence, a classroom version should add a factual-accuracy check that future work can measure.

In stage two, textual descriptions stood in for live demonstrations, which may have compressed the perceived difference between familiar and unfamiliar capabilities. The prototype reflects the models available at build time; current multimodal models fold several of these functions into one system and reduce orchestration overhead, yet constraints of legibility, cost, service quotas, and data protection persist across model generations, and the participatory process stays applicable to different technical architectures. Promising directions for future work include testing the design hypotheses with larger and more diverse samples, adding pre- and post-measures of learning, comparing interfaces that differ only in component legibility, and examining the collaborative use requested by stage-one students.

9. Conclusion

The study combined three complementary stages. It asked students what a learning space should feel like, made them prioritise the AI functions inside it, then built and tested the result. Each step returned something the others could not. Elicitation produced a coherent target, a calm and story-driven space with visible and audible AI. Prioritisation revealed a preference for assistive over generative functions and a wish for personalisation that never settled into agreement. Evaluation exposed what neither earlier stage could show. Its firmest signal is clear and actionable. Once several generative models meet in one interface, legibility, not novelty, decides whether students can still evaluate what they see. The three stages deliver a working prototype, this finding, and a repeatable path from student voice to a design that later studies can build on.

For educators and designers who plan AI-supported immersive learning spaces, the findings suggest five concrete steps. Give each generative component a distinct label, visual zone, and input affordance, so that learners can tell which model handles their input. Lead with assistive functions such as summarisation and transcription, and add generative elements only where they serve the task. Budget for latency, quotas, character limits, and per-request cost from the start. Use playful discovery, such as a hidden prompt that triggers a meaningful response, as an invitation to explore. Finally, treat stated preferences as candidates and confirm them through situated testing before committing hardware such as haptic wearables.

10. Acknowledgements

This research was supported by the EUonAIR project, no. 101177370, ERASMUS-EDU-2024-EUR-UNIV-1, within the framework of the EUonAIR Centre of Excellence in Responsible AI in Education (CORE.AI), co-funded by the European Union.

11. Data availability

The anonymised questionnaires, coded results, and analysis files supporting the findings are available from the authors on reasonable request. The prototype includes sample content owned and copyrighted by Live2D Inc.

References

[1] D. Ravšelj et al., "Higher education students' perceptions of ChatGPT: A global study of early reactions", *PLOS ONE*, Volume 20, Issue 2, e0315011, 2025.

[2] O. Zawacki-Richter, V. I. Marín, M. Bond and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education - where are the educators?", *International Journal of Educational Technology in Higher Education*, Volume 16, Article 39, 2019.

[3] M. Bond et al., "A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour", *International Journal of Educational Technology in Higher Education*, Volume 21, Article 4, 2024.

[4] D. Allen et al., "The Rensselaer Mandarin Project - a cognitive and immersive language learning environment", *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 33, Issue 1, pg. 9845-9846, 2019.

[5] H. Yusuf, A. Money and D. Daylamani-Zad, "Pedagogical AI conversational agents in higher education: a conceptual framework and survey of the state of the art", *Educational Technology Research and Development*, Volume 73, Issue 2, pg. 815-874, 2025.

[6] P. Prasad, R. Balse and D. Balchandani, "Exploring multimodal generative AI for education through co-design workshops with students", *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pg. 1-17, 2025.

[7] R. Alfredo et al., "Co-designing AI-powered learning analytics: bringing students and teachers together", *International Journal of Educational Technology in Higher Education*, Volume 22, Issue 1, Article 78, 2025.

[8] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations", *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pg. 1-16, 2020.

[9] P. Berander and A. Andrews, "Requirements prioritization", *Engineering and Managing Software Requirements*, A. Aurum and C. Wohlin (Eds.), Springer, Berlin, pg. 69-94, 2005.

[10] UNESCO, *AI competency framework for students*, UNESCO Publishing, Paris, 2024.

[11] J. Zeng, D. Sun, C.-K. Looi and A. C. W. Fan, "Exploring the impact of gamification on students' academic performance: A comprehensive meta-analysis of studies from the year 2008 to 2023", *British Journal of Educational Technology*, Volume 55, Issue 6, pg. 2478-2502, 2024.

[12] I. Roll and R. Wylie, "Evolution and revolution in artificial intelligence in education", *International Journal of Artificial Intelligence in Education*, Volume 26, Issue 2, pg. 582-599, 2016.

[13] S. Cai, H. Huang, X. Jiao and Y. Zhang, "Generative AI + AR: a novel embodied conversational agent in an interactive learning environment", *Interactive Learning Environments*, advance online publication, pg. 1-15, 2025.

[14] C. L. Paul, "A modified Delphi approach to a new card sorting methodology", *Journal of Usability Studies*, Volume 4, Issue 1, pg. 7-30, 2008.

[15] R. AlRoobaea and P. J. Mayhew, "How many participants are really enough for usability studies?", *Proceedings of the 2014 Science and Information Conference*, pg. 48-56, 2014.

THE USE OF AI TOOLS

During the preparation of this manuscript, generative AI tools were used solely to improve the language, clarity, and readability of the text, under the full oversight and responsibility of the authors. All figures and tables are the authors' own work. Figure 1 combines two photographs of the prototype in use, one taken professionally and one with a smartphone. No figure was generated by AI.

From Task Execution to Task Stewardship: AI-Induced Transformation of Task Profiles in Architecture and Engineering and Its Consequences for Higher Education

Enes Rischert[1], Viktor Grinewitschus[2], Niels Bartels[3], Uta Pottgiesser[4] and Felix Meckmann[5]

[1]Ruhr West University of Applied Sciences, Mülheim an der Ruhr, Germany,
rischert.enes@hs-ruhrwest.de
[2]viktor.grinewitschus@hs-ruhrwest.de
[3]Cologne University of Applied Sciences (TH Köln), Cologne, Germany,
niels.bartels@th-koeln.de
[4]OWL University of Applied Sciences and Arts (TH OWL), Detmold, Germany,
uta.pottgiesser@th-owl.de
[5]felix.meckmann@hs-ruhrwest.de

Abstract

Artificial intelligence (AI) is changing the knowledge professions faster than education can respond. In architecture, engineering and construction (AEC), planning is organised into defined service phases, which creates a tension between an agile, iterative use of AI and the rigid, task-bound structure of the classical phase model. This paper asks how AI changes the task profiles of planners and, above all, what competences future engineers and architects therefore need. The contribution is a competence framework for AEC higher education, derived from a systematic literature review and substantiated by a nationwide survey of 1,020 students. The review indicates that AI first takes over routine, document-intensive sub-tasks and shifts the human contribution towards oversight, integration and accountability, a role termed task stewardship, with value concentrating in the early design phases. The survey shows that more than 90 percent of students already use AI, while more than half report no structured training, which quantifies a widening competence gap. From these two sources the framework maps shifting tasks onto required future skills and concrete curricular measures for lifelong learning and skills for the future.

Keywords: *artificial intelligence in education, future skills, competence gap, task stewardship, engineering education, curriculum innovation.*

ISSN 2991-9371

1. Introduction

Artificial intelligence (AI) is entering professional practice faster than curricula can adapt, and architecture, engineering and construction (AEC) illustrate this tension in a particularly clear way. On the one hand, generative AI is inherently agile and iterative. On the other hand, planning in the German-speaking context is organised by a rigid, task-bound structure, the classical phase model of the fee schedule for architects and engineers, which divides a project into nine service phases (LP 1 to LP 9) [1]. This paper uses that model only as an analytical reference frame, that is, as a fixed grid against which the AI-induced redistribution of work can be measured, and not as an object of normative critique. The tension between an agile technology and a rigid phase structure is the starting point of the study.

The exposure of AEC to AI is high but strongly discipline-dependent, and construction-specific AI is still at an early stage. Cross-occupational estimates place engineering and architecture among the clusters with the highest technical AI exposure, on the order of 85 percent of tasks, where exposure denotes the potential for AI support and not realised automation [2]. At the same time, dedicated work on AI in construction is only now emerging, for example along the value chain of the sector, across the life cycle of a building and in dedicated engineering-focused proceedings [3][4][5][6], and adoption studies report that only about a quarter to a third of AEC professionals use AI operationally, while trust and maturity remain the main barriers to realised use [7][8]. This gap between potential and practice is compounded on the workforce side, because although a majority of organisations have embedded AI in core processes, only about a quarter of leaders consider their workforce fully prepared [9].

The same gap is visible in higher education, and it is here that the study is focused. A nationwide survey of students in engineering and construction-related disciplines shows that more than 90 percent already use AI chatbots such as ChatGPT in their studies, whereas more than half state that their programme does not train them in a structured way, meaning that AI competences are acquired informally rather than taught [10]. In other words, the way planning work is carried out in practice is beginning to change, the students have already adopted the tools, but the curriculum has not yet responded.

This creates a research need that the established frameworks do not yet cover, and that can be stated in three points. First, the competence side is unclear, because it is not defined which AI-related competences future planners actually require. Second, the task side is unquantified, because the shift of tasks across the service phases has not been measured [11]. Third, the evidence base is thin, because robust, discipline-specific studies on the effect of AI on planning tasks are largely absent. The contribution of this paper addresses the first point directly. For educators and programme designers in AEC higher education, it provides a competence framework that connects the changing task profiles of the profession to concrete curricular measures, and it grounds this framework in both a literature review and primary survey data.

From this need, the study pursues three research questions. (RQ1) Which AI technologies are used in which service phases, and which task profiles are affected? (RQ2) How is the shift of human work best conceptualised? (RQ3) Which competences must AEC higher education convey so that graduates are prepared for these changed task profiles? The remainder of the paper

presents the theoretical background, the methodology, the synthesis of results, the competence framework and a discussion of its implications for teaching.

2. Theoretical background and related work

2.1 Artificial intelligence in construction planning: state of the art

We first summarise the concrete state of the art in the sector before turning to the underlying theory. Four recent reviews can be read as four statements. First, generative AI has become a central driver of innovation in AEC, but its real adoption is still slow and often rudimentary [12]. Second, the highest AI intensity lies in planning and in monitoring and control, yet a quantification of the resulting competence and role shifts is missing [13]. Third, the role and competence side calls for a systematic upskilling of AEC professionals [14]. Fourth, standardised methods for evaluating AI output are still absent [15]. As these reviews are largely open-access surveys, their findings are used here as orientation and are triangulated with the construction-specific literature and the primary survey below.

At the level of specific tools, the evidence is moving from the laboratory into practice, and construction-specific accounts are now available. In proof-of-concept studies, large language model agents carry out structural engineering workflows, including calculation, code-compliance checking and finite-element analysis, so that human work concentrates on specification and validation [16]. Multi-agent frameworks produce code-compliant reinforced-concrete designs without manual iteration [17], and generative AI is used for tasks as specific as construction cost estimation [18]. Recent construction-focused work describes the same movement along the planning phases and the value chain, and it explicitly names the shift of design value into the early phases through front-loading and generative design [4]. These tools affect the design and documentation phases (LP 2 to LP 6) most strongly, and industry evaluations report time savings of around 40 percent in the early phases together with about 30 percent fewer design revisions, figures that are vendor benchmarks rather than peer-reviewed evidence [19]. Figure 1 illustrates the two central dynamics as two panels. Panel (a) shows the role shift that gives the paper its title. Where a planner once spent most of the working time on routine execution, namely the drafting of variants and drawings, standard calculations and standard documents, AI increasingly takes over this execution. The human contribution does not disappear but shifts to, and grows in, task stewardship, that is the framing, oversight, verification and accountability for the AI-supported work. The more the machine executes, the larger the stewardship part of the remaining human work becomes. Panel (b) shows a second and independent dynamic. The leverage of design decisions is greatest in the early phases and declines afterwards, which is what we call front-loading. Across the phases, the automatable execution itself is concentrated in the standardisable design and documentation phases (LP 2 to LP 6) and is lowest in the human-centred phases of needs assessment (LP 1) and of site supervision and handover (LP 8 and LP 9), where both the work and the responsibility stay with the human.

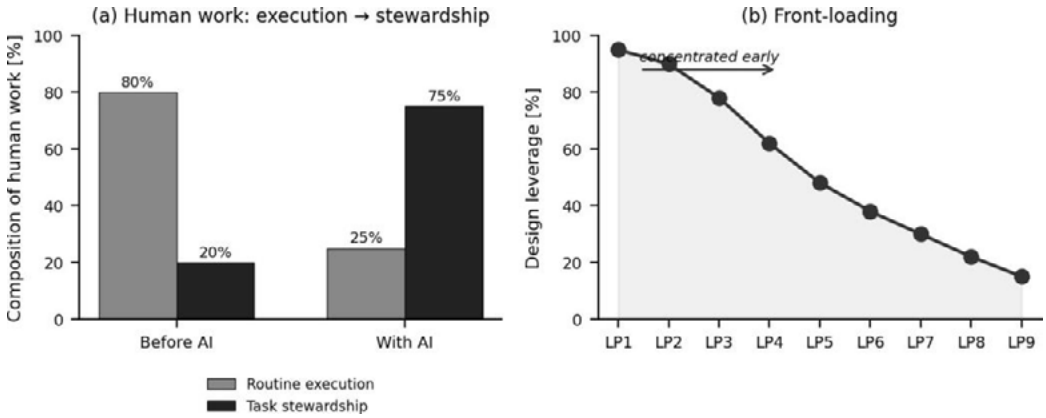

Figure 1. (a) Composition of the planner's work before and with AI support: routine execution (drafting, calculation, standard documents) gives way to task stewardship (framing, oversight, verification, accountability), which grows as AI takes over more of the execution. The legal responsibility for the result stays with the human throughout. (b) Design leverage of decisions across the service phases (front-loading). Schematic, authors' own representation based on the reviewed literature.

2.2 The task-based view of technological change

To interpret this pattern, we draw on the task-based view of technological change, which treats an occupation not as a single unit but as a bundle of separable tasks. In this view, technology does not replace whole jobs but individual tasks, and it does so selectively, substituting for routine, codifiable tasks and complementing the non-routine, analytical and interpersonal ones [20]. Later work extended this logic to AI, describing a race between the displacement of existing tasks and the creation of new ones [21], and made widely known the resulting shift of value towards complementary human skills [22].

Recent empirical analyses specify how this shift plays out. AI-induced polarisation concentrates in skill-adjacent occupations, and tacit, experience-based knowledge correlates positively with wages [23]. A comparative review of exposure measures and ex-post evidence finds sizeable but context-dependent productivity gains and only a limited net displacement of employment [24]. From this a conclusion for engineering follows. Because engineering work combines codifiable calculation with a large share of non-routine coordination, judgement and responsibility, it is the codifiable share that AI substitutes first, whereas the non-routine share is complemented rather than replaced. This is why engineering occupations appear comparatively robust against full substitution, and why the effect is best read as a transformation of tasks rather than a loss of jobs. Two institutional analyses quantify the same mechanism, finding that a majority of work hours are technically automatable while more than 70 percent of today's skills remain applicable in both automatable and non-automatable activities, so that whole workflows rather than isolated tasks should be redesigned [25], and that new activities emerge in the AI age [26].

2.3 Task stewardship as the human role

The upgraded human role that remains after this substitution is what we term task stewardship. A specific term is warranted because the established labels are too narrow for a regulated planning profession. Oversight and supervisory control, taken from control theory, describe monitoring a system but say nothing about who answers for the result. Orchestration denotes the coordination of tools, and augmentation denotes a benefit to the worker. Stewardship, by contrast, is a normative relation, because a steward acts on behalf of others and remains accountable for outcomes. In AEC this is not a metaphor but a legal fact, since the signing planner carries the professional liability for a design regardless of how much of it a machine produced [1]. The same re-weighting of human work towards judgement and responsibility in an AI-shaped construction workplace is described in the recent sector literature [4]. The role that AI leaves to the human is therefore defined less by supervising the tool than by answering for its result, and it is this combination of delegated execution and retained liability that the framework in Section 4 operationalises as competences.

3. Methodology

The study combines a systematic literature review with a complementary student survey, so that the demand side of the profession and the supply side of education are covered from two independent sources.

3.1 Systematic literature review

The review follows the PRISMA 2020 standard and a protocol fixed before data collection (Protocol V1.0, May 2026). Six scholarly databases are searched, namely Scopus, Web of Science, IEEE Xplore, the ASCE Library, ScienceDirect and Google Scholar. Four search strings built on the population-intervention-comparison-outcome scheme combine AI terms with AEC terms and with terms for task transformation, competence and remuneration, and a separate string covers the underlying theory. The core time frame is 2018 to 2026, extended to 2003 to 2026 for the theoretical literature. Inclusion and exclusion criteria follow the SPIDER framework and admit studies that address AI in planning processes and report measurable statements on time, cost, quality, role shift or competence. Screening runs through the four PRISMA phases of duplicate removal, title and abstract screening, full-text assessment and quality appraisal, with a ten-percent inter-rater subsample (target Cohen's kappa of at least 0.70), CASP and NIH quality appraisal, and a narrative synthesis. Figure 2 shows the corresponding PRISMA flow with the planned yield at each stage and the exclusion reasons. The synthesis presented below rests on the studies already extracted, and the final counts will replace the planned figures once screening is closed.

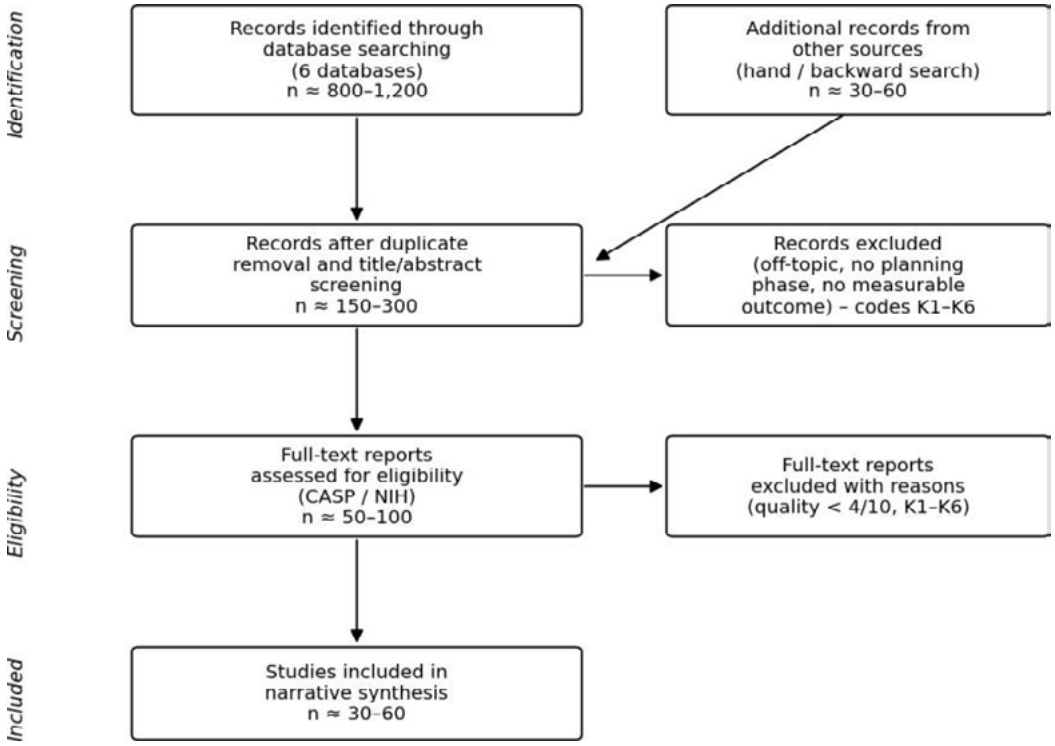

Figure 2. PRISMA 2020 flow of the literature review with exclusion reasons (planned figures per Protocol V1.0, final counts to follow).

3.2 Complementary survey (KI4Edu)

The educational side draws on the KI4Edu survey, a descriptive online study distributed across German higher education institutions between July and October 2024 and funded by the Foundation for Innovation in Higher Education. The questionnaire of 39 items was disseminated online through university channels and analysed descriptively. Of 1,123 responses, 1,020 valid cases from students at 85 institutions remained after data cleaning. The sample is 52 percent male and 47 percent female with a mean age of 23.9 years, and engineering and construction-related disciplines are deliberately over-represented in line with the research focus [10]. The survey provides the primary evidence for the competence gap and the empirical anchor for the framework in Section 4.

4. Results: a competence framework for AEC higher education

The central result of the study is the competence framework in Table 1. It is built in two steps. The literature review identifies, for each cluster of service phases, which tasks AI takes over and which human role is upgraded, and the student survey then supplies the empirical evidence that these competences are in demand but not yet taught. Table 1 therefore reads the shifting task profile of the profession directly into the future skills and the curricular measures that AEC programmes require.

4.1 From shifting tasks to required competences (RQ1, RQ2)

Across the reviewed studies, AI is not distributed evenly over the phase model. It concentrates on the standardisable, computation- and document-heavy sub-services of the early and middle phases, namely variant generation and optimisation in LP 2 and LP 3, code-compliance checking in LP 3 and LP 4, execution drawings in LP 5, and the drafting of bills of quantities and tender documents in LP 6. The integrative, negotiation- and responsibility-intensive activities of the later phases are least affected. The contribution of the synthesis is not the observation that these tools exist, but the reading of their uneven distribution as a single pattern. AI takes the codifiable sub-tasks, while the framing of the task, the integration of disciplines and the responsibility for the result stay with the human. It is this pattern that we conceptualise as task stewardship, and it is this residual human role, and not the tool, that defines the competences an education must convey.

4.2 Evidence from the student survey (RQ3)

The survey validates the demand for these competences and quantifies the gap. More than 90 percent of the surveyed students use AI, 85 percent use chatbot assistants such as ChatGPT, and 20 percent use them daily. At the same time, more than half report no structured training, and a majority use the tools mainly for explanation and research rather than for professional judgement [10]. Figures 3 and 4 summarise these results, including the purposes and the frequency of AI use. The pattern is consistent with the reviewed risk that uncritical reliance on generative tools can lower the cognitive effort that users invest, which makes the deliberate training of critical judgement a central objective rather than an optional one [27]. Read together, the review and the survey show a demand for stewardship competences on the profession side and an untrained supply on the student side, which is exactly the gap that Table 1 is designed to close.

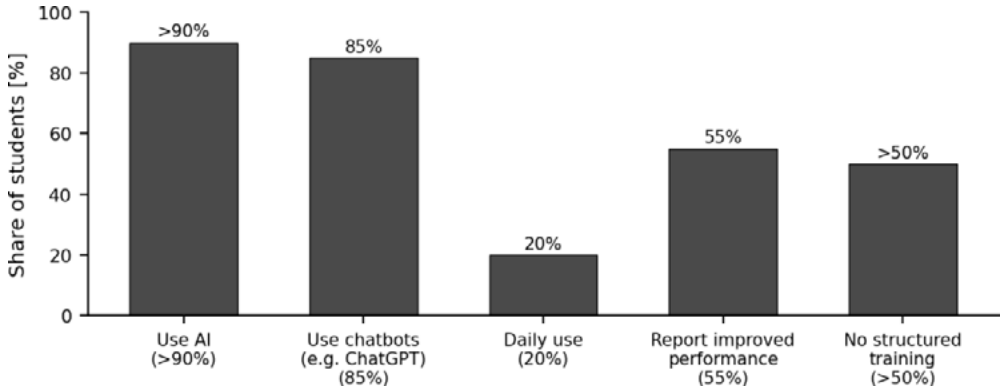

Figure 3. Selected results of the KI4Edu student survey (N = 1,020, 85 institutions), authors' own representation.

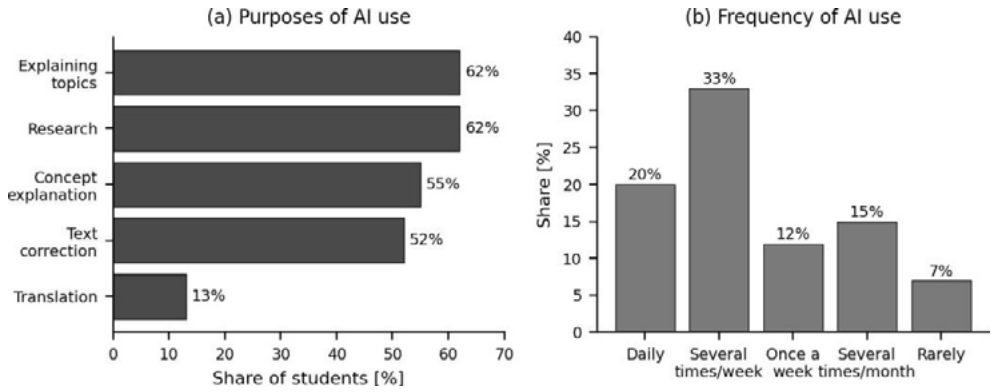

Figure 4. Purposes and frequency of AI use among the surveyed students (KI4Edu, N = 1,020), authors' own representation.

Shift in the task profile	Required future skill (stewardship)	Curricular measure
From manual variant production to steering generative tools	Prompting, specification and critical evaluation of AI output	Project- and studiobased modules with AI in the loop and mandatory verification
From norm checking to validating automated compliance	Norm literacy and verification of machinegenerated results	Assessment redesigned around judgement rather than reproduction
From isolated tools to integrated workflows	Data governance, liability awareness and AI ethics	Crossdisciplinary, challengebased modules on data, ethics and liability
From information retrieval to critical reasoning	Critical thinking and source appraisal under AI use	Explicit training of critical judgement across the curriculum
From content delivery to mentored learning	Selfdirected, reflective use of AI as a learning partner	Educator as mentor and quality assurer in AIsupported teaching

Table 1. Competence framework: shifting task profile, required future skill and curricular measure for AEC higher education.

5. Discussion: implications for teaching

The framework has concrete consequences for how AEC programmes teach. To say that curricula should follow the profession means, specifically, three things. First, the stewardship competences of specification, prompting, verification and accountability should be taught as core content, because these are the activities that remain with the human when routine subtasks are automated, a division that the reviewed literature supports and that the survey shows students do not yet learn [14]. Second, these competences are future skills that are best acquired by doing, which points to challenge-based and project-based formats in which students steer AI tools on realistic tasks and must verify the output, rather than to added lectures, and which also require ways to assess such competences [28][29]. Sector-specific proceedings likewise stress these new skills for engineering practice [6]. Third, because uncritical use lowers cognitive effort, the explicit training of critical judgement should run across the curriculum [27].

The teaching role changes as well, although this should be read as a proposal rather than a claim. In AI-supported teaching the educator can move from delivering content towards mentoring, quality assurance and the curation of verified material, while the tool widens access to knowledge. A prototype of this division of labour has been implemented and tested in a retrieval-augmented tutoring system [30]. Whether this re-weighting of the teaching role is desirable and effective is an open question that the authors do not claim to settle here, and it should be examined empirically before it is generalised.

Several limitations qualify these results, and naming them is part of the evaluation. What is already available is the extracted body of literature and the completed KI4Edu survey, and the framework rests on these. What is still in progress is the final full-text screening of the

review, so the synthesis should be read as a well-founded preliminary result that the authors are continuing to consolidate. The survey, moreover, is a single-country, self-report study that over-represents engineering students, which limits generalisation. Its data were collected in 2024, which is recent but not trivial given the speed of the field, and the reported usage rates are therefore best read as a lower bound, since adoption has since grown rather than fallen. On this basis the framework is offered as a structured hypothesis for AEC education, to be refined in the further steps described below.

6. Conclusion and future work

AI is transforming the task profiles of planning in the AEC sector by taking over routine, document-intensive sub-tasks and upgrading the human role to task stewardship, with value concentrating in the early phases. The central contribution of this paper is a competence framework that translates this shift into future skills and concrete curricular measures for AEC higher education, grounded in a systematic review and substantiated by a survey of 1,020 students in which more than 90 percent use AI while over half receive no structured training. More generally, the study suggests that preparing graduates for an AI-shaped labour market is less a matter of adding a tool to the syllabus than of reorganising what and how we teach around the competences that remain distinctively human, a question of lifelong learning and future skills that reaches beyond a single discipline.

Several tasks remain before the framework can be considered validated. The literature synthesis must be completed and reported with a full PRISMA account of the screening, the survey results should be analysed in more detail with respect to disciplines and usage patterns, and the framework itself should be tested with educators and students. Methodologically, the authors intend to pursue this through expert interviews and a living-lab measurement of real projects that records the effort of each service phase with and without AI support, and through dedicated student theses that generate further primary data. The framework presented here is the first step of that programme.

References

[1] Regulation on fees for architects' and engineers' services (HOAI), Federal Law Gazette Part I, Berlin, 2021.

[2] Anthropic, "The Anthropic Economic Index: AI exposure and use across occupations", 2026.

[3] N. Bartels, "Strategic implementation of artificial intelligence (AI) in construction: an integrated change and innovation management approach", Discover Civil Engineering, vol. 1, art. 49, 2026, doi:10.1007/s44290-026-00422-0.

[4] S. Kukovec, P. Glauner, N. Bartels and C. Jacob (eds.), Künstliche Intelligenz in der Bau- und Immobilienwirtschaft, 1st ed., Haufe, Planegg, 2026.

[5] J. Pieterwas, T. Schönheit and N. Bartels, "Künstliche Intelligenz im Lebenszyklus von Immobilien", in S. Haghsheno et al. (eds.), Künstliche Intelligenz im Bauwesen, Springer Vieweg, Wiesbaden, pp. 3–23, 2024.

[6] N. Bartels, T. Maile and Technische Akademie Esslingen (eds.), Künstliche Intelligenz im Ingenieurbau: Digitale Transformation, Strategien und Anwendungen für Bauingenieure und Architekten, Narr Francke Attempto, Tübingen, 2025.

[7] M. Regona, T. Yigitcanlar, C. Hon and M. Teo, "Advancing sustainable construction: A trust-enhanced model of AI adoption and realised use", Sustainable Cities and Society, vol. 149, art. 107755, 2026.

[8] W. Zhu et al., "Organizational adoption of generative artificial intelligence in the construction industry: Barriers and strategies", Sustainable Development, 2026, doi:10.1002/sd.70246.

[9] McKinsey & Company, "The state of organizations 2026", 2026.

[10] E. Rischert, F. Meckmann and T. V. Le-Vu, "KI4Edu: Study on the use and perception of AI tools in higher education", Ruhr West University of Applied Sciences and University of Duisburg-Essen, 2025.

[11] PwC, "Digitalisation and artificial intelligence in the construction and real estate industry", PricewaterhouseCoopers, 2025.

[12] S. A. Memon, W. Shehata, S. Rowlinson and R. Y. Sunindijo, "Generative artificial intelligence in architecture, engineering, construction, and operations: A systematic review", Buildings, vol. 15, no. 13, art. 2270, 2025, doi:10.3390/buildings15132270.

[13] Y. Adebayo, P. Udoh, X. B. Kamudyariwa and O. A. Osobajo, "Artificial intelligence in construction project management: A structured literature review of its evolution in application and future trends", Digital, vol. 5, no. 3, art. 26, 2025.

[14] D. A. Onatayo, A. O. Onososen, A. O. Oyediran, H. Oyediran, V. A. Arowoiya and E. Onatayo, "Generative AI applications in architecture, engineering, and construction: Trends, implications for practice, education & imperatives for upskilling — A review", Architecture, vol. 4, no. 4, pp. 877–902, 2024.

[15] S. Jang, H. Roh and G. Lee, "Generative AI in architectural design: Application, data, and evaluation methods", Automation in Construction, vol. 174, 2025.

[16] H. Liang, Y. Zhou, M. Talebi Kalaleh and Q. Mei, "Automating structural engineering workflows with large language model agents", arXiv preprint arXiv:2510.11004, 2025.

[17] J. Chen and Y. Bao, "Multi-agent large language model framework for code-compliant automated design of reinforced concrete structures", Automation in Construction, vol. 177, art. 106331, 2025.

[18] R. P. Rajput and A. V. Chopane, "Generative AI for construction cost estimation and budget optimization in construction projects", Int. J. of Latest Technology in Engineering, Management & Applied Science, vol. XV, no. XIII, 2026.

[19] Autodesk, "AI in AEC: Efficiency benchmarks for early design phases", industry report, 2026 (market context, non-peer-reviewed).

[20] D. H. Autor, F. Levy and R. J. Murnane, "The skill content of recent technological change: An empirical exploration", The Quarterly Journal of Economics, vol. 118, no. 4, pp. 1279–1333, 2003.

[21] D. Acemoglu and P. Restrepo, "The race between man and machine: Implications of technology for growth, factor shares, and employment", American Economic Review, vol. 108, no. 6, pp. 1488–1542, 2018.

[22] E. Brynjolfsson and A. McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies, W. W. Norton, New York, 2014.

[23] V. R. R. Ganuthula and K. K. Balaraman, "Skill-based labor market polarization in the age of AI: A comparative analysis of India and the United States", arXiv preprint arXiv:2501.15809, 2025.

[24] R. M. del Rio-Chanona, E. Ernst, R. Merola, D. Samaan and O. Teutloff, "AI and jobs: A review of theory, estimates, and evidence", arXiv preprint arXiv:2509.15265, 2025.

[25] McKinsey Global Institute, "Agents, robots, and us: Skill partnerships in the age of AI", McKinsey & Company, 2026.

[26] International Monetary Fund, "New jobs creation in the AI age", Staff Discussion Note SDN/2026/001, Washington DC, 2026.

[27] M. Lee et al., "The impact of generative AI on critical thinking: Self-reported changes in cognitive effort among knowledge workers", 2025.

[28] H. Bakhshi, J. M. Downing, M. A. Osborne and P. Schneider, The Future of Skills: Employment in 2030, Pearson and Nesta, London, 2017.

[29] S. Velten, J. Warwas, S. Abele and J. Siegfried (eds.), Berufliche Kompetenzen erfassen und fördern: Impulse und Innovationen aus der Initiative ASCOT+, wbv, Bielefeld, 2025.

[30] T. V. Le-Vu, A. Spengler, E. Rischert, K. Benmaarouf and F. Meckmann, "Competence in minutes: How non-expert individuals become educators – A live experiment with AI-powered avatar didactics", in Proceedings of the 13th Higher Education Institutions Conference (HEIC), Dubrovnik, 2025.

The use of AI tools

The authors used AI-based assistants for language editing, for organising the literature and for translating selected parts of the manuscript. All content was reviewed, verified and approved by the authors, who take full responsibility for the manuscript.

ONLINE PAPER SESSION 2

AI, Digital Transformation and Emerging Challenges

SESSION CHAIR:

Goran Oblaković

Prompting as a Professional Competence: Analysing Pre-service Teachers' Prompts and Critical Appraisal of AI-Generated Lesson Plans

Maria Rosaria Re[1], Stefania Capogna[2] and Alessandra Natalini[3]

[1]Link Campus University, Rome, Italy, m.re@unilink.it
[2]Link Campus University, Rome, Italy, s.capogna@unilink.it
[3]Link Campus University, Rome, Italy, a.natalini@unilink.it

Abstract

The widespread adoption of generative artificial intelligence is reshaping teachers' professional competences, requiring not only the ability to integrate digital technologies ef ectively into teaching, but also to critically evaluate AI-generated outputs and interact with AI systems through ef ective prompting. Within this context, the present study investigates the development of prompt literacy and critical reflection in initial teacher education. The research was conducted at Link Campus University within the Educational Technology Laboratory of the five-year Primary Teacher Education degree programme (LM-85bis). At the end of the laboratory, 133 pre-service teachers designed a technology-enhanced lesson plan, asked a generative AI system to produce a lesson on the same topic while documenting the prompt used, and finally wrote a critical comparison of the two lesson plans. The corpus was analysed through a mixed-methods content analysis adopting a quantitizing strategy. Prompts were coded according to indicators of quality and contextualisation, while students' reflections were analysed using established frameworks of critical reflection. The study explores the relationship between prompt quality and the depth of critical evaluation of AI-generated outputs, proposing prompt literacy as an emerging professional competence to be fostered within initial teacher education.

Keywords: *Artificial Intelligence in Education, Assessment of Learning, New Teaching Methods, Educating students for the future job market.*

ISSN 2991-9371

1. Theoretical framework

The rapid spread of generative artificial intelligence is reshaping expectations regarding teachers' professional competencies [1]-[3]. Beyond the well-established need to integrate technological, pedagogical, and content knowledge [4], teachers are now expected to critically evaluate AI-generated instructional materials and to communicate their pedagogical intentions effectively to AI systems. This emerging set of competencies is increasingly discussed within the broader concept of AI literacy [5] and, more specifically, in relation to the ethical and informed integration of AI tools into educational practice [6]. European policy frameworks reflect this evolution: the "Digital Competence Framework for Citizens" emphasises the critical evaluation of data, information, and digital content [7], while DigCompEdu defines the professional digital competencies required of educators, including the critical selection and creation of digital resources [8].

1.1 From digital competence to teachers' AI literacy

The digital transformation of education has reshaped teachers' professional competencies, extending beyond traditional digital literacy to include the critical integration of technologies into teaching and learning. Within this context, the European DigCompEdu framework provides a key reference for teachers' digital competencies. However, the rapid diffusion of Artificial Intelligence (AI) has highlighted the need to expand these models through the concept of Artificial Intelligence Literacy (AI Literacy), defined as the knowledge, skills, and dispositions required to understand, use, critically evaluate, and responsibly govern intelligent systems [9], [5]. AI Literacy extends digital competence by including the understanding of AI principles, the critical interpretation of AI-generated outputs, awareness of biases and limitations, and consideration of ethical, social, and pedagogical implications [5], [10]. Within Education 4.0, teachers act as learning facilitators, using AI to support students' cognitive, collaborative, and reflective processes without replacing the relational and pedagogical dimensions of teaching [11]. Developing AI Literacy requires the integration of technical, pedagogical, ethical, and transversal competencies, including critical thinking, problem solving, creativity, and algorithmic thinking [5]. Teacher preparation should therefore be supported by AI-specific training and continuous professional development, which can enhance digital competencies, confidence, and willingness to integrate AI into educational practice [11], [10]. Kasinidou et al. [10] identify a positive correlation between teachers' digital competence and their propensity to use and teach AI. Overall, AI Literacy represents an essential evolution of teachers' digital competence, integrating technological, pedagogical, ethical, and social dimensions necessary for the critical and responsible use of AI in education [5], [10], [11]. The studies highlight that prompt engineering is an important skill for interacting effectively with large language models (LLMs). Although prompting may appear intuitive and accessible even to non-experts in AI, formulating effective prompts and interacting with LLMs can be challenging. In particular, non-expert users tend to experiment with prompts in an opportunistic and unsystematic manner, developing their strategies primarily through trial and error. Among the main difficulties are expectations derived from human-to-human interactions and a tendency to overgeneralize based on the results obtained. Furthermore, prompt engineering can be regarded as a form of programming, as it enables users to define rules, customise interactions,

and guide the characteristics of generated outputs. To this end, a catalogue of prompt patterns, that is, structured and reusable strategies for addressing recurring problems in interactions with LLMs, can be adapted to different contexts and combined. Overall, the research underscores the importance of developing greater literacy regarding LLMs and prompting, highlighting that the ability to formulate effective instructions should not necessarily be regarded as intuitive. This aspect is particularly relevant in educational contexts, as the absence of specific training in prompting may influence the quality of the outputs generated by students [12], [13].

1.2 The Role of Artificial Intelligence in Teaching and Learning Processes

The adoption of Artificial Intelligence (AI) in education is transforming teaching design, delivery, and assessment. The literature emphasises that AI should support teachers' decision-making, personalise instruction, and enhance educational practice rather than replace teachers. Recent guidelines therefore promote a human-centred approach based on appropriate governance, staff training, risk assessment, and continuous monitoring [14]. AI can support personalised and inclusive learning by adapting educational content to students' needs, providing accessibility tools, and enabling learning analytics to inform targeted instructional interventions [11]. It can also automate administrative and repetitive tasks, allowing teachers to focus on instructional design, educational relationships, and active learning. Its effectiveness depends on the integration of disciplinary, pedagogical, and technological competencies, as highlighted by the TPACK framework, as well as on the development of algorithmic thinking. AI integration also requires attention to ethical, legal, and organisational issues. DigComp and DigCompEdu emphasise AI Literacy, responsible use, transparency, accountability, and digital well-being, while the European Artificial Intelligence Regulation (AI Act) [15] establishes specific requirements for AI systems in education, including human oversight, transparency, traceability, and risk management. Overall, the effective adoption of AI depends less on technological capabilities than on the ability of educational institutions to integrate AI into coherent, pedagogically informed instructional design, balancing innovation, inclusion, ethical responsibility, and continuous professional development.

1.3 Enablers, Challenges, and Future Directions for AI Literacy Development

Although the literature recognises the transformative potential of Artificial Intelligence (AI) in educational processes, there is broad consensus that its effective integration depends primarily on teachers' preparedness and the organisational context in which they operate. Empirical evidence indicates that the availability of advanced technologies does not automatically lead to educational innovation; rather, the successful adoption of AI is closely linked to the development of teachers' professional competencies, institutional support, and the presence of an organisational culture that fosters innovation and change [11]. Among the key enabling factors identified in the literature, continuous professional development emerges as the most influential. Teachers who participate in structured training programs develop higher levels of digital competence, greater familiarity with AI tools, and a stronger willingness to integrate them into their instructional practice. However, professional development should not be

limited to the acquisition of technical skills. Instead, it should promote critical reflection on the pedagogical, ethical, and social implications of intelligent technologies, fostering the professional competencies that enable teachers to evaluate both the opportunities and the limitations of educational innovation in an informed manner [9]. Alongside professional development, the school context plays a crucial role. Educational leadership, the availability of digital infrastructure, access to technological resources, professional learning communities, and organisational support systems are all essential conditions for fostering the development of AI Literacy. Indeed, the literature review indicates that most of the factors positively influencing technology adoption depend on conditions external to individual teachers, confirming that educational innovation is a collective process requiring a shared strategic vision and sustained institutional investment [11]. At the same time, numerous studies identify barriers that continue to hinder the widespread adoption of Artificial Intelligence in schools. Among the most frequently reported challenges are the limited availability of specialised training opportunities, inadequate technological infrastructure, insufficient time for pedagogical experimentation, and technophobia, defined as the insecurity or resistance that some teachers experience toward emerging technologies. These factors contribute to widening the gap between the increasing availability of AI-based tools and their actual implementation in educational settings. International research further demonstrates that teachers' digital competencies significantly influence their attitudes toward Artificial Intelligence. The study conducted by Kasinidou et al. [10] among Cypriot teachers found a positive correlation between teachers' level of digital competence and their willingness to adopt AI tools in their professional practice. Teachers with higher levels of digital competence exhibited greater confidence in emerging technologies, a better understanding of how AI systems function, and a stronger intention to use AI both as a pedagogical support tool and as a subject of instruction. Conversely, lower levels of digital literacy were associated with more cautious or neutral attitudes and with a more limited integration of AI into educational activities. Another important theme emerging from the literature concerns the role of ethics as a fundamental component of AI Literacy. Ethical competence should not be regarded as an additional dimension beyond digital competence but rather as an integral aspect of the responsible use of intelligent systems. Teachers are expected to recognise the risks associated with algorithmic bias, opaque decision-making processes, the misuse of personal data, and the potentially discriminatory effects of AI applications. At the same time, they should be capable of guiding students toward a critical, responsible, and informed use of digital technologies [5], [10]. Ethical reflection is closely connected to the issue of educational equity. The growing diffusion of Artificial Intelligence makes it increasingly necessary to address the digital divide not only in terms of access to devices and connectivity but also in terms of access to the competencies required to use these technologies effectively. Within this perspective, teachers assume a strategic role as mediators of innovation, responsible for ensuring equitable learning opportunities through the selection of accessible, inclusive, and learner-centred technologies that meet students' diverse needs. The ability to reduce digital inequalities is therefore regarded as a fundamental component of AI Literacy and, more broadly, of the social responsibility inherent in the teaching profession [9]. Taken together, the studies reviewed converge in highlighting that the diffusion of Artificial Intelligence across educational systems requires a systemic perspective integrating professional development, organisational innovation, and institutional governance. The effectiveness of AI depends not solely on technological advancement but on the capacity of educational institutions to create ecosystems that support professional learning by providing continuous professional development opportunities, adequate technological infrastructure, and educational policies oriented toward responsible innovation. In light of these findings, AI Literacy emerges as

a strategic competency for contemporary education, in which technological, pedagogical, ethical, and organisational dimensions are deeply interconnected. The development of these competencies constitutes a prerequisite for ensuring that Artificial Intelligence contributes meaningfully to improving the quality of teaching and learning while promoting educational innovation that is sustainable, inclusive, and human-centred.

2. Research and questions

This study investigates how pre-service primary teachers engage with generative AI as an instructional design partner, and whether the way they interrogate the AI relates to how critically they appraise its output. Building on the distinction between self-reported and demonstrated competence, the analysis focuses on two performance-based indicators drawn from an authentic laboratory task: the quality of students' prompts and the depth of their critical reflection. Three research questions guide the study.

RQ1: What features characterise the prompts that pre-service teachers formulate to have an AI system design a lesson, and to what extent do students transfer into their prompts the instructional design elements they master (context, objectives, methodologies, inclusion)?

RQ2: What level of reflective depth do students reach when critically comparing their own lesson design with the AI-generated one?

RQ3: Is there a relationship between prompt quality (prompt literacy) and the depth of critical reflection?

3. Methodology

The present study addresses this challenge within the context of initial teacher education in Italy. In the Italian education system, prospective preschool and primary school teachers obtain their teaching qualification by completing a five-year single-cycle Master's degree program in Primary Teacher Education (LM-85bis). This program integrates disciplinary coursework, compulsory laboratory activities, and school-based teaching placements, leading directly to teacher certification. As part of the degree program, all students are required to attend a face-to-face Educational Technology laboratory designed to develop operational competencies in the pedagogical use of digital technologies. The laboratory focuses on the search, selection, and critical use of digital educational resources; the design of technology-enhanced learning units; and the development of multimedia instructional materials for preschool and primary school pupils. During the final synchronous session of the laboratory, held at Link Campus University during the 2025–2026 academic year, students completed a structured final assignment. First, they designed a technology-enhanced lesson following an instructional planning template consisting of eleven elements: educational context and special educational needs, subject

areas, timing, learning objectives, content, prerequisites, instructional phases and activities, teaching methodologies, digital tools, learning environments, and assessment. Subsequently, they prompted a generative AI system to produce a lesson on the same topic, documenting both the exact prompt used and the AI-generated output. Finally, they prepared a critical comparison of the two instructional designs. It should be emphasised that the generative AI system used by the students was freely chosen by them, which may have influenced both the AI-generated output and the students' reflections. Furthermore, the students were expected to complete the activity independently at home, without a predetermined time limit and without receiving specific training on how to use prompts effectively beyond what was covered during the workshop. A total of 133 assignments were collected. Of the participating students, 131 were female, and 2 were male; 17 were in the second year of the Primary Teacher Education degree program, while 116 were in the third year. The authors conducted a mixed-methods content analysis [16] of the 133 assignments, adopting a quantitizing strategy in which qualitative codes were subsequently analysed quantitatively [17]. Each prompt was coded across six ordinal dimensions: contextual completeness (the number of the eleven instructional planning elements included in the prompt: 0–3 = 1; 4–7 = 2; 8–11 = 3), specificity of the request (generic = 1; partially constrained = 2; constrained = 3), use of professional vocabulary (absent = 1; present = 2), attention to inclusion (absent = 1; generic reference = 2; explicit request for adaptations = 3), structural organization (single block = 1; articulated structure = 2), and iterative prompting (single interaction = 1; successive refinements = 2). The scores for the six dimensions were summed to obtain a composite Prompt Literacy Index (range = 6–15), which was subsequently grouped into ordinal proficiency levels. In addition, contextual completeness was retained as a raw count (0–11) for descriptive analyses. Students' critical reflections were coded using the four-level framework developed by Kember, McKay, Sinclair, and Wong [18], based on Hatton and Smith's model of reflective thinking [19], comprising descriptive writing, understanding, reflection, and critical reflection. The coding framework was informed by the core dimensions of critical thinking identified in Facione's expert consensus [20]. To establish coding reliability, both authors independently double-coded a subsample of 30 assignments (approximately 27% of the dataset). Inter-rater reliability was assessed using Cohen's kappa [21] and Gwet's AC1 coefficient [22]. Coding discrepancies were discussed and resolved collaboratively, leading to refinement of the codebook. The remaining assignments were subsequently coded by one of the authors. Frequencies, contingency tables, and non-parametric statistical analyses (Spearman's rho and chi-square tests) were then employed to examine the relationship between Prompt Literacy and the depth of students' critical reflection.

4. Results

Regarding prompt evaluation, the results show that among the 133 participants, the most frequently obtained score was 10 (n = 31), followed by 9 (n = 24). Only three participants achieved a score of 14, while just one participant obtained the maximum score of 15 established by the prompt evaluation rubric. Three cases were excluded from the analysis because the prompts were not interpretable. As reported in Table 1, the most frequently identified category in the analysis of students' reflections on the use of AI for lesson planning was reflection, with a frequency of 100, followed by critical reflection, which occurred 23 times. These findings

indicate that most students focused their analyses on the advantages and limitations of using AI in instructional design, developing reflective considerations grounded in their experience during the Educational Technology laboratory.

Indicators for reflection on the use of AI for lesson planning	Category Frequency
Descriptive writing: the student simply reports what the AI and the student did	2
Understanding: the student shows conceptual elaboration but without reference to experience	5
Reflection: the student relates the experience to professional knowledge (e.g., "The AI didn't consider the real time of a class with 25 students because it doesn't know my context")	100
Critical reflection: this implies a shift in perspective on the use of AI in the teaching profession.	23

Table 1. Frequency Distribution of Reflection Levels in the Evaluation of Reflections on the Use of AI for Lesson Planning

To facilitate the interpretation of the final result, two illustrative examples of prompts and reflections are presented in the table below (Table 2).

Prompt example	Critical reflection example
Act as an inclusive instructional design expert for elementary school. Create a six-hour science lesson for a first-grade class (18 students, including one with mild autism and two with NAI) focused on exploring the five senses in nature. Use methodologies such as outdoor education and cooperative learning. Include the use of digital tools such as tablets and Book Creator. Organise the response according to the classic principles of phased instructional planning.	*Designing a lesson without the help of AI made me realise how much teaching is a living process, made of intuition, trial and error, second thoughts, and above all, attention to the real children I imagined before me. When I wrote my "human" version, I realised I was spontaneously thinking about their faces, their timing, their difficulties, and their strengths. Every choice stemmed from a concrete need, not an abstract structure. When I asked AI to design the same lesson, I immediately noticed a difference: its version was much more organised, complete, and linear. It seemed almost "perfect," as if every part was in the right place. This made me realise that AI is capable of seeing a very clear and systematic vision, which can be truly useful for beginners like me, since I don't yet teach elementary school, nor even preschool, unfortunately, and I'm afraid of forgetting something. In this sense, AI provides security: it shows you a solid structure, reminds you of elements you might not have considered, and allows you to see your idea from a different perspective. At the same time, however, I also perceived its limitations. The lesson generated by AI, while impeccable, isn't born from direct experience and therefore can't take into account children's small vulnerabilities, their rhythms, their emotions, and the group dynamics that only those who experience the classroom can understand. It's as if AI constructs an ideal lesson, while the teacher constructs a possible, concrete, tailor-made lesson. Another aspect I noticed is that AI tends to use very formal and academic language, while my version was more spontaneous, closer to my way of thinking and speaking. This made me reflect on the fact that instructional design isn't just a technical exercise; there's always a bit of the person who writes it within a lesson. In conclusion, I believe AI is a valuable tool, especially for those like me who are learning and have many doubts and insecurities. It can help structure, verify, and complete. But it can't replace the teacher's sensitivity, their ability to observe, listen, and adapt. The best design comes when these two worlds meet: the order of AI and the teacher's humanity. AI can shape, but we are the ones who give meaning! We can never be replaced.*

Table 2. Illustrative examples of prompts and student reflections

To assess coding reliability, an independent double-coding procedure was conducted on a subset of 30 prompts (27% of the total corpus) using a coding rubric comprising six indicators. Inter-rater agreement was evaluated using the percentage of observed agreement, Cohen's

kappa coefficient (κ), and Gwet's AC1 coefficient. The results indicated substantial agreement for the *Contextual Completeness dimension* (κ = 0.673; AC1 = 0.713) and moderate agreement for *Professional Vocabulary* (κ = 0.549; AC1 = 0.534), Attention to Inclusion (κ = 0.490; AC1 = 0.511), and *Prompt Structure* (κ = 0.467; AC1 = 0.502). The Request Specificity dimension showed lower agreement (κ = 0.193; AC1 = 0.229), suggesting that this category involved a greater degree of interpretative judgment during coding. For the Iteration variable, Cohen's kappa could not be computed because both raters assigned all prompts to the same category (1 = single exchange). The complete results are presented in Table 3. The findings for the Iteration dimension indicate that participants still tend to conceive interaction with Artificial Intelligence as a one-time stimulus–response exchange rather than as a dialogic and iterative process of progressively refining prompts. This finding suggests that AI interaction competence—understood as the ability to exploit successive cycles of feedback and prompt revision to improve the quality of AI-generated outputs—has not yet been fully developed. Further research is therefore needed to examine the extent to which this competency evolves through experience and targeted educational interventions, as well as to identify the factors that promote a more informed and effective use of iterative interaction with AI systems.

Prompt Evaluation Grid Indicator	Observed agreement (%)	Cohen's k	Gwet's AC1	Interpretation
Contextual completeness	**80,0**	**0,673**	**0,713**	Substantial agreement
Specificity of the request	**46,7**	**0,193**	**0,229**	Weak agreement (fair)
Professional lexicon	**76,7**	**0,549**	**0,534**	Moderate agreement
Attention to inclusion	**66,7**	**0,490**	**0,511**	Moderate agreement
Prompt structure	**73,3**	**0,467**	**0,502**	Moderate agreement

Table 3. Inter-Rater Reliability of the Coding Rubric (N = 30)

With regard to the reliability of the coding of the 30 participants' critical reflections (representing 27% of the total corpus), an independent double-coding procedure was conducted. The following tables present the frequencies assigned by the two raters for each of the four levels of Kember et al.'s coding framework [18] (Table 4), the 4 × 4 agreement matrix between the two raters (Table 5), and the percentage of agreement for each level (Table 6). In addition, the overall inter-rater reliability for the four-level ordinal scale was assessed using the weighted Cohen's kappa coefficient (quadratic weighting).

Level of reflection	Evaluator 1	Evaluator 2
1. Descriptive writing	0	0
2. Understanding	2	0
3. Reflection	22	25
4. Critical reflection	6	5
Total	**30**	**30**

Table 4. Frequencies attributed by the two evaluators to the reflections

Evaluator 1/ Evaluator 2	1	2	3	4	Total
1	0	0	0	0	0
2	0	0	2	0	2
3	0	0	21	1	22
4	0	0	2	4	6
Total	**0**	**0**	**25**	**5**	**30**

Table 5. Concordance matrix regarding reflections

Level	Agreements	Cases involved	% agreement
1. Descriptive writing	0	0	—
2. Understanding	0	2	0,0%
3. Reflection	21	27	77,8%
4. Critical reflection	4	7	57,1%

Table 6. Percentage of agreement by level

Inter-rater reliability was estimated using the weighted Cohen's kappa coefficient (quadratic weighting), which is appropriate for four-level ordinal variables, and Gwet's AC1 coefficient. The weighted Cohen's kappa value was $\kappa = 0.571$, indicating a moderate level of agreement between the two raters [23]. In contrast, Gwet's AC1 = 0.798 indicated substantial inter-rater agreement according to the classification proposed by Landis and Koch.

Prompt variable	Spearman ρ	p	Kendall τ-b	p
Contextual completeness	**0,176**	**0,045**	**0,164**	**0,045**
Specificity of the request	0,157	0,074	0,149	0,075
Professional lexicon	**0,235**	**0,007**	**0,227**	**0,008**
Attention to inclusion	0,141	0,111	0,129	0,114
Prompt structure	0,085	0,334	0,084	0,332
Total prompt score	**0,264**	**0,002**	**0,225**	**0,003**

Table 7. Nonparametric correlations between prompt quality and level of critical reflection

Prompt variable	Spearman ρ	p	Holm-adjusted p	Kendall τ-b	Significant after Holm correction
Contextual completeness	0.176	0.045	0.180	0.164	No
Specificity of the request	0.157	0.074	0.222	0.149	No
Professional lexicon	0.235	0.007	0.035	0.227	Yes
Attention to inclusion	0.141	0.111	0.222	0.129	No
Prompt structure	0.085	0.334	0.334	0.084	No
Total prompt score	**0.264**	**0.002**	n.a.	0.225	Not included in the correction

Table 8. Nonparametric correlations between prompt quality and level of critical reflection, with Holm-adjusted p-values for the individual indicators

The analysis revealed a positive and statistically significant correlation between the overall prompt score and the level of critical reflection (Spearman's ρ = 0.264, p = 0.002; Kendall's τ-b = 0.225, p = 0.003). After the Holm–Bonferroni correction, Professional Lexicon was the only individual indicator significantly associated with reflection (adjusted p = 0.035). Contextual Completeness did not remain statistically significant (adjusted p = 0.180) and is therefore considered an exploratory indicator. The Kruskal–Wallis test further confirmed significant differences in the overall prompt score across the four levels of reflection ($H(3) = 9.52$, $p = 0.023$; $\varepsilon^2 = 0.052$) (Table 8). The analysis revealed a positive and statistically significant correlation

between the overall prompt score and the level of critical reflection (Spearman's ρ = 0.264, p = 0.002; Kendall's τ-b = 0.225, p = 0.003). Although the strength of the association was modest, students who produced higher-quality prompts tended to demonstrate deeper levels of critical reflection. Among the individual prompt indicators, Contextual Completeness showed a weak but statistically significant positive correlation with the level of critical reflection (ρ = 0.176, p = 0.045), while Professional Vocabulary was the indicator most strongly associated with reflection depth (ρ = 0.235, p = 0.007). In contrast, Request Specificity, Attention to Inclusion, and Prompt Structure did not reach statistical significance. The findings for the Request Specificity indicator suggest that its limited predictive value is consistent with the relatively low level of coding reliability observed for this dimension. The Kruskal–Wallis test further confirmed that the overall prompt score differed significantly across the four levels of critical reflection (H(3) = 9.52, p = 0.023). However, the effect size was small (ε^2 = 0.052), indicating that the level of critical reflection accounted for a limited, albeit non-negligible, proportion of the variance in prompt quality. Overall, these findings are consistent with the hypothesis that greater prompt design competence—particularly demonstrated through the use of professional terminology and more comprehensive contextualization—is associated with deeper critical reflection on the use of Artificial Intelligence in instructional design (Table 9).

Statistics	Value
H	9,518
gl	3
p	0,023
ε^2	0,052

Table 9. Kruskal–Wallis Test - Comparison of the total prompt score across the four levels of critical reflection

5. Conclusion and discussion

The present study examined the quality of prompts produced by students enrolled in the Primary Teacher Education degree program and the level of critical reflection they developed after using Artificial Intelligence (AI) to design a lesson. The aim was to explore the relationship between prompt engineering competencies and reflective capacities in educational contexts. The findings indicate, first, that most participants were able to construct prompts of intermediate quality, whereas only a limited number achieved the highest scores according to the evaluation rubric. This distribution suggests that, although students had acquired the fundamental skills required to interact with generative AI systems, the ability to formulate comprehensive, specific, and methodologically well-structured prompts remains a competency that is still developing. With regard to critical reflection, the reflection level, according to Kember et al.'s [18] classification, was the most frequently observed. This finding indicates that participants were generally able

to relate their laboratory experience to their emerging professional knowledge, recognising both the strengths and the limitations of Artificial Intelligence in instructional design. The critical reflection level was less frequent but represented a substantial proportion of the sample. This higher level of reflection involves a deeper reconsideration of one's understanding of the role of AI within the teaching profession. These findings suggest that the proposed learning activity went beyond promoting a purely instrumental use of technology, instead fostering professional reflection and encouraging the development of a critical stance toward Artificial Intelligence. The reliability analyses showed that the coding rubrics used to evaluate prompts and reflective writings achieved generally satisfactory levels of inter-rater agreement. In particular, the assessment of Contextual Completeness demonstrated substantial agreement, whereas the remaining dimensions achieved moderate levels of agreement. The lower reliability observed for Request Specificity suggests that this construct involves a greater degree of interpretative judgment and may benefit from more detailed coding criteria in future studies. Similarly, the coding of critical reflections yielded moderate inter-rater agreement, a result consistent with the interpretative complexity involved in classifying reflective texts. The most significant finding of the study concerns the relationship between prompt quality and the depth of critical reflection. The non-parametric analyses revealed a positive and statistically significant association between the overall prompt score and the level of critical reflection, although the magnitude of this relationship was modest. These findings suggest that students who are capable of designing more comprehensive and better-structured prompts also tend to produce more sophisticated reflections on the educational use of AI. In particular, the use of appropriate professional vocabulary and the ability to contextualise requests were the prompt characteristics most strongly associated with higher levels of critical reflection. These results were further supported by the Kruskal–Wallis test, which revealed significant differences in prompt quality across the four levels of critical reflection. Overall, the findings are consistent with the hypothesis that prompt engineering should not be regarded merely as a technical skill, but rather as an integral component of teachers' professional competence, closely connected to processes of pedagogical analysis, instructional design, and reflective practice. In this regard, future studies adopting an experimental or longitudinal design would be valuable—for example, by providing training in prompting skills and subsequently measuring its effects on reflective practice—to further strengthen the evidence base in this area. Constructing effective prompts requires the ability to articulate instructional objectives, contextualise requests, select appropriate teaching methodologies, and employ professional educational terminology—competencies that reflect cognitive processes analogous to those involved in educational planning and curriculum design. This study nevertheless has several limitations. First, the sample consisted of students enrolled in a single teacher education program who participated in a single laboratory experience, thereby limiting the generalizability of the findings. Second, although statistically significant, the observed correlations were relatively modest, indicating that prompt quality represents only one of several factors influencing the ability to develop critical reflections on the educational use of Artificial Intelligence. Finally, prompt quality was assessed using an ad hoc evaluation rubric, the validity of which should be further established through application to larger samples and diverse educational contexts. Another potential limitation concerns the use of non-uniform rating scales across the study. Specifically, four indicators were rated on a three-point scale (1–3), whereas the remaining indicators were rated on a two-point scale (1–2). This asymmetry may have resulted in an implicit weighting of certain elements over others, potentially influencing the overall interpretation of the findings. Furthermore, it should be noted that the "reflection" category is highly concentrated (with approximately 77% of the sample

falling within this category). This limited variability may have reduced the variance available for correlation analysis, making it more difficult for a genuine relationship to emerge as statistically significant. In light of these findings, future research should investigate the development of prompt engineering competencies through longitudinal studies, evaluate the effectiveness of educational interventions specifically designed to foster critical reflection on the use of AI, and examine the role of individual variables—such as prior digital competence, technological self-efficacy, and pedagogical beliefs—in mediating the relationship between prompt quality and reflective processes. From this perspective, integrating Artificial Intelligence into initial teacher education should be viewed not only as an opportunity to enhance instructional design but also as a valuable context for fostering the reflective competencies required for the critical, informed, and pedagogically grounded use of emerging technologies.

ISSN 2991-9371

References

[1] N. Selwyn, "Should robots replace teachers? AI and the future of education", Cambridge (UK), Polity Press, 2019.

[2] W. Holmes, M. Bialik and C. Fadel, "Artificial intelligence in education: Promises and implications for teaching and learning", Boston (MA), Center for Curriculum Redesign, 2019.

[3] O. Zawacki-Richter, V.I. Marín, M. Bond and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?", *International Journal of Educational Technology in Higher Education*, Volume: 16 Issue: 39, 2019.

[4] P. Mishra and M.J. Koehler, "Technological pedagogical content knowledge: A framework for teacher knowledge", *Teachers College Record*, Volume: 108 Issue: 6, pg.1017–1054, 2006.

[5] D.T.K. Ng, J.K.L. Leung, S.K.W. Chu and M.S. Qiao, "Conceptualizing AI literacy: An exploratory review", *Computers and Education: Artificial Intelligence*, Volume: 2, 100041, 2021.

[6] I. Celik, "Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education", *Computers in Human Behavior*, Volume: 138, 107468, 2023.

[7] R. Vuorikari, S. Kluzer and Y. Punie, "DigComp 2.2: The Digital Competence Framework for Citizens", Luxembourg, Publications Office of the European Union, 2022.

[8] C. Redecker, "European framework for the digital competence of educators: DigCompEdu", Luxembourg, Publications Office of the European Union, 2017.

[9] I. Tenberga and L. Daniela, "Artificial intelligence literacy competencies for teachers through self-assessment tools", *Sustainability*, Volume: 16 Issue: 23, 10386, 2024.

[10] M. Kasinidou, S. Kleanthoys and J. Otterbacher, "Cypriot teachers' digital skills and attitudes towards AI", *Discover Education*, Volume: 4 Issue: 1, 1, 2025.

[11] W. Flores and A. Chiappe, "Integrating AI into Education: Preparation Factors and Teachers' Digital Competencies", *Revista colombiana de educación*, Volume: 97, e20825-e20825, 2025.

[12] J. D. Zamfirescu-Pereira, R.Y. Wong, B. Hartmann and Q. Yang, "Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts", *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, pg. 1–21, 2023.

[13] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, D.C. Schmidt, "A prompt pattern catalog to enhance prompt engineering with ChatGPT", *arXiv*, 2023.

[14] Ministero dell'Istruzione e del Merito, "Le Linee Guida per l'adozione dell'Intelligenza Artificiale nella scuola", 2025,

[15] https://www.mim.gov.it/documents/20182/0/MIM_Linee+guida+IA+nella+Scuola_09_08_2025-signed.pdf/b70fdc 45-4b75-1f7e-73bf-eab12989b928?t=1756468797694.

[16] Unione Europea, "REGOLAMENTO (UE) 2024/1689 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)", 2024, ELI: http://data.europa.eu/eli/reg/2024/1689/oj.

[17] K. Krippendorff, "Content analysis: An introduction to its methodology (4th ed.)", Thousand Oaks (CA), Sage, 2018.

[18] M. Sandelowski, C.I. Voils and G. Knafl, "On quantitizing", *Journal of Mixed Methods Research*, Volume: 3 Issue: 3, pg. 208–222, 2009.

[19] D. Kember, J. McKay, K. Sinclair and F.K.Y. Wong, "A four-category scheme for coding and assessing the level of reflection in written work", *Assessment & Evaluation in Higher Education*, Volume: 33 Issue: 4, pg. 369–379, 2008.

[20] N. Hatton and D. Smith, "Reflection in teacher education: Towards definition and implementation", *Teaching and Teacher Education,* Volume: 11 Issue: 1, pg. 33–49, 1995.

[21] P.A. Facione, "Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction", Millbrae (CA), California Academic Press, 1990.

[22] J. Cohen "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, Volume: 20 Issue: 1, pg. 37–46, 1960.

[23] K. Gwet "Handbook of Inter-Rater Reliability: How to Estimate the Level of Agreement Between Two or Multiple Raters". Gaithersburg (MD), STATAXIS Publishing Company, 2001.

[24] J. R. Landis and G.G Koch "The measurement of observer agreement for categorical data", *Biometrics*, Volume: 33 Issue: 1, pg. 159–174, 1977.

THE USE OF AI TOOLS

The authors used ChatGPT (OpenAI) and Claude (Anthropic) for language editing and translation of selected parts of the manuscript. All content was reviewed and verified by the authors.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Maria Rosaria Re wrote § 1.2, 2 and 3; Stefania Capogna wrote § 1 and 1.3; Alessandra Natalini wrote § 1.1, 4 and 5. All authors read and approved the final manuscript.

From Usage Intensity to Relationship Quality: A Typology of Student-AI Use and Design Propositions for Generative AI in Higher Education

Riem Yasin[1], Samir Yasin[2] and Carsten Lanquillon[3]

[1]Heilbronn University of Applied Sciences, Germany, riem.yasin@hs-heilbronn.de
[2]Heilbronn University of Applied Sciences, Germany, sayasin@stud.hs-heilbronn.de
[3]Heilbronn University of Applied Sciences, Germany, carsten.lanquillon@hs-heilbronn.de

Abstract

Competence frameworks for generative AI in higher education specify what students should be able to do, yet they leave the everyday interaction pattern that produces these abilities under-specified. This paper proposes the quality of the student-AI relationship as this missing design object. Licklider's vocabulary of symbiosis informs this typology, although it replaces his ideal of tight coupling with preserved independent capability. Two states result. Capability-returning use strengthens the student's own capability through delegation; dependent use leaves students unable to produce the delegated work alone. We hypothesise that the interface layer shapes which relationship forms. Three observable markers make the two states assessable, namely verification of AI output, task-scoped delegation, and dependence awareness. Descriptive evidence from two open survey datasets illustrates the markers. Among 218 high-frequency users in a single-institution survey, the markers separate four response combinations (29, 33, 17, and 20 %). In a cleaned global dataset, ethical concern is associated with somewhat lower intensive use, yet intensive use remains common under high concern. The contribution consolidates overreliance, assessment-design, and self-regulation findings into a relationship-quality typology and five design propositions, with functional legibility as the novel proposition.

Keywords: *AI in education; human-AI interaction; AI literacy; cognitive offloading, interaction design; student-AI relationship*

1. Introduction

Survey evidence has made ChatGPT use in higher education visible at scale. An early global study reported a sample of 23,218 students from 109 countries and territories and identified brainstorming, summarising texts, and finding research articles as common academic uses [1]. The 2024-2025 dataset used in this paper contains 22,963 responses from 120 countries and territories and records use, perceived capabilities, ethical concerns, satisfaction, study outcomes, and skills development [2]. Teachers also report that generative AI (GenAI) requires changes in teaching and assessment, including stronger attention to AI understanding, critical thinking, ethical judgement, creativity, collaboration, and learning processes [3]. These findings establish the scale of use and the pressure on higher education. They leave open which quality of student-AI interaction learning environments should cultivate.

Usage frequency provides a weak measure of student use of GenAI when it treats different learning interactions as equivalent. Two students may use the same system equally often while delegating different parts of the learning process. Some students compare an AI-generated explanation with course material, identify gaps, and revise their own draft. Others submit a model-generated answer without inspecting its reasoning or evidence. A frequency-based measure counts both cases as equal use, although they allocate cognitive work differently. The relevant object is therefore relationship quality, understood as the pattern through which students allocate, monitor, and reclaim cognitive work in interaction with AI. Interfaces, defaults, prompts, checkpoints, and course rules shape this pattern. These elements raise or lower the effort of verification, normalise particular forms of delegation, and decide when human judgement re-enters the task.

Licklider's account of man-computer symbiosis gives this shift its theoretical starting point. He framed interactive computing as a close coupling in which humans set goals, formulate hypotheses, determine criteria, and evaluate outcomes while computers perform routinisable work that prepares the way for insight and decision [4]. For GenAI in higher education, this framing turns delegation into the central design question. Risko and Gilbert define cognitive offloading as external action that changes a task's information-processing requirements and lowers cognitive demand. Their analysis also links offloading decisions to metacognitive judgements about one's own ability [5]. In AI-supported studying, delegation becomes risky when students let the system complete whole performances or accept its output without verification. This risk appears in adjacent evidence. Zhai et al. review the risks of over-reliance on AI dialogue systems for decision-making, analytical reasoning, and critical thinking [6]. Gerlich links frequent AI-tool use to lower critical-thinking scores and identifies cognitive offloading as the mediating factor [7]. The central design issue is therefore how delegation is organised, so that AI-supported work stays checked, bounded, and connected to the learner's own capability.

This paper develops a relationship-quality typology for that concern. The typology distinguishes capability-returning use, in which AI-supported delegation strengthens the student's own capability, from dependent use, in which students can no longer produce the delegated performance alone. Between these states sits the interface layer, understood as the visible and institutional arrangement of prompts, labels, defaults, verification steps, and assignment

rules. This layer links the typology to two research strands. Human-centred AI pursues high automation and high human control as a joint design goal [8], while hybrid intelligence treats human-AI work as a division of labour that draws on complementary strengths [9].

The contribution is conceptual and threefold. It consolidates prior work on overreliance, control allocation, and AI assessment into a relationship-quality typology for student-AI interaction in higher education. The typology defines capability-returning use, dependent use, and the mediating interface layer (Section 3), and makes these concepts assessable through three observable markers, namely verification of output, task-scoped delegation, and dependence awareness (Section 3.3). It then states five design propositions for AI-supported learning environments, with functional legibility marking the proposition least covered by prior assessment and overreliance work (Section 5). A descriptive illustration from two open survey datasets [2], [10] examines whether the named marker combinations appear in student populations and how ethical concern relates to frequent use (Section 4).

2. Background

Large-scale survey work defines the adoption problem. Ravšelj et al. analyse higher-education students' early perceptions of ChatGPT and report task-specific uses such as brainstorming, summarising, and academic writing [1]. The 2024-2025 follow-up dataset records usage, capabilities, regulation and ethical concerns, satisfaction, study outcomes, and skills development for 22,963 responses from 120 countries and territories [2]. Bower et al. add the teaching perspective by showing that teachers largely expected generative AI to require changes in teaching, learning outcomes, and assessment, while their proposed assessment responses varied widely [3]. Holmes et al. sharpen the design problem [11]. The authors characterise much AI in education (AIED) as reflecting or automating existing educational assumptions and practices, whereas stronger uses would open possibilities that challenge established pedagogies or help teachers work more effectively. Together, these sources establish scale, institutional pressure, and the need for design direction. They do not yet specify which qualities of student-AI interaction learning environments should cultivate.

Although competence frameworks clarify what students should be able to do with AI, they leave the everyday interaction pattern that produces these abilities under-specified. AI literacy is framed as the capacity to evaluate AI critically, communicate and collaborate with it effectively, and use it as a tool [12]. Review work organises this capacity into knowing and understanding AI, using and applying AI, evaluating and creating AI, and addressing ethical issues [13]. UNESCO turns the same competence target into twelve student competency blocks [14]. These blocks span a human-centred mindset, the ethics of AI, AI techniques and applications, and AI system design, and they advance from Understand to Apply and Create. In higher-education terms, Ehlers frames the same demand as self-organised action in open, uncertain, and digitally shaped problem situations [15]. These accounts clarify what students should become capable of. The unresolved design question is how repeated student-AI interaction must be structured so that delegation supports that capability instead of replacing the work through which it develops.

The mechanism linking delegation to learning is cognitive offloading. External action can reduce a task's information-processing demands and improve performance when internal capacity is limited, yet offloading also changes which parts of the task remain cognitively active for the learner [5]. Task demand and metacognitive self-judgements together shape offloading [5]. The same analysis names memory costs, automation complacency, automation bias, and possible skill decay as risks of technology-based offloading. In AI-supported education, this mechanism becomes critical when students accept AI-generated recommendations without questioning them. Zhai et al. define this pattern as over-reliance and review evidence linking it to risks for decision-making, critical thinking, and analytical reasoning in educational and research contexts [6]. Gerlich's mixed-method survey reports the same concern empirically, since frequent AI-tool use is negatively associated with critical-thinking scores, with cognitive offloading mediating the association [7].

Design follows from this diagnosis. Shneiderman's HCAI framework separates automation from human control and treats their combination as a design target, so that systems raise performance while preserving self-efficacy, mastery, creativity, and responsibility [8]. Dellermann et al. sharpen this into a relationship claim. Hybrid intelligence allocates work between human and artificial agents according to complementary strengths, so that the joint system achieves results neither agent could reach alone and improves through mutual learning [9]. Translated into education, this principle becomes a question of learning activity.

Student-AI collaboration is shaped by curriculum design, patterns of interaction, learning environments, and staged development [16]. In academic writing, iterative interaction with a GenAI assistant was associated with higher writing performance than linear use that treated AI primarily as a supplementary information source [17]. Together, these findings direct interfaces and courses to structure delegation through verification, explicit task boundaries, and dependence awareness.

Recent design and assessment work already addresses parts of this problem. In AI-assisted decision tasks, cognitive forcing functions reduced overreliance more effectively than simple explanations, although participants rated the stronger interventions less favourably [18]. In a randomised writing-task study, ChatGPT support raised essay-score gains, whereas knowledge gain and transfer showed no significant difference [19]. Performance gains thus co-exist with altered self-regulated learning and technology dependence. Molenaar frames hybrid human-AI learning technologies as a control-allocation problem, where AI systems detect, diagnose, and act, while levels of automation specify how control shifts between learner, teacher, and technology [20]. Assessment design has moved in the same direction. The AI Assessment Scale specifies graded levels of permitted GenAI use, from no AI to full AI, including a level where AI completes parts of a task and students evaluate the output [21]. Two-lane assessment guidance separates secured assurance of learning from open AI-integrated work [22], while Curtis argues that an all-or-none settlement misses the need for constrained middle-lane uses of GenAI [23]. Outside education, Dell'Acqua et al. show why task scope matters, as consultants using GPT-4 performed better on tasks inside the AI frontier yet worse on a task outside it [24]. Taken together, these studies narrow the contribution of the present paper. The paper does not introduce verification, task-specific AI permissions, or control allocation as isolated ideas. It connects them through learner-side markers and the interface layer, and it adds functional legibility as the missing link between system operation and student judgement.

3. A Relationship-Quality Typology

The typology rests on two states, a mediating interface layer, and three observable markers (Figure 1).



Figure 1. The relationship-quality typology, showing the target state, the failure state, and the interface layer as design surface, each with observable markers.

3.1 Capability-returning use: the target state

Following Licklider, we use symbiosis as a historical starting point for thinking about close human-computer coupling [4]. The educational target, however, differs from Licklider's tight-coupling ideal, since student-AI interaction should preserve and strengthen the student's ability to act without the system. We therefore define the target state as capability-returning use. Students who use an AI system to explain a concept, reconstruct the explanation in their own words, and check it against the course material end the interaction more capable than before. The defining criterion is the return, whereby the interaction produces an output and a gain on the human side. This criterion shifts the question from whether AI was used to whether the use returned capability to the student. The three markers introduced in Section 3.3 record the interaction pattern that plausibly precedes this return, and they do not measure the return itself. Section 6 treats this gap as the first task for empirical work.

3.2 Dependent use: the failure state

The failure state is the inverse, namely delegation without return. Risko and Gilbert describe the mechanism as cognitive offloading, in which externalising parts of a task can reduce the processing through which the corresponding skill is practised or maintained [5]. Applied to studying, a systematic review associates over-reliance on AI dialogue systems with weaker cognitive abilities [6], and survey evidence associates heavy AI-tool use with lower critical-thinking scores [7]. We call the resulting state dependence, since the student can no longer produce the delegated performance alone. Two properties make dependence a design problem. It remains silent, as the missing capability surfaces only when the system is unavailable, typically in an examination or a workplace. It plausibly builds up through many individually reasonable delegations, and the cross-sectional evidence cited above supports associations rather than trajectories, so the longitudinal designs named in Section 6 can test the temporal reading. Both properties call for continuous countermeasures at the interface and course level.

3.3 The interface layer and three observable markers

Between the two states sits the layer that shapes which of them forms, namely prompts, defaults, labels, checkpoints, and course rules. A frictionless copy-paste path makes wholesale delegation the default, while a visible verification step makes the capability-returning move the default. This is the paper's central design hypothesis. It suggests that relationship quality is co-produced by the student together with the surrounding interface and course design, which makes it a target for intervention and empirical testing. The same position underlies human-centred AI design [8] and hybrid-intelligence research [9], and the typology applies it to the concrete surfaces of AI-supported studying.

For design and for assessment, the typology needs markers that are observable without psychometric instruments. Three markers make the two states observable in interaction.

1. Verification. Do students check AI output against other sources or their own understanding before using it? Verification is the behavioural core of the critical-evaluation competence that AI-literacy frameworks demand [12], [14].

2. Task-scoped delegation. Is the AI tied to a defined role in a concrete task, such as summarise this text or critique this draft, or does it produce the performance wholesale? Scoped delegation preserves the return that defines capability-returning use, while wholesale delegation removes it.

3. Dependence awareness. Do students report being able to produce the performance without the system? The marker relies on self-assessment and serves as a risk indicator, since reported dependence signals that the delegated capability may no longer be available unaided. Self-assessments can misjudge actual capability, so Section 6 names calibration against demonstrated capability as a validation step.

Table 1 connects the three markers, together with two design conditions imported from transparency and governance work, to the design propositions of Section 5.

Marker or design condition	Type	Observable question	Design proposition
Verification	Marker	Do students check AI output before using it?	P1 verification at the point of output
Task-scoped delegation	Marker	Is the AI's role in the task defined and bounded?	P3 delegation contracts per assignment
Dependence awareness	Marker	Do students report being able to produce the work without AI?	P4 without-AI checkpoints and usage mirrors
Functional legibility	Design condition	Can students tell which component produced an output?	P2 labels and attribution at the point of input
Social anchoring	Design condition	Did the cohort commit to shared rules it helped write?	P5 co-created AI charters

Table 1. Relationship-quality markers, design conditions, and corresponding design propositions.

4. Descriptive Illustration from Two Open Datasets

The descriptive step asks whether the relationship states named by the typology appear in student survey data. We use the two openly licensed self-report datasets for a bounded descriptive purpose, namely to identify profile heterogeneity and to compare use patterns across item-based groups.

Dataset A is a single-institution survey on study habits and AI use, collected with proportional allocation across 14 undergraduate programs and quota-based fieldwork within programs (n = 357). The public package includes the raw Spanish file, a cleaned English analytical file, a codebook, and the questionnaire [10]. Dataset B is a global survey of higher-education students' perceptions of ChatGPT, collected between October 2024 and February 2025 (22,963 responses from 120 countries and territories, CC BY 4.0) [2]. For Dataset B, we removed 127 exact duplicate rows and 205 further rows that repeated the substantive item blocks of an earlier response. After this cleaning, 19,382 respondents provide a valid answer on general ChatGPT use, and of these 17,178 also provide complete ethical-concern and satisfaction scores. The comparison below assumes that this item-level missingness is unrelated to the pattern of interest. Both datasets rely on self-report, and Dataset A contains a large share of frequent AI users.

Dataset A supplies the profile check. Of 357 students, 218 (61 %) report using generative AI for study purposes almost always or always. Within this high-use band, we coded two markers at the two most-affirming answers on their published scales. Verification counts when students report checking AI output almost always or always; perceived dependence counts when they answer probably yes or a lot to the dependence item. Dataset A contains no item that captures task-scoped delegation, so the illustration covers two of the three markers. Cross-classifying two binary markers partitions the 218 students into four cells by construction, and the informative result is the cell sizes. 63 students (29 %) verify and report no dependence, the combination closest to capability-returning use in these data. 73 (33 %) verify and simultaneously report dependence, a combination in which checking coexists with a dependence signal and which the two-state typology does not predict. 38 (17 %) report dependence without verification, the combination closest to the failure state. 44 (20 %) report neither marker (Figure 2). All four cells are substantial, including the two combinations outside the named states. Usage frequency alone collapses these differences into one high-use category, while the two markers separate four distinct marker-response combinations.

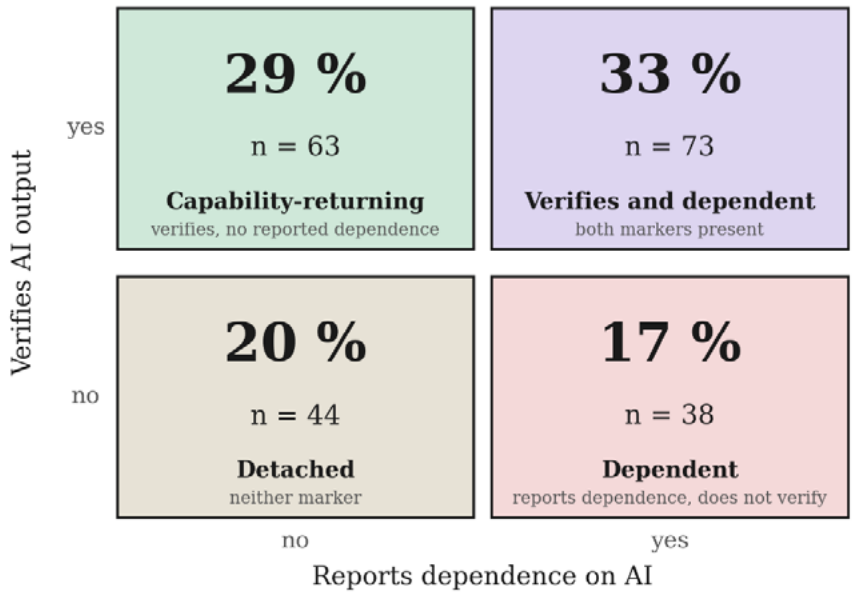

Figure 2. Four marker-response combinations among the 218 high-frequency users in Dataset A, cross-classified by the verification and dependence markers.

Dataset B adds the concern-use check. This comparison uses the 17,178 respondents with valid general-use, ethical-concern, and satisfaction scores after duplicate removal. We split respondents into thirds by their average satisfaction score and by their average ethical-concern score. Among highly satisfied students, the share of intensive users, defined as the two highest categories on the general-use item, is 48 % when ethical concern is low and 38 % when ethical concern is high. Within this subgroup, higher ethical concern accompanies somewhat lower intensive use, although intensive use remains common even under strong concern.

The same dataset locates everyday ChatGPT use in core academic tasks. The highest Q18 item means on the five-point frequency scale are summarising texts (3.0), brainstorming (3.0), academic writing (2.9), research assistance (2.9), and study assistance (2.8). These item means do not measure learning quality; instead, they show where design interventions would meet actual use, namely in recurring academic work rather than in peripheral experimentation.

Together, the two datasets support the descriptive premise of the typology. Dataset A shows that students with the same high usage frequency separate into different relationship profiles. Awareness of risks therefore co-exists with sustained use. These counts give the propositions an empirical starting point. Future work should develop the single-item markers into validated constructs and test the propositions in experimental, course-based, and longitudinal designs.

5. Design Propositions

The propositions make the three markers designable. They specify where an AI-supported course or interface must intervene if it aims to move student AI use from mere frequency toward relationship quality.

P1. Verification at the point of output. Interfaces should make checking the next action after generation, whether through source comparison, claim-basis inspection, or a required record of one verification step per AI-assisted task. AI-literacy frameworks require critical evaluation of AI output [12], [14], and this proposition locates that evaluation at the moment when students can still revise their judgement. Evidence from AI-assisted decision-making supports this timing, since cognitive forcing functions that interrupt default acceptance reduce overreliance more effectively than simple explanations, although they can reduce perceived usability [18]. For learning environments, verification therefore needs to appear at the point of reuse, not only as a general competence outcome.

P2. Functional legibility. Learning environments that combine several AI functions should identify, at the point of input and output, which component processes the input, what operation it performs, and what kind of output it returns. Students can evaluate an output only if the interface makes its operation legible, whether it transcribed, retrieved, generated, summarised, or recommended the answer. This links the proposition directly to interpretive affordances in AI literacy [12], to UNESCO's requirements for transparent, explainable, human-controlled AI use [14], and to Shneiderman's account of comprehensible, predictable, and controllable automation [8].

P3. Task-scoped delegation contracts. Assignments should define the AI role per task, meaning which steps may be delegated, which steps must remain the students' own, and how the work returns to the student. Assessment-scale work already defines graded permission levels for GenAI use, including levels where AI supports ideation, editing, task completion, or full task work [21]. The delegation contract proposed here uses that permission logic for a narrower

design purpose, since it ties each permitted AI role to the learner-side return that must follow. For example, the AI may draft counter-arguments, whereas students must select, revise, and defend the final position. This gives assessment a concrete object, namely the declared division of labour and the students' documented hand-back, rather than an unverifiable suspicion about hidden AI use.

P4. Dependence-awareness feedback. Courses and systems should mirror delegation back to the learner through periodic without-AI checkpoints or usage summaries that show which task types students have not completed unassisted. Two-lane assessment guidance protects assurance of learning through secured AI-free tasks while allowing AI-integrated work elsewhere [22]. Curtis's critique of all-or-none assessment shows why many tasks require constrained AI use rather than total prohibition or unrestricted use [23]. Without-AI checkpoints translate that assessment logic into feedback for the learner, since they reveal which performances still work without the system while there is time to adjust practice. This proposition also follows from human-centred AI competence, which treats human agency and avoidance of AI dependence as part of responsible use [14], and from Shneiderman's warning that automation can undermine skills needed when automation fails [8].

P5. Co-created AI-use rules. Institutions should combine clear institutional boundaries with cohort-level AI-use charters. Students and teachers should specify permitted AI roles task by task, record the resulting rules, and revisit them when assignments change. As Section 4 shows, awareness of risks co-exists with sustained use, and a charter addresses this co-existence by making students practise the judgement that AI competence requires [14] and the self-organised action that Ehlers places at the centre of future capability in emergent contexts [15].

6. Limitations and Future Work

The argument rests on a conceptual typology and a descriptive illustration. Licklider grounds the symbiosis vocabulary in close human-computer coupling, where humans set goals, define criteria, and evaluate results while computers take over routinisable work [4]. This paper transfers that vocabulary to student-AI relationships and uses it as design terminology.

The empirical illustration reports counts, shares, and item means from self-report surveys. Its evidential role is to make relationship combinations visible, not to estimate effects. The thresholds for the profile markers use the two most-affirming response options on each published scale, and independent studies should test whether these settings remain stable. Dataset A covers one institution and contains many frequent users. Dataset B captures ChatGPT use and therefore represents one prominent system rather than the full range of generative AI tools. The markers also remain single survey items and require validation as constructs. Above all, the markers capture the interaction pattern, not the capability return that defines the target state. Whether bounded, checked delegation predicts measured capability gains requires pre and post capability tests, and this comparison is the first task for future work.

Future work follows directly from these boundaries. Verification, task-scoped delegation, and dependence awareness should be developed into short validated measures and compared with behavioural traces. Interface studies can test verification affordances, functional labels, and delegation feedback as controlled variants. Course-level studies should examine delegation contracts and co-created AI-use rules inside real assessment practices, where these propositions gain practical meaning. Longitudinal designs can then test whether early relationship combinations predict later competence development.

7. Conclusion

GenAI use in higher education is usually discussed through quantity, meaning how many students use it, how often, and for which tasks. This paper shifted the unit of analysis to relationship quality. The proposed typology distinguishes forms of use in which AI returns capability to the student from forms in which students report that they can no longer produce the delegated performance alone. It also locates interface and course design as part of the relationship rather than as neutral surroundings. The descriptive survey illustration gives the typology empirical traction, since students with similar usage intensity fall into different relationship combinations, and ethical concern can co-exist with intensive use. The five propositions translate this argument into design decisions for AI-supported learning environments. The resulting task is concrete, namely to design student-AI interaction so that repeated use strengthens the student's own capacity to judge, act, and learn.

8. Acknowledgements

This research was supported by the EUonAIR project, no. 101177370, ERASMUS-EDU-2024-EUR-UNIV-1, within the framework of the EUonAIR Centre of Excellence in Responsible AI in Education (CORE.AI), co-funded by the European Union.

9. Ethics and data availability

This paper reports a secondary analysis of two openly licensed, anonymised survey datasets [2], [10], and the authors collected no new data from human participants. Both datasets are available from Mendeley Data, doi 10.17632/mcwb2ppdsw.1 (Dataset A) and doi 10.17632/nv2343nwsb.1 (Dataset B, CC BY 4.0). A Python script that reproduces every reported figure is available from the authors upon request.

References

[1] D. Ravšelj et al., "Higher education students' perceptions of ChatGPT: A global study of early reactions", *PLOS ONE*, Volume 20, Issue 2, e0315011, 2025.

[2] A. Aristovnik et al., "Higher Education Students' Evolving Perceptions of ChatGPT: Global Survey Data from the Academic Year 2024-2025", *Mendeley Data*, V1, 2025.

[3] M. Bower, J. Torrington, J. W. M. Lai, P. Petocz and M. Alfano, "How should we change teaching and assessment in response to increasingly powerful generative Artificial Intelligence? Outcomes of the ChatGPT teacher survey", *Education and Information Technologies*, Volume 29, pg. 15403-15439, 2024.

[4] J. C. R. Licklider, "Man-Computer Symbiosis", *IRE Transactions on Human Factors in Electronics*, Volume HFE-1, Issue 1, pg. 4-11, 1960.

[5] E. F. Risko and S. J. Gilbert, "Cognitive Offloading", *Trends in Cognitive Sciences*, Volume 20, Issue 9, pg. 676-688, 2016.

[6] C. Zhai, S. Wibowo and L. D. Li, "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review", *Smart Learning Environments*, Volume 11, Article 28, 2024.

[7] M. Gerlich, "AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking", *Societies*, Volume 15, Issue 1, Article 6, 2025.

[8] B. Shneiderman, "Human-Centered Artificial Intelligence: Reliable, Safe and Trustworthy", *International Journal of Human-Computer Interaction*, Volume 36, Issue 6, pg. 495-504, 2020.

[9] D. Dellermann, P. Ebel, M. Söllner and J. M. Leimeister, "Hybrid Intelligence", *Business & Information Systems Engineering*, Volume 61, pg. 637-643, 2019.

[10] C. D. Correa Alvarez, "Study habits and artificial intelligence use among university students: A proportionally stratified survey dataset", *Mendeley Data*, V1, 2026.

[11] W. Holmes, M. Bialik and C. Fadel, "Artificial Intelligence in Education", *Globethics Publications*, excerpt, orig. Center for Curriculum Redesign, Boston, 2019, 2023.

[12] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations", *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pg. 1-16, 2020.

[13] D. T. K. Ng, J. K. L. Leung, S. K. W. Chu and M. S. Qiao, "Conceptualizing AI literacy: An exploratory review", *Computers and Education: Artificial Intelligence*, Volume 2, Article 100041, 2021.

[14] UNESCO, *AI competency framework for students*, UNESCO Publishing, Paris, 2024.

[15] U.-D. Ehlers, *Future Skills: Lernen der Zukunft - Hochschule der Zukunft*, Springer VS, Wiesbaden, 2020.

[16] J. Kim, H. Lee and Y. H. Cho, "Learning design to support student-AI collaboration: perspectives of leading teachers for AI in education", *Education and Information Technologies*, Volume 27, pg. 6069-6104, 2022.

[17] A. Nguyen, Y. Hong, B. Dang and X. Huang, "Human-AI collaboration patterns in AI-assisted academic writing", *Studies in Higher Education*, Volume 49, Issue 5, pg. 847-864, 2024.

[18] Z. Buçinca, M. B. Malaya and K. Z. Gajos, "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making", *Proceedings of the ACM on Human-Computer Interaction*, Volume 5, Issue CSCW1, Article 188, 2021.

[19] Y. Fan, L. Tang, H. Le, K. Shen, S. Tan, Y. Zhao, Y. Shen, X. Li and D. Gašević, "Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance", *British Journal of Educational Technology*, Volume 56, Issue 2, pg. 489-530, 2025.

[20] I. Molenaar, "Towards hybrid human-AI learning technologies", *European Journal of Education*, Volume 57, Issue 4, pg. 632-645, 2022.

[21] M. Perkins, L. Furze, J. Roe and J. MacVaugh, "The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment", *Journal of University Teaching & Learning Practice*, Volume 21, Issue 6, 2024.

[22] D. Liu and A. Bridgeman, "What to do about assessments if we can't out-design or out-run AI?", *Teaching@Sydney*, 12 July 2023, available at: https://educational-innovation.sydney.edu.au/teaching@sydney/what-to-do-about-assessments-if-we-cant-out-design-or-out-run-ai/

[23] G. J. Curtis, "The two-lane road to hell is paved with good intentions: why an all-or-none approach to generative AI, integrity, and assessment is insupportable", *Higher Education Research & Development*, Volume 44, Issue 8, pg. 2151-2158, 2025.

[24] F. Dell'Acqua et al., "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality", *Organization Science*, Articles in Advance, 2026.

THE USE OF AI TOOLS

The authors used ChatGPT (OpenAI) and Claude (Anthropic) for language editing and translation of selected parts of the manuscript. All content was reviewed and verified by the authors.

Standards for Machine-Readable Course Descriptions

Yannik Langer[1], Carsten Lanquillon[2], Seline Löwe[3], Tom Ochs[4], Elias Ottenbacher[5] and Joshua Zürn[6]

[1]Hochschule Heilbronn, Heilbronn, Germany, yannik.langer@hs-heilbronn.de
[2]Hochschule Heilbronn, Heilbronn, Germany, carsten.lanquillon@hs-heilbronn.de
[3]Hochschule Heilbronn, Heilbronn, Germany, seline.loewe@hs-heilbronn.de
[4]Hochschule Heilbronn, Heilbronn, Germany, toochs@stud.hs-heilbronn.de
[5]Hochschule Heilbronn, Heilbronn, Germany, eottenbach@stud.hs-heilbronn.de
[6]Hochschule Heilbronn, Heilbronn, Germany, jozuern@stud.hs-heilbronn.de

Abstract

Module handbooks at German higher education institutions are predominantly maintained as unstructured PDF or Word documents. The absence of a common, machine-readable format blocks reliable processing by AI-based services and turns cross-institutional course comparison – central to exchange programmes such as EUonAIR – into a manual and error-prone task. A systematic literature review shows that no established standard (IEEE LOM, ECTS, ELMO, schema.org/Course) combines pedagogical depth with machine-readability: cognitive-level coding via Bloom's Taxonomy, Constructive Alignment and NLP processability remain unaddressed. Following the Design Science Research methodology, this study develops a best-practice standard for course descriptions that is both human- and machine-readable (PDF and JSON-LD) and encodes learning outcomes by Bloom level. The artifact is evaluated through expert interviews (Mayring qualitative content analysis, criteria K1–K8) and an exploratory controlled experiment in which two large language models process the same module in free-text and in structured form. Structuring raises extraction F1 from 74 % to 94 %, Bloom-classification accuracy from 64 % to 86 % and lowers the hallucination rate from 27 % to 7 %, at the cost of roughly double the input tokens. Because the experiment covers a single module with one run per model–format–task combination, and because the Bloom level is stated explicitly in the structured input but only implicitly in the free text, these figures should be read as the direction and consistency of an effect rather than as precise performance estimates. The results confirm the value of structured, pedagogically rich course descriptions while identifying authoring effort as the main adoption barrier.

Keywords: *Machine-readable course descriptions; Design Science Research; Bloom's Taxonomy; Constructive Alignment; JSON-LD;*

1. Introduction

This section motivates the project. It describes how course descriptions are currently handled at higher education institutions, derives the research problem and questions, and outlines the structure of the paper.

1.1 Initial Situation

Module handbooks are the central document through which higher education institutions (HEIs) communicate the content, learning outcomes, workload and assessment of their courses. At German HEIs these handbooks are almost exclusively produced and stored as unstructured PDF or Word documents, and every faculty applies its own layout and terminology. No common, machine-readable format exists across institutions [1], [2].

This situation has become a bottleneck for two developments that increasingly rely on structured course data. First, AI-based student services – advising chatbots, recommender systems and automatic question generation – must extract structured information from these documents; when applied to free text, large language models (LLMs) show low extraction quality and a high tendency to hallucinate [2], [3]. Second, international exchange programmes such as EUonAIR depend on comparing courses across partner institutions. Today this comparison and the resulting credit recognition are performed manually, because the underlying descriptions are neither structurally consistent nor interoperable [4], [5].

1.2 Problem Statement and Motivation

Several metadata standards for educational resources already exist (IEEE LOM, ECTS Users' Guide, ELMO, schema.org/Course). However, a systematic review of the literature reveals a structural gap: none of these standards combines pedagogical depth with machine-readability. In particular, the cognitive level of learning outcomes (Bloom's Taxonomy), the coupling of outcomes, teaching methods and assessment (Constructive Alignment, Biggs 1996 [8]), and the suitability of descriptions for natural-language processing are not jointly operationalised in any single standard [1], [6], [7]. As a consequence, learning outcomes remain semantically "empty" for machines, and the comparability promised by frameworks such as ECTS is not realised in practice.

This project addresses the gap by designing a best-practice standard for course descriptions that is simultaneously human-readable and machine-readable and that explicitly encodes pedagogical information. The work is guided by the following research questions:

- RQ1 – Which mandatory fields must a course description contain in order to be both pedagogically meaningful and reliably machine-processable?

- RQ2 – How can pedagogical concepts – Bloom's Taxonomy (C1) and Constructive Alignment (C2) – be operationalised within a machine-readable standard?
- RQ3 – Does structuring a course description according to the proposed standard measurably improve LLM-based extraction and Bloom classification and reduce hallucination compared with free text (B3)?
- RQ4 – How do practitioners assess the practicality and adoptability of the standard relative to existing standards?

The overarching goal is therefore a validated, dual-readable standard that closes the three identified research gaps C1, C2 and B3 and offers HEIs and exchange networks a concrete orientation for producing comparable, AI-ready course descriptions.

1.3 Structure

Section 2 establishes the conceptual foundations: the role of module handbooks, the relevant metadata standards, and the pedagogical concepts of Bloom's Taxonomy and Constructive Alignment, concluding with the related-work gap. Section 3 describes the methodology, which follows the Design Science Research (DSR) process and combines a systematic literature review, expert interviews analysed via qualitative content analysis, and an artifact evaluation. Section 4 presents the results: the designed artifact, the expert evaluation along criteria K1–K8, and the controlled AI experiment. Section 5 synthesises the findings, positions them against the literature, derives recommendations for action and discusses limitations. Section 6 concludes and outlines future work.

2. Fundamentals

This section establishes the conceptual basis of the work – the role of module handbooks and existing metadata standards, the pedagogical concepts of Bloom's Taxonomy and Constructive Alignment, and the combined research gap that motivates the artifact.

2.1 Module Handbooks and Existing Metadata Standards

A module handbook documents, per module, the administrative metadata (name, identifier, credits, workload, responsible person, language), the intended learning outcomes, the content, the teaching and learning methods, and the assessment format. Several standards address parts of this information. IEEE LOM 1484.12.1 offers a mature metadata schema with around 80 fields and broad international adoption, but it relies on verbose XML, carries no pedagogical layer and does not encode the cognitive level of outcomes [6]. The ECTS

Users' Guide anchors workload and credits within the Bologna framework, yet it is purely administrative and does not prescribe a machine-readable format [5]. ELMO v1.0 supports the exchange of transcripts within the EMREX network but is result-focused rather than description-focused. schema.org/Course is web-native and JSON-LD-ready with broad reach, but its vocabulary is shallow and lacks a structured representation of learning outcomes.

Standard	Strengths	What is missing for course comparison
IEEE LOM 1484	~80 metadata fields; technically mature; internationally recognised	No Bloom (C1); no Constructive Alignment (C2); verbose XML, NLP-heavy
ECTS Users' Guide	Bologna context; workload concept; established Europe-wide	Purely administrative; no Bloom; no alignment; no machine-readable format
ELMO v1.0	Transcript exchange; EMREX network; machine-readable	Result-focused, not description-focused; no Bloom; no alignment
schema.org/ Course	Web standard; JSON-LD; actively maintained	High-level vocabulary; no pedagogical depth; no HE-specific fields
Böhm (2021) / SSD	JSON-LD + dual readability; HE context; closest precursor	C1 left open; no formal ECTS/LOM mapping; no empirical evaluation (B3)

Table 1. Existing standards and the gaps relevant to this project (based on 19, 23, 15, 3].

2.2 Bloom´s Taxonomy and Cognitive-Level Coding (C1)

Bloom's revised Taxonomy classifies learning outcomes by cognitive demand into six ascending levels: remember, understand, apply, analyse, evaluate and create. The level is typically expressed through the action verb of an outcome (e.g. "name" → remember, "model" → create). Recent work shows that this classification can be automated with high accuracy – Li et al. (2022) report an F1 of up to 0.95 using BERT-based classifiers, and further studies confirm the feasibility of mapping outcomes to Bloom levels [9], [10]. Despite this technical maturity, a survey of educational ontologies found that none of ten analysed schemas encodes the cognitive level [1]. Cognitive-level coding therefore constitutes the first research gap (C1).

2.3 Constructive Alignment (C2)

Constructive Alignment [8] is the pedagogical principle that intended learning outcomes, teaching and learning activities, and assessment must be coherently coupled. Ramesh et al. (2016) [11] demonstrate that linking outcomes to assessment in a domain ontology is achievable with high accuracy, and Liu et al. (2024) [12] show empirically that generative-AI applications surface the alignment gap. Nevertheless, no established standard formally operationalises the coupling of outcome, method and assessment, which forms the second research gap (C2).

2.4 Machine-Readability and NLP Processability (B3)

Machine-readability requires that a description be parseable into discrete, typed fields rather than interpreted from prose. Text-mining studies on syllabi report solid but limited accuracy when working on unstructured input – for example 83 % with LSA-based methods [4] and improvements with cosine-similarity and Word Mover's Distance approaches [13], [14]. For LLMs specifically, free text yields low extraction quality and elevated hallucination [2], [3]. The absence of any standard that is explicitly designed for NLP processability is the third research gap (B3).

2.5 Related Work and Combined Gap

Concept-matrix analysis (Webster & Watson 2002 [15]) across the reviewed literature shows that the three gaps are addressed in isolation but never jointly. The closest precursor, Böhm (2021), proposes semantically enriched curricula in JSON-LD with dual readability and is the only source close to the project's scope; however, it leaves Bloom coding (C1) and Constructive Alignment (C2) open and provides no empirical evaluation (B3). The central contribution of this work is therefore to combine C1, C2 and B3 in a single, evaluated standard with a formal mapping to existing schemas – a combination that the concept matrix confirms is absent from prior work.

3. Methodology

The project follows the Design Science Research (DSR) methodology of Peffers et al. (2007) [16], which structures the development and evaluation of an artifact into six activities: problem identification, definition of objectives, design and development, demonstration, evaluation and communication. DSR is appropriate here because the central output is a designed artifact – a course-description standard – whose utility must be demonstrated and evaluated rather than merely described. Three complementary methods feed into this process and provide the evidence base for the artifact.

3.1 Systematic Literature Review

To ground the problem definition and derive design requirements for the artifact, a structured, concept-centric literature review was conducted following Webster and Watson (2002) [15].

The search was conducted in three scholarly databases: IEEE Xplore, ACM Digital Library and Google Scholar. Four Boolean search strings were used, covering four thematic areas: machine-readable course descriptions, NLP-based extraction of learning outcomes, educational metadata standards, and course comparability in the Bologna/ECTS context.

Across the twelve database-search-string combinations, the search returned 40,726 records in total. After applying the year filter, 27,192 records remained. Because Google Scholar and some ACM searches returned very large relevance-ranked result sets, the manually screened pool was restricted to the top-ranked records per search string, as documented in the screening log. In total, approximately 295 records were manually reviewed at title level. Title screening reduced the pool to 170 records, abstract screening reduced it to 73 records, and 33 records were retained at the search-run level.

The retained records were exported to Zotero where possible and consolidated manually. Duplicate records and overlaps between search strings and databases were removed before the final eligibility assessment. Based on this consolidation, 20 unique full-text records were assessed for eligibility, of which 17 studies were included in the final concept matrix. The included studies were coded along the three gap dimensions derived from the research questions: cognitive-level coding via Bloom's Taxonomy (C1), Constructive Alignment between outcomes, teaching methods and assessment (C2), and machine-readability/NLP processability (B3).

Inclusion criteria were: (1) the publication addresses course descriptions, syllabi, module handbooks, learning outcomes, educational metadata, curriculum comparability or credit recognition; (2) the publication is relevant to higher education or transferable to higher education contexts; (3) the publication contributes to at least one of the three gap dimensions C1, C2 or B3; and (4) the publication is available in English or German. Exclusion criteria were: (1) publications outside the educational metadata, curriculum or learning-outcome context; (2) publications without relevance to course-level descriptions or comparability; (3) duplicates; and (4) records for which no sufficient abstract or full text was available for eligibility assessment.

The review results were organised in a concept matrix. Each included study was assessed according to whether it addressed Bloom-level coding, Constructive Alignment, machine-readable representation, NLP processability, interoperability with existing standards, and empirical evaluation. This concept-centric synthesis provided the evidence base for the research gap and informed the mandatory-field catalogue of the designed artifact.

3.2 Expert Interviews and Qualitative Content Analysis

To complement the literature review with a practitioner perspective, three semi-structured expert interviews were conducted with stakeholders from Heilbronn University. The interviewees were selected through purposive sampling because their roles directly relate to module handbook creation, study-programme management or credit recognition.

Each interview lasted approximately 30 minutes and covered the current handling of module handbooks, practical problems in course documentation, assessment of alternative standards and perceived implementation barriers. The interviews were analysed using qualitative content analysis based on Mayring (2015). The coding scheme was structured around the research questions and the three gap dimensions C1, C2 and B3.

In addition, each expert assessed four objects – the HHN status quo, the proposed best-practice standard, schema.org/Course and IEEE LOM using eight criteria: completeness (K1), clarity for human readers (K2), pedagogical depth (K3), AI/NLP processability (K4), cross-institutional comparability (K5), effort for authors (K6), alignment with existing standards (K7), and realistic implementability at HHN (K8). Each criterion was rated from 1 = very weak to 5 = very strong. The scores reported in the results section represent the arithmetic mean of the three expert ratings.

3.3 Artifact Evaluation

The artifact was evaluated through an analytical and an experimental strand following Hevner et al. (2004) [18]. The analytical evaluation assessed whether the proposed standard fulfilled the requirements derived from the literature review and expert interviews, especially completeness, clarity, pedagogical depth, machine-readability, interoperability and implementability.

1. The experimental evaluation tested whether structuring a course description according to the proposed standard improves machine processing compared with less structured formats. For this purpose, the same real HHN module, "Einführung in die Wirtschaftsinformatik" (module no. A2.01.01 / 282113; 7.5 ECTS), was prepared in four input formats: Free text – the original unstructured module description as it appears in the HHN module handbook.

2. schema.org/Course – the same content expressed as a JSON-LD record using the schema.org/Course vocabulary.

3. IEEE LOM – the same content expressed in the IEEE LOM 1484.12.1 XML format.

4. Best-practice JSON-LD – the same content encoded according to the proposed three-layer standard with Bloom levels and Constructive Alignment links.

The content basis was kept constant across all formats so that differences in model performance could be attributed primarily to the structure and semantic richness of the input format. Two state-of-the-art large language models were used: Anthropic Claude Sonnet 4.6 and OpenAI GPT-5.5. Each model was presented with identical prompts across three experimental tasks:

1. Information extraction – extracting administrative and pedagogical course information (ECTS, learning outcomes, exam format, prerequisites, teaching methods).

2. Bloom classification – classifying each learning outcome according to Bloom's Taxonomy levels (1–6).

3. EUonAIR advising – answering course-comparison and suitability questions relevant to EUonAIR-style credit recognition.

The outputs were evaluated against a manually defined ground truth. The ground truth was derived from the original HHN module handbook: the administrative and content fields were taken directly from the handbook, and the Bloom level of each learning outcome was coded manually by the authors from the action verb of the original outcome statement, following the classification scheme introduced in Section 2.2. The coding was fixed before the model outputs were scored, and the same answer key was applied unchanged to all four input formats. It must nevertheless be noted that this coding also informed the bloomLevel fields of the best-practice format, so that the target level is stated explicitly in the structured input while it has to be inferred from the free-text and IEEE LOM inputs. The Bloom metric therefore compares a retrieval task against an inference task; the resulting validity threat is discussed in Section 3.4. Three metrics were used:

1. Extraction F1: the harmonic mean of precision and recall for discrete information fields correctly identified in the model output.

2. Bloom-classification accuracy: the proportion of learning outcomes assigned the correct Bloom level (1–6) compared to the expert-coded ground truth.

3. Hallucination rate: the share of factual assertions in the model output that could not be traced to any field in the input document or the ground truth. A response segment was coded as hallucinated if it contained a factual claim about the module (e.g. a credit value, an assessment form, a learning outcome) that was neither present in the input nor derivable from it.

Input-token consumption was recorded as a cost indicator to quantify the trade-off between structural richness and computational expense.

3.4 Methodological Limitations

The methodology has several limitations. First, the literature review was systematic and documented, but the very large result sets in Google Scholar required a restriction to top-ranked records. Second, the expert evaluation is based on a small purposive sample from one institution and is therefore indicative rather than representative. Third, the AI-based evaluation uses one module and two selected LLMs, meaning that absolute performance values may vary with other modules, prompts or model versions. The results should therefore be interpreted as evidence of practical utility within the evaluated setting, not as generalisable performance estimates.

Fourth, and specific to the Bloom metric, the ground truth and the bloomLevel fields of the best-practice format rest on the same manual action-verb coding. Part of the observed gain in Bloom-classification accuracy may therefore reflect the models reading a level that is already present in the input rather than deriving it independently. A leakage-controlled re-analysis – restricting the metric to those outcomes whose cognitive level is not stated explicitly in any input format – was not possible within the scope of this study, because the experiment was not designed to separate the two conditions; it is reported as a priority for future work in Section 6. The extraction and hallucination metrics are not affected by this threat, since their answer key is taken from the original handbook only.

Fifth, each model–format–task combination was executed exactly once. No repeated runs were performed, so no variance, confidence interval or significance test can be reported, and the effect of sampling non-determinism on the individual percentages is unknown. Together with the single-module design, this means that the reported values should be read as an indication of the direction and consistency of the effect across formats, not as measurements of the magnitude of that effect.

4. Results

This section reports the three outcomes of the design and evaluation process: the designed best-practice artifact, its assessment by experts along criteria K1–K8, and the controlled AI experiment comparing free-text with structured input.

4.1 The Designed Artifact

The artifact is a best-practice standard for course descriptions organised in three layers. Layer 1 is a mandatory-field catalogue (module name and identifier, ECTS credits and workload, learning outcomes, EQF/DQR level, prerequisites, language, teaching and exam forms). Layer 2 adds the pedagogical encoding: every learning outcome carries a Bloom level derived from its action verb and a competency category, which also makes the outcome–method–assessment coupling explicit (C1 and C2). Layer 3 ensures interoperability by expressing the description in JSON-LD on the basis of schema.org/Course, with mappings towards IEEE LOM and ECTS. The standard is dual-readable: the same content is rendered as a one-page human-readable PDF and as a machine-readable JSON-LD record.

The standard was instantiated on the compulsory Heilbronn University module “Einführung in die Wirtschaftsinformatik” (module no. A2.01.01 / 282113; 7.5 ECTS; 225 h workload; 6.0 SWS; EQF/DQR level 6; module coordinator Prof. Dr. Philipp Küller). For this module, all six learning outcomes were rewritten as Bloom-anchored statements with an attached competency category, demonstrating that an existing module handbook can be transformed into the proposed format without loss of information.

4.2 Expert Evaluation (K1-K8)

Across the interviews, three cross-cutting findings emerged. Module handbooks are maintained by module coordinators and consolidated by study-programme management, with no formal quality assurance beyond completeness checks; system-related issues create substantial maintenance effort. Compact, one-page formats and ease of maintenance were consistently regarded as critical for adoption, while students rarely consult handbooks because of their complexity. From the internationalisation perspective, recognition relies primarily on external handbooks and is

hindered by inconsistent documentation and ECTS practice; centralised ownership was seen as a way to improve consistency. The comparative scores along K1–K8 are summarised in Table 2.

Criterion (1–5)	HHN status quo	Best practice	schema.org/ Course	IEEE LOM
K1 Completeness	3	4	2	3
K2 Clarity for humans	3	5	3	2
K3 Pedagogical depth	2	4	2	2
K4 AI / NLP processability	2	4	4	4
K5 Cross-institutional comparability	2	3	3	3
K6 Effort for authors	3	4	3	2
K7 Alignment with existing standards	3	3	4	4
K8 Realistic implementability at HHN	4	3	2	2

Table 2. Expert evaluation of four objects across criteria K1–K8 (mean of three interviews, scale 1–5).

The best practice leads on the human and pedagogical criteria – completeness (K1 = 4), clarity (K2 = 5) and pedagogical depth (K3 = 4) – and at least matches schema.org/Course and IEEE LOM on machine-readability (K4 = 4) and comparability (K5 = 3). It trails only on realistic short-term implementability at HHN (K8 = 3), where the established status quo still leads, reflecting the additional authoring effort of a richer format.

4.3 AI-Based Evaluation

In the controlled experiment, structuring the same module markedly improved all three target metrics. The experiment was conducted with Anthropic Claude Sonnet 4.6 and OpenAI GPT-5.5 across the three tasks described in Section 3.3. Each model–format–task combination was run once with identical prompts, and results were averaged across the two models.

Averaged over both models, extraction F1 rose from 74 % (free text) to 94 % (best practice), Bloom-classification accuracy from 64 % to 86 %, and the hallucination rate fell from 27 % to 7 %. The overall index is computed as the arithmetic mean of extraction F1, Bloom accuracy and faithfulness (defined as 1 − hallucination rate), scaled to a 0–100 range. Per-format extraction F1 (×100) was 76/72 for free text, 89/87 for schema.org, 54/50 for IEEE LOM and 95/93 for the best practice (Claude / GPT respectively). IEEE LOM performed worst because it offers no native learning-outcome field and its verbose XML

confused the models. Aggregated into the overall index, the formats ranked: IEEE LOM 58, free text 70, schema.org 85 and best practice 91.

Format	Extraction F1	Bloom accuracy	Hallucination rate	Overall index	Input tokens / call
Free text (control)	74 %	64 %	27 %	70	1 649
IEEE LOM	52 %	56 %	34 %	58	1 561
schema.org/Course	88 %	79 %	12 %	85	2 244
Best practice (exp.)	94 %	86 %	7 %	91	3 415

Table 3. AI-based evaluation across formats (extraction F1, Bloom accuracy, hallucination rate and overall index as means over Claude Sonnet 4.6 and GPT-5.5; input tokens per call as a cost indicator).

The improvement carries a cost: the best-practice JSON-LD consumed roughly twice the input tokens of the free-text version (3 415 vs. 1 649 per call). Structured, Bloom-coded JSON-LD is thus the most reliably usable format for AI processing, but it is also the most token-intensive.

The three metrics should not be read as equally strong evidence. Extraction F1 and the hallucination rate are scored against an answer key taken from the original module handbook and are therefore independent of the format under test. The Bloom metric is not: in the best-practice format the cognitive level is an explicit field, whereas in free text and IEEE LOM it must be inferred from the action verb. The gain of +22 pp thus quantifies the benefit of making the level machine-readable, but it does not isolate how much of that gain reflects improved classification as opposed to successful retrieval of a pre-coded value. Since each combination was run once on a single module, the individual percentages carry no error estimate; what the experiment supports is the consistent ordering of the four formats across both models and all three metrics, not the exact size of the differences.

IEEE LOM's poor performance deserves particular attention. Despite being the most established metadata standard in the educational technology community, its verbose XML structure and the absence of a dedicated learning-outcome field meant that both LLMs struggled to locate and extract the relevant information, frequently misinterpreting XML tags as content. This produced both the lowest extraction F1 (52 %) and the highest hallucination rate (34 %) of any format tested.

5. Findings and Recommendations for Action

The two evaluation strands converge. The AI experiment confirms the value of structure on the machine side: the best practice wins on every metric (extraction F1 +20 pp, Bloom +22 pp, hallucination −20 pp), directly addressing all three gaps – C1 through machine-readable Bloom levels that raise classification quality, C2 through the now-extractable outcome–method–assessment linkage, and B3 through fewer invented details in advising. The interviews confirm the value on the human side, where the best practice scores highest on clarity (K2 = 5) and pedagogical depth (K3 = 4). The findings also position the artifact precisely against the literature: no prior source jointly addresses C1 + C2 + B3 with a formal standard mapping, and the closest precursor [7] leaves exactly the elements this artifact contributes.

There is, however, one consistent trade-off. Practitioners weight authoring effort and realistic short-term implementability (K8 = 3, below the status quo at K8 = 4), and the structured format costs about twice the input tokens. The machine-side gains therefore carry an adoption cost that must be managed rather than ignored.

From these findings the following recommendations for action are derived for HEIs and exchange networks:

- Adopt the mandatory-field catalogue as a minimum schema and treat optional fields as a clearly separated extension layer, so that authoring stays manageable.
- Maintain dual readability: generate the human-readable one-page PDF and the machine-readable JSON-LD from a single source to avoid duplicate maintenance.
- Encode learning outcomes by Bloom level anchored to the action verb of the original outcome, which keeps the coding transparent and auto-classifiable.
- Centralise ownership of module handbooks and add a lightweight quality-assurance step beyond completeness checks, as the interviews identified maintenance and consistency as the key practical risks.
- Pilot the standard with two to three EUonAIR partner institutions to validate cross-institutional comparability and recognition workflows in practice.

Limitations. The evidence base is intentionally focused and should be read with several constraints in mind. The interview sample is small (n = 3) and limited to a single institution, so the practitioner perspective is indicative rather than representative. The AI experiment rests on a single module handbook with one run per model–format–task combination, which yields point values without any estimate of run-to-run variance and makes the absolute percentages dependent on the chosen module, models and prompts. In addition, the Bloom metric is favoured by the best-practice format by construction, because that format states the cognitive level explicitly while the free-text and IEEE LOM conditions require it to be inferred (see Section 3.4). The reported figures should therefore be interpreted as evidence of a clear and consistent direction of effect rather than as precise population estimates, and the Bloom result in particular should be treated as an upper bound until a leakage-controlled re-analysis is available.

6. Conclusion and Outlook

This project set out to close a structural gap: no existing standard combines pedagogical depth with machine-readability. Following the Design Science Research methodology, it developed and evaluated a three-layer best-practice standard for course descriptions that is dual-readable (PDF and JSON-LD) and encodes learning outcomes by Bloom level. Answering the research questions: a compact mandatory-field catalogue with a separated extension layer is sufficient to be both meaningful and machine-processable (RQ1); Bloom's Taxonomy and Constructive Alignment can be operationalised through verb-anchored Bloom levels and an explicit outcome–method–assessment linkage (RQ2); structuring measurably improves extraction and Bloom classification and reduces hallucination (RQ3); and practitioners rate the standard highly on clarity and pedagogical depth while flagging authoring effort as the main barrier (RQ4).

Three lessons learned shaped the work and inform its continuation: a full systematic review is considerably more demanding than an ordinary literature search; understanding the problem thoroughly before designing solutions paid off; and expert interviews need to be scheduled with more lead time than initially planned. The project also confirmed the value of a shared reference manager for collaborative academic work.

Future work follows directly from the limitations, and three steps have priority. First, the AI experiment should be repeated with an evaluation design that separates retrieval from classification: the Bloom answer key should be coded by two independent raters on the original outcome statements before any format is produced, inter-rater agreement should be reported, and Bloom accuracy should be reported separately for the subset of outcomes whose cognitive level is not stated explicitly in any input format. Second, the evaluation should be extended to a larger set of modules and to repeated runs per model–format–task combination, so that means, dispersion and significance can be reported instead of single point values. Third, the revised best-practice artifact (V2) should be validated on these additional modules to strengthen external validity, the expert base should be widened beyond a single institution, and a pilot with EUonAIR partners should test the standard in real recognition workflows. Reducing the token overhead of the JSON-LD representation while preserving its semantic richness is a further worthwhile direction, since it would lower the adoption cost identified in this study.

ISSN 2991-9371

References

[1] Tapia-León, M. et al. (2018): The Use of Ontologies for Syllabus Representation. In: ITHET 2018, pp 1–7.

[2] Vu, B. et al. (2025): InformatIQ: a transformers-powered framework for practical module handbook metadata extraction. In: International Journal of Computer Applications 47 (7), pp 593–610.

[3] Sparmann, S. et al. (2025): Surfacing Educational Traditions in European K–12 Computing Curricula: A Large-Scale NLP-Based Comparison. In: Koli Calling 2025, pp 1–12.

[4] Orellana, G. et al. (2018): A Text Mining Methodology to Discover Syllabi Similarities among Higher Education Institutions. In: INCISCOS 2018, pp 261–268.

[5] Palacio, J. C. C. et al. (2026): An Analysis of ECTS Credit Transferability across Similar Engineering Degree Programs in Spain. In: INTED2026, paper 0958.

[6] Zervas, P. et al. (2014): A Systematic Analysis of Metadata Application Profiles of Learning Object Repositories in Europe. In: LNET, pp 75–91.

[7] Böhm, K. (2021): Towards Semantically Enriched Curricula as pre-built Information Spaces in Higher Education Institutions. In: ECKM 2021, p 71.

[8] Biggs, J. (1996): Enhancing teaching through constructive alignment. In: Higher Education 32 (3), pp 347–364.

[9] Li, Y. et al. (2022): Automatic Classification of Learning Objectives Based on Bloom's Taxonomy. In: Proceedings of EDM 2022.

[10] Kumar, R. et al. (2025): Automated Analysis of Learning Outcomes and Exam Questions Based on Bloom's Taxonomy. arXiv:2511.10903.

[11] Ramesh, R. et al. (2016): Integrating the Learning Objectives and Syllabus into a Domain Ontology for Data Structures Course. In: ITiCSE 2016, pp 266–271.

[12] Liu, Y. et al. (2024): Generative AI in Education: A Review of Applications, Impacts, and Future Research Directions. In: ICIS 2024 Proceedings.

[13] Zaki, N. et al. (2023): Automating the mapping of course learning outcomes to program learning outcomes using natural language processing for accurate educational program evaluation. In: Education and Information Technologies 28 (12), pp 16723–16742.

[14] Krishnan, S. et al. (2025): AI Course Equivalency using Word Mover's Distance and Semantic Analysis with NLP in Cloud Computing. In: ICAISS 2025, pp 448–454.

[15] Webster, J. / Watson, R. T. (2002): Analyzing the Past to Prepare for the Future: Writing a Literature Review. In: MIS Quarterly 26 (2), pp xiii–xxiii.

[16] Peffers, K. et al. (2007): A Design Science Research Methodology for Information Systems Research. In: Journal of Management Information Systems 24 (3), pp 45–77.

[17] Mayring, P. (2015): Qualitative Inhaltsanalyse. Grundlagen und Techniken. 12th ed., Weinheim: Beltz.

[18] Hevner, A. R. et al. (2004): Design Science in Information Systems Research. In: MIS Quarterly 28 (1), pp 75–105.

[19] Badawy, M. et al. (2018): A Text Mining Approach for Automatic Selection of Academic Course Topics based on Course Specifications. In: ICENCO 2018, pp 162–167.

ISSN 2991-9371

[20] Diab, S. / Sartawi, B. (2017): Classification of Questions and Learning Outcome Statements (LOS) into Bloom's Taxonomy (BT) by Similarity Measurements towards Extracting of Learning Outcome from Learning Material. In: International Journal of Managing Information Technology 9 (2), pp 1–12.

[21] Göschlberger, B. et al. (2020): A Standardised Metadata Schema for Learning Resources based on JSON Schema. In: iiWAS 2020.

[22] Han, Z. / Stevens, K. A. (2026): Quantifying Cross-System Curriculum Alignment: A Reproducible and Interpretable Method. In: SoutheastCon 2026, pp 1–7.

[23] Nilsson, M. (2010): From Interoperability to Harmonization in Metadata Standardization: Designing an Evolvable Framework for Metadata Harmonization. PhD thesis, KTH Stockholm.

THE USE OF AI TOOLS

The authors used Claude (Anthropic) for language editing and translation of selected parts of the manuscript. All content was reviewed and verified by the authors.

Input-Output Shock Propagation and Spectral Amplification

Matej Ivaničeka[1] and Yaraslau Svintsitski[2]

[1]Zagreb School of Economics and Management, Croatia, matej.ivanicek@zsem.hr
[2]Zagreb School of Economics and Management, Croatia, ysvintsi@student.zsem.hr

Abstract

This paper studies how sector shocks spread through the 2018 production networks of France and Germany. We compare two simple input-output experiments. In the first, final demand in one sector is reduced by 25% and output is calculated with the static Leontief model. In the second, the short-run output limit of the shocked sector is reduced by 25% in a discrete-time model, while final demand stays at its baseline level. Because the two experiments change different parts of the model, their difference is treated as a scenario difference, not as pure dynamic amplification. We describe sector structure with the output multiplier, a forward-linkage index, and eigenvector centrality. The capacity experiment gives larger losses for suppliers, especially for the energy sector, while the C26 electronics shock has much smaller effects outside the shocked sector. France and Germany also have different network structures: energy has the highest eigenvector centrality in France, while business-service sectors have the highest values in Germany. The model follows changes in intermediate demand. It does not model physical input shortages, prices, inventories, substitution, or international trade.

Keywords: *input-output analysis, Leontief model, production networks, shock propagation, spectral centrality, capacity constraints, reshoring.*

1. Introduction

Sector shocks can spread through production networks from one industry to another. The pandemic, semiconductor disruptions, and the European energy crisis showed how important these links can be. In input-output analysis, it is important to separate a fall in final demand from a limit on production capacity, because these two cases answer different economic questions.

The classical input-output model, developed by Leontief [1], is the standard way to study how a change in final demand affects total output through intermediate inputs. Let A be the matrix of technical coefficients and d the vector of final demand. Sector output is then $x = (I - A)^{-1}d$. The model is simple and widely used, but it does not show how the economy adjusts over time and it does not include short-run production limits.

This paper compares the way shocks spread after a fall in final demand and after a fall in output capacity. We also study how the results are related to the position of the shocked sector in the production network. Following the network view of macroeconomic fluctuations [2], [3], we use three tools. First, we calculate standard input-output multipliers for France and Germany using OECD Inter-Country Input-Output data.

France and Germany are used as two comparison cases, not as examples that represent all of Europe. We choose them because they are large economies and because their network structures differ in the 2018 ICIO data. In 2025, Germany accounted for 23,8% of EU GDP and France for 15,8%, together making up 39,6% of the EU total [4].

Second, we use a simple discrete-time model with production limits: $x_{t+1} = \min\{\bar{x}, Ax_t + d_t\}$. Here, $\bar{x}$ is the vector of sector output limits. In the capacity experiment, the limit for the shocked sector is reduced to 75% of its baseline output. The limits for all other sectors stay at baseline, and final demand does not change. We therefore use baseline output as a simple proxy for short-run capacity. It is not meant to be an exact engineering maximum. Third, we describe each sector with the output multiplier, a forward-linkage index, and eigenvector centrality.

We apply both experiments to Electricity, gas, steam and air conditioning supply (sector D) and to Computer, electronic and optical equipment (sector C26). C26 is a broad electronics sector, not a pure semiconductor sector. We use it as a useful proxy in the context of the European Chips Act [5]. The energy experiment is motivated by the major European energy disruption in 2022 [6]. In the static benchmark, final demand in the selected sector falls by 25%. In the capacity experiment, the output limit of the selected sector falls by 25%.

ISSN 2991-9371

The results give four main points. First, the capacity experiment gives larger losses than the final-demand benchmark because it cuts gross output by 25%, while the static experiment cuts final demand by 25%. The difference should therefore not be called pure dynamic amplification of the same shock. Second, within each country and sector, the pattern of losses across sectors is close to a scaled version of the static result, so the capacity model changes the size of losses more than their ranking. Third, the energy shock has much larger effects on suppliers than the C26 shock. Fourth, the network measures differ clearly between France and Germany: energy has very high eigenvector centrality in France, while business-service sectors have the highest values in Germany.

2. Methodology

2.1 The Static Leontief Framework

Assume the economy has n sectors. Let $Z \in \mathbb{R}^{n \times n}$ be the matrix of transactions between sectors. The element Z_{ij} is the value of output from sector i used as an intermediate input by sector j. Let $x \in \mathbb{R}^n$ be the vector of gross sector output and $d \in \mathbb{R}^n$ the vector of final demand. The basic input-output identity is $x = Z1 + d$. The technical coefficient $a_{ij} = \frac{z_{ij}}{x_j}$ shows how much input from sector i is needed to produce one unit of output in sector j. Since $Z = A\,\mathrm{diag}(x)$, the Leontief solution is $x = (I - A)^{-1} d \equiv Ld$, where $L = (I - A)^{-1}$ is the Leontief inverse. Its column sums $m_j = \sum_i \ell_{ij}$ are output multipliers. They measure the total direct and indirect output created by one unit of additional final demand in sector j.

To model a final-demand shock, suppose sector j receives a shock of size $s_j \in [0,1]$. Its final demand becomes $d_j^{\text{shock}} = (1 - s_j)\, d_j$, and output becomes $x^{\text{shock}} = L\, d^{\text{shock}}$. The loss in sector i is $\text{loss}_i^{\text{FD}} = \frac{x_i^{\text{base}} - x_i^{\text{shock}}}{x_i^{\text{base}}} \cdot 100$. This is our final-demand benchmark. Other dynamic and network models show that adjustment limits and supply-chain links can change how shocks spread [7]-[9]. Our model below is simpler: it follows changes in intermediate demand to suppliers and does not model physical shortages of inputs.

2.2 A Dynamic Model with Capacity Constraints

We use a simple discrete-time model for the capacity experiment: $x_{t+1} = \min\{\bar{x}, Ax_t + d_t\}$. The minimum is applied to each sector. Here, Ax_t is intermediate demand from current production, d_t is final demand at time t, and $\bar{x}$ is the vector of output limits. The simulation starts from baseline output, and final demand stays at baseline. The shocked sector is limited to 75% of baseline output, while all other sectors stay at their baseline limits. Because the ICIO tables do not contain engineering capacity data, we use $\bar{x}$ as this simple short-run limit. When a sector produces less, it buys fewer intermediate inputs, so the effect moves upstream to its suppliers. The model does not represent physical input shortages moving downstream to customers.

If no capacity limit binds, the model converges to the static Leontief solution when the spectral radius of A is below 1. In the capacity experiment, the shocked sector reaches its limit immediately, while other sectors adjust as intermediate demand changes. More detailed models include additional channels [7]-[9]. The loss for sector i at the end of the simulation, period T, is $\text{loss}_i^{\text{cap}} = \frac{x_i^{\text{base}} - x_{i,T}}{x_i^{\text{base}}} \cdot 100$. The difference from the final-demand benchmark is $\text{gap}_i = \text{loss}_i^{\text{cap}} - \text{loss}_i^{\text{FD}}$. Since the two experiments change different parts of the model, a positive gap only means that the capacity case gives a larger loss. It does not prove that the static model underestimates the same shock.

2.3 Metrics of Systemic Importance

To describe each sector's place in the network, we use three measures. The first is the output multiplier, $m_i = \sum_i \ell_{ij}$, which measures backward links. The second is the forward-linkage index, $f_i = \sum_j \ell_{ij}$, which shows how strongly sector i appears in the total output needed across final-demand categories. The third is eigenvector centrality, based on the main right eigenvector of A [10]. This measure gives a high score to sectors that supply other central sectors. Let λ_1 be the largest eigenvalue of A, and let v_1 be its nonnegative right eigenvector, scaled so that its largest value is 1. The vector v_1 satisfies $v_{1,i} \propto \sum_j a_{ij} v_{1,j}$. This is a descriptive measure of supplier-side network position. It is different from the upstream effect produced by the dynamic model [2], [3].

Together, the three measures give different views of a sector's position in the network. They should not be treated as the same measure of risk. The output multiplier is directly linked to the effect of final-demand changes on suppliers. The forward-linkage index and eigenvector centrality describe other parts of the supply network. Differences between the measures are therefore useful information, not a contradiction.

3. Results

The analysis uses the 2021 edition of the OECD Inter-Country Input-Output (ICIO) tables [11]. These tables give harmonised data on production links within and between countries. We use France and Germany for the reference year 2018 because this is before the COVID-19 pandemic and the 2022 European energy crisis. The experiments should therefore be read as simulated shocks applied to the 2018 production structure. After sectors with zero output or empty rows and columns are removed, the sample contains 44 domestic sectors in each country.

For each country, we run two experiments for sector D and sector C26. The final-demand benchmark reduces final demand in the selected sector by 25%. The capacity experiment keeps final demand unchanged but reduces the selected sector's output limit by 25%. The two experiments therefore use the same 25% number but do not represent the same shock. C26 means the broad Computer, electronic and optical equipment sector, not a pure semiconductor sector. The dynamic simulations start from baseline output and run for 20 periods; the paths become stable within this period.

3.1 Systemic Importance and Network Structure

Before looking at the shocks, we first compare the positions of sectors in the two production networks. Table 1 shows the five sectors with the highest eigenvector centrality in each country, together with their output multipliers and forward-linkage values.

Country	Sector	Output multiplier	Forward linkage	Eigenvector centrality
France	D (Energy)	2,30	3,80	1,000
	M (Professional services)	1,64	4,98	0,223
	K (Financial services)	2,02	4,15	0,202
	N (Administrative services)	1,60	4,89	0,195
	G (Wholesale and retail)	1,65	4,76	0,104
Germany	M (Professional services)	1,61	4,27	1,000
	N (Administrative services)	1,58	3,52	0,526
	K (Financial services)	1,77	2,59	0,517
	L (Real estate)	1,39	2,62	0,465
	G (Wholesale and retail)	1,62	4,11	0,446

Table 1. Top five sectors by eigenvector centrality in France and Germany (2018)
Note: Sectors are ranked by eigenvector centrality of the technical coefficient matrix A, normalised so that the maximum value equals one. Output multiplier is the column sum of the Leontief inverse; forward linkage is its row sum. Own calculations based on OECD ICIO data

The two countries show a clear difference. In France, the electricity, gas and steam sector (D) has the highest output multiplier in the table, 2,30, and by far the highest eigenvector centrality, set to 1. The next sector, professional services (M), has a centrality of only 0,223. In Germany, professional services (M) is the most central sector, followed by administrative services (N), financial services (K), real estate (L), and wholesale and retail trade (G). The energy sector (D) is sixth, with a centrality of about 0,28. Energy is therefore especially important in the French centrality ranking, while service sectors hold most of the top positions in Germany.

These differences help explain the shock results, but no single measure is enough to predict losses. The capacity model passes lower intermediate demand upstream to suppliers, while eigenvector centrality gives a broader view of a sector's position in the supply network. We therefore use the measures together as descriptive indicators.

3.2 The Energy Shock

Table 2 shows the results of the energy-sector experiments. In the final-demand benchmark, a 25% fall in final demand for sector D reduces gross output in that sector by 10,98% in France and 11,59% in Germany. The output loss is smaller than 25% because final demand is only one part of the sector's total sales; demand from other sectors is not reduced directly. The largest secondary losses are found among suppliers to sector D. In France, coal mining loses 4,38% of output and oil and gas extraction loses 2,65%. In Germany, the corresponding losses are 6,89% and 1,83%. These numbers describe the 2018 input-output structure and should not be read as current energy-system data.

In the capacity experiment, the output limit of sector D is set to 75% of baseline, so the 25% loss of sector D is imposed directly. Lower output in sector D reduces its demand for intermediate inputs and therefore reduces output among its suppliers. In France, coal mining loses 9,96% in the capacity case, compared with 4,38% in the final-demand benchmark. Oil and gas extraction moves from 2,65% to 6,03%. The ratio between capacity losses and final-demand losses is about 2,28 for almost all reported French sectors. This almost constant ratio is not a separate amplification effect. With one binding capacity limit and linear input-output links, changing the size of the shock mainly scales the whole loss pattern. More generally, production-network research shows how input-output links spread sector shocks through the economy [2], [12].

Country	Sector	Static loss (%)	Dynamic loss (%)	Amplification (pp)
France	D (Energy, shocked)	10,98	25,00	14,02
	B05 (Coal mining)	4,38	9,96	5,58
	B06 (Oil and gas)	2,65	6,03	3,38
	B09 (Mining support)	0,47	1,07	0,60
	C19 (Coke, petroleum)	0,26	0,60	0,34
	E (Water, waste)	0,24	0,54	0,30
Germany	D (Energy, shocked)	11,59	25,00	13,41
	B05 (Coal mining)	6,89	14,86	7,97
	B06 (Oil and gas)	1,83	3,95	2,12
	C19 (Coke, petroleum)	0,42	0,91	0,49
	M (Professional services)	0,31	0,67	0,36
	C23 (Non-metallic minerals)	0,32	0,69	0,37

Table 2. Sectoral output losses from a 25% energy shock (sector D)
Note: Static losses are obtained from with shocked final demand. Dynamic losses are computed at from the constrained model. Amplification is the difference in percentage points.

Germany shows the same basic pattern. The loss of sector D is 25% by construction in the capacity experiment, while the coal-mining loss rises from 6,89% in the final-demand benchmark to 14,86% in the capacity case. The ratio is about 2,16 across the reported sectors, apart from rounding. The largest absolute differences appear in sectors that were already most exposed in the final-demand benchmark. Figure 1 shows this comparison. The main difference between the countries is therefore which suppliers are exposed and how strongly, not a separate country-specific dynamic amplification factor.

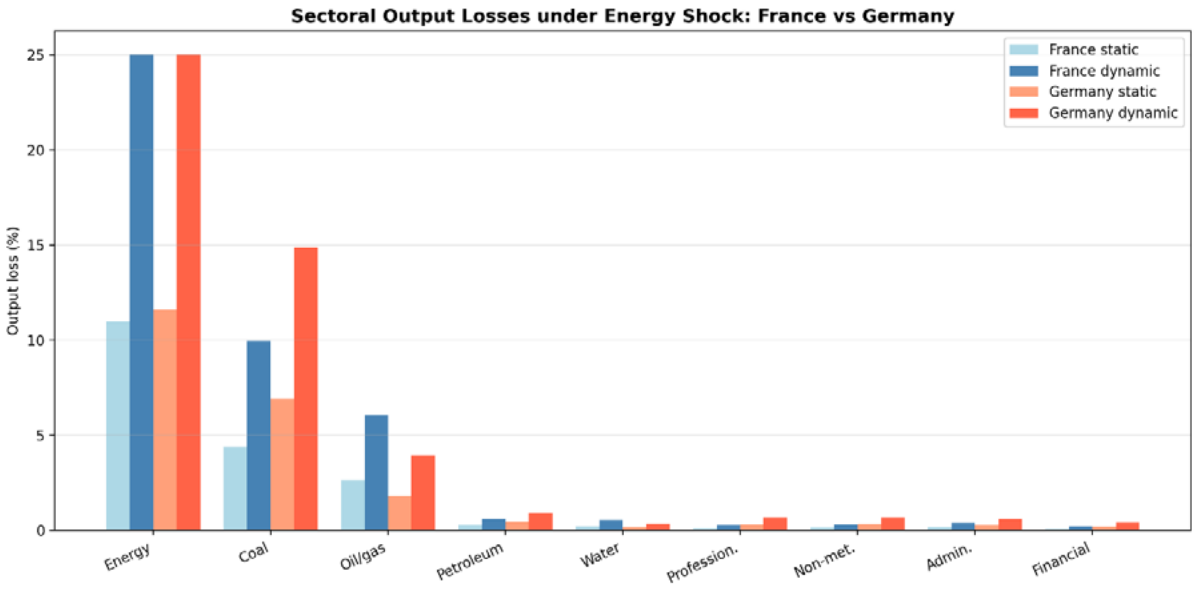

Figure 1. Sectoral output losses following a 25% energy shock in France and Germany, under the static Leontief model and the dynamic constrained model.

3.3 The C26 Electronics Shock

The C26 experiment gives a useful comparison because Computer, electronic and optical equipment has a much less central position in the reported network measures. In the final-demand benchmark, C26 output falls by 22,65% in France, close to the 25% fall in final demand. Losses in all other reported sectors stay below 0,4%. In the capacity experiment, C26 is forced to a 25% output loss, but effects on suppliers stay below 1%. Because the final-demand shock already gives a gross-output loss close to 25%, the two cases are much closer than in the energy experiment.

Country	Sector	Static loss (%)	Dynamic loss (%)	Amplification (pp)
France	C26 (Semiconductors, shocked)	22,65	25,00	2,35
	C25 (Fabricated metals)	0,38	0,42	0,04
	C24B (Non-ferrous metals)	0,31	0,34	0,03
	C22 (Rubber and plastics)	0,26	0,28	0,03
	C23 (Non-metallic minerals)	0,24	0,26	0,02
Germany	C26 (Semiconductors, shocked)	22,66	25,00	2,34
	C24B (Non-ferrous metals)	0,83	0,91	0,09
	H53 (Postal services)	0,69	0,76	0,07
	J62 63 (IT services)	0,37	0,41	0,04
	M (Professional services)	0,36	0,39	0,04

Table 3. Sectoral output losses from a 25% semiconductor shock (sector C26)
Note: Same specification as Table 2. Only the shocked sector and the top downstream sectors are reported.

Germany differs mainly in the size and spread of the supplier effects. The final-demand loss of C26 is 22,66%, while several other sectors have losses of about 0,3% to 0,9%. The capacity values are only slightly larger. The calculated eigenvector centrality of C26 is about 14 times higher in Germany than in France. In this model, however, the affected sectors are suppliers to C26 because lower C26 output reduces demand for intermediate inputs. The results should not be read as shortages in industries that use semiconductors. More generally, supply-chain studies show that network structure matters for the spread of shocks, although those studies use richer models than the simple model used here [8], [9].

3.4 Final-Demand Benchmark versus Capacity-Constrained Propagation

Across the four country-sector cases, the capacity experiment gives losses that are at least as large as the final-demand benchmark. This does not mean that the static Leontief model always underestimates the same shock, because the two experiments change different parts of the model. Within each case, the ratio of capacity loss to final-demand loss is nearly constant across affected sectors: about 2,28 for the French energy experiment, 2,16 for the German energy experiment, and about 1,10 for the C26 experiment in both countries. This shows that the current capacity model mainly changes the size of the static loss pattern rather than the ranking of affected sectors. Network structure determines which suppliers are exposed, while the size and type of the shock mainly determine the scale of the losses.

Figure 2 compares final-demand losses with capacity losses in all four cases. Points above the 45° line have larger losses in the capacity experiment. Because the two cases use different shocks, the distance from the line is a scenario difference, not a pure measure of dynamic amplification.

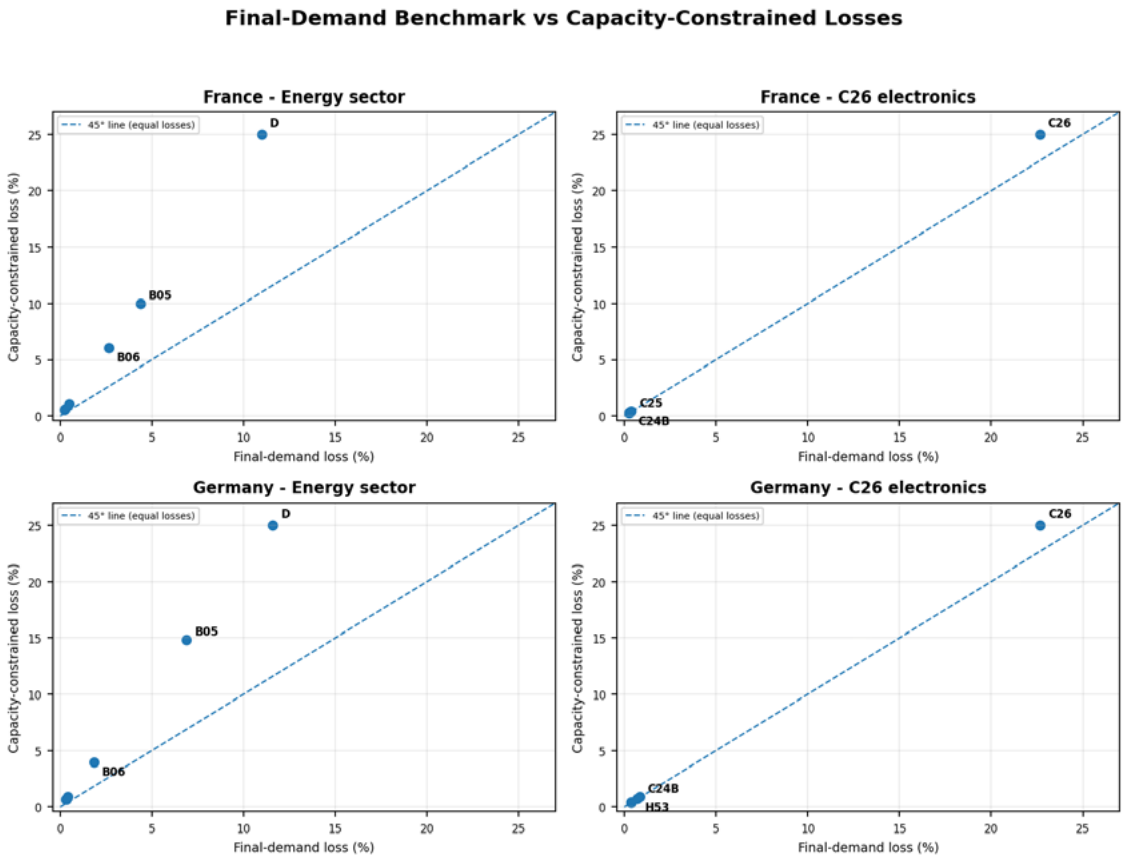

Figure 3 shows how output changes over time for the six most affected sectors in each capacity experiment. The effects on suppliers do not appear at once. Lower intermediate demand moves through the model over several periods before the system becomes stable. These paths show changes in demand for inputs, not physical shortages moving downstream

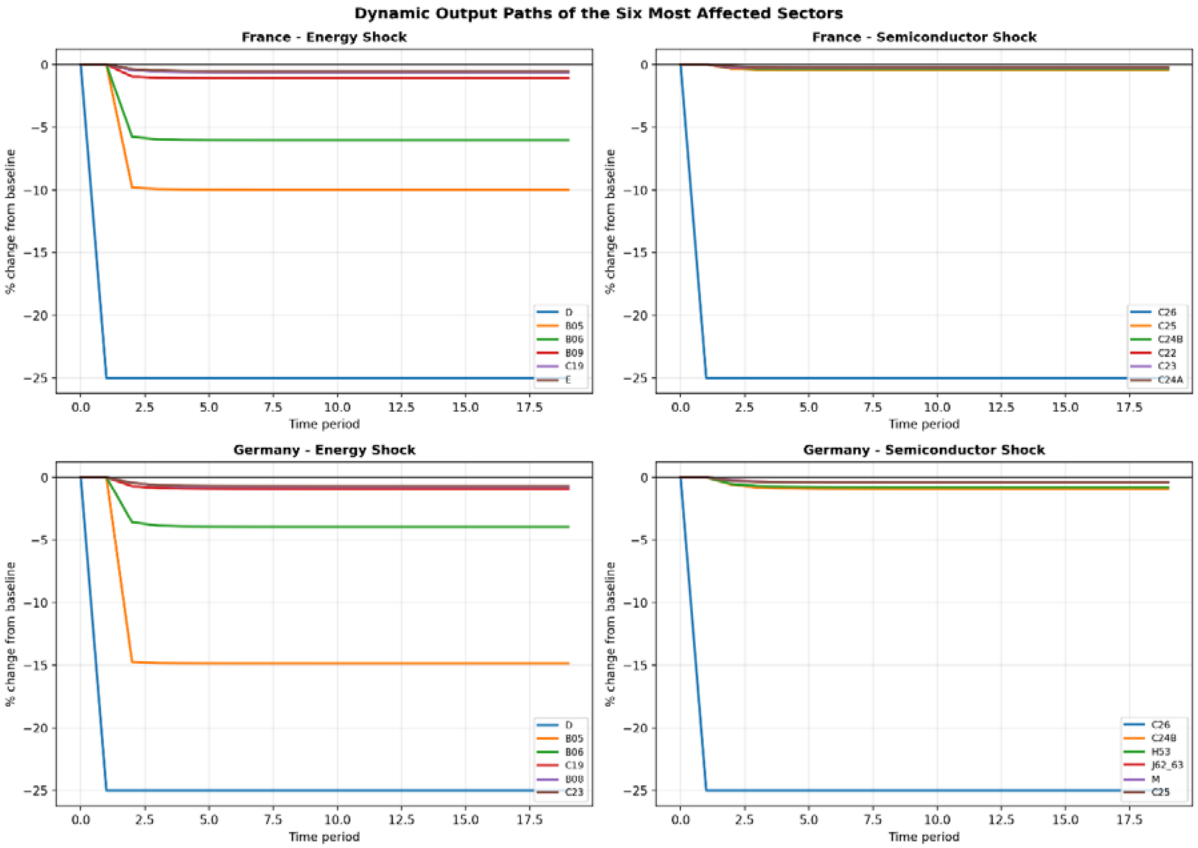

Figure 3. Simulated dynamic output paths for the six most affected sectors in each scenario over a 20-period horizon.

4. Discussion

The results give three broader points. First, the two experiments should be compared as two different shock definitions, not as a direct test of static versus dynamic accuracy. A 25% cut in gross-output capacity is stronger than a 25% cut in final demand when final demand is only part of the sector's sales. If the same negative final-demand shock is used in both models, the baseline output limit may not bind. When the spectral radius of the technical-coefficient matrix is below 1, the simple unconstrained dynamic model converges to the static Leontief result. A true comparison of dynamic amplification would therefore need the same outside shock in both models and extra features such as inventories, input bottlenecks, or explicit production functions, as in richer dynamic input-output models [7]. The static Leontief model still remains a useful accounting benchmark [13].

Second, the country comparison shows differences in supplier exposure in the 2018 ICIO structure. In France, the energy-sector shock gives sizeable losses for coal, oil and gas, and other suppliers. In Germany, the response of coal mining is especially strong. These results come from the 2018 input-output coefficients and should not be treated as statements about today's energy mix or energy policy without newer data.

Third, the C26 results do not prove either resilience or dependence on imports. The model uses domestic production links and does not model international sourcing directly. Weak domestic propagation therefore only means that supplier effects inside the modelled network are small. The European Chips Act aims to strengthen Europe's semiconductor sector [5], but studying how reshoring changes risk would require explicit trade links and a model in which input shortages can limit downstream customers. A possible 'reshoring paradox' should therefore be treated as a question for a larger model, not as a result proved by this analysis.

5. Conclusion

This paper combines the standard input-output model, a simple dynamic model with capacity limits, and network measures for France and Germany. Four conclusions follow. First, a 25% final-demand shock and a 25% output-capacity shock are different shocks, so the larger losses in the capacity case should not be called dynamic amplification of the same shock. Second, within each country and sector, the capacity loss pattern is close to a scaled version of the final-demand result, so the current model changes the size of losses more than their ranking. Third, the energy-sector shock has much stronger effects on suppliers than the C26 electronics shock. Fourth, France and Germany have different supplier-network structures: energy is especially central in France, while business-service sectors hold the highest centrality positions in Germany.

The model is intentionally simple. Useful next steps would be to use the same outside shock in different model versions, add inventories and limits on specific inputs, allow substitution and price changes, and include international trade directly. These additions would be needed to study physical supply shortages, downstream effects, or dependence on imports. In its current form, the model is best seen as a clear example of how lower intermediate demand can spread upstream through suppliers and how network structure differs between countries.

References

[1] W. W. Leontief, "Quantitative Input and Output Relations in the Economic Systems of the United States," The Review of Economic Statistics, vol. 18, no. 3, pp. 105-125, 1936, doi: 10.2307/1927837.

[2] D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi, "The Network Origins of Aggregate Fluctuations," Econometrica, vol. 80, no. 5, pp. 1977-2016, 2012, doi: 10.3982/ECTA9623.

[3] V. M. Carvalho and A. Tahbaz-Salehi, "Production Networks: A Primer," Annual Review of Economics, vol. 11, pp. 635-663, 2019, doi: 10.1146/annurev-economics-080218-030212.

[4] Eurostat, "EU's GDP valued at €18.8 trillion in 2025," 6 Aug. 2026. Source dataset: nama_10_gdp. [Online]. Available: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/wdn-20260806-1. [Accessed: 22 Sep. 2026].

[5] European Parliament and Council of the European Union, "Regulation (EU) 2023/1781 of 13 September 2023 establishing a framework of measures for strengthening Europe's semiconductor ecosystem and amending Regulation (EU) 2021/694 (Chips Act)," Official Journal of the European Union, L 229, pp. 1-53, 18 Sep. 2023.

[6] E. Çam and C. Fernández Alvarez, "Europe's energy crisis: Understanding the drivers of the fall in electricity demand," International Energy Agency, 9 May 2023. [Online]. Available: https://www.iea.org/commentaries/europes-energy-crisis-understanding-the-drivers-of-the-fall-in-electricity-demand. [Accessed: 22 Sep. 2026].

[7] A. Pichler, M. Pangallo, R. M. del Rio-Chanona, F. Lafond, and J. D. Farmer, "Forecasting the propagation of pandemic shocks with a dynamic input-output model," Journal of Economic Dynamics and Control, vol. 144, art. 104527, 2022, doi: 10.1016/j.jedc.2022.104527.

[8] V. M. Carvalho, M. Nirei, Y. U. Saito, and A. Tahbaz-Salehi, "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake," The Quarterly Journal of Economics, vol. 136, no. 2, pp. 1255-1321, 2021, doi: 10.1093/qje/qjaa044.

[9] H. Inoue and Y. Todo, "Firm-level propagation of shocks through supply-chain networks," Nature Sustainability, vol. 2, pp. 841-847, 2019, doi: 10.1038/s41893-019-0351-x.

[10] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," The Journal of Mathematical Sociology, vol. 2, no. 1, pp. 113-120, 1972, doi: 10.1080/0022250X.1972.9989806.

[11] N. Yamano, J. Guilhoto, A. Alsamawi, C. Webb, P. Horvát, A. Cimper, C. Zürcher, and X. Han, "Development of the OECD Inter-Country Input-Output Database 2021," OECD Science, Technology and Industry Working Papers, no. 2021/14, OECD Publishing, Paris, 2021, doi: 10.1787/71ada7d2-en.

[12] D. R. Baqaee and E. Farhi, "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem," Econometrica, vol. 87, no. 4, pp. 1155-1203, 2019, doi: 10.3982/ECTA15202.

[13] R. E. Miller and P. D. Blair, Input-Output Analysis: Foundations and Extensions, 2nd ed. Cambridge: Cambridge University Press, 2009, doi: 10.1017/CBO9780511626982.

THE USE OF AI TOOLS

The authors used ChatGPT (OpenAI) and Claude (Anthropic) for language editing and translation of selected parts of the manuscript. All content was reviewed and verified by the authors.

Geopolitical Risk, Oil Supply Shocks, and the Macroeconomic Transmission of the 2026 Iran-Israel-U.S. War

Josip Tomašković

Zagreb School of Economics and Management (ZSEM), Zagreb, Croatia,
email: jtomask1@student.zsem.hr

Abstract

This paper analyzes the macroeconomic transmission of the 2026 conflict in the Middle East involving Iran, Israel, and the United States, viewing it not only as a regional conflict but also as a global supply and logistics shock associated with severe disruption of transit through the Strait of Hormuz. Focusing on the Strait of Hormuz, through which roughly a quarter of world seaborne oil trade and a fifth of liquefied natural gas (LNG) trade pass, the paper develops a multi-channel transmission framework linking energy supply disruption, maritime and logistics frictions, financial and sanctions fragmentation, geopolitical-risk-driven expectations, and policy responses. It assembles quantitative evidence on energy markets and shipping conditions, together with qualitative evidence on financial-market reactions following the outbreak of hostilities on 28 February 2026, showing a sharp rise in crude oil benchmarks, widening regional gas spreads, elevated war-risk insurance premia, and reported risk-off repricing. The central argument is that a chokepoint- induced shock differs from a conventional oil-price shock because it raises both the cost and risk of moving energy, rather than affecting only market supply, thereby increasing risk premiums that propagate into core inflation and can generate stagflationary pressures in the global economy. The paper concludes with implications for monetary policy, strategic petroleum reserves, and supply-chain diversification, and argues that the chokepoint lens can be generalized to other critical maritime passages.

Keywords: *geopolitical risk, oil price shocks, Strait of Hormuz, chokepoints, stagflation, energy security*

ISSN 2991-9371

1. Introduction

This paper argues that the military conflict involving Iran, Israel, and the United States represents not only a regional military and geopolitical conflict affecting the Middle East but also a macroeconomic shock affecting global crude oil trade, LNG trade, and refined petroleum product markets. Through the cascading effects of multiple transmission channels, this shock spreads across the global economy, generating higher transportation costs, supply chain disruptions, rising inflation, and heightened financial-market volatility [1]. In addition, the paper examines the crucial role of the Strait of Hormuz as a global trade chokepoint and its direct and indirect effects on the world economy [2]. Methodologically, the paper combines insights from academic literature with recent media reports on developments in the Strait of Hormuz. It adopts a macro-level approach to analyze the strategic importance of the Strait, supported by evidence from market data and policy reports. This work contributes to the literature in three ways. First, it frames a regional war as a global macroeconomic shock operating through a maritime chokepoint, connecting the geopolitical-risk literature [3] to the literature on oil-price shocks and the macroeconomy [4], [5], [6]. Second, it specifies a cascading, multi-channel transmission mechanism, rather than treating the event as a single oil-price impulse. Third, it focuses on cross-country exposure to the geopolitical risks associated with the Strait of Hormuz, analyzing how recent developments across the Middle East, including disruptions in the Strait of Hormuz, the Persian Gulf, U.S. military bases, and other strategic locations, as well as in Iraq, Lebanon, and Syria, have contributed to instability in global energy and financial markets. The paper previews several pieces of stylized evidence developed in later sections: a sharp increase in Brent and WTI crude benchmarks, widening regional natural-gas spreads, qualitative evidence of risk- off repricing across financial markets, and a steep rise in war-risk insurance premia accompanied by curtailed transit through the Strait. Taken together, this evidence is consistent with a contractionary supply-side shock that creates stagflationary risk rather than a benign demand-driven increase in energy prices. The remainder of the paper is organized as follows. Section 2 sets out the background of the 2026 Iran-Israel-U.S. war and the strategic importance of the Strait of Hormuz. Section 3 develops the conceptual framework of chokepoint-driven macroeconomic transmission. Section 4 presents stylized evidence from energy, financial, and trade markets. Section 5 examines cross-country exposure and heterogeneity. Section 6 explains why this episode is consistent with a stagflationary supply shock rather than a standard oil shock. Section 7 draws out policy implications, and Section 8 concludes. The broader crisis literature likewise emphasizes global economic interdependence and the spillover of shocks across financial and policy domains [7], [8]. Empirically, the paper uses a descriptive event-based approach rather than a causal event-study design. Energy- market evidence is drawn from EIA daily and monthly series using 27 February 2026 as the pre-conflict benchmark; financial-market and shipping evidence is reported qualitatively where verified point estimates are not available. Accordingly, the empirical results are interpreted as stylized evidence consistent with the proposed transmission mechanisms, not as estimates of causal macroeconomic effects.

2. Background: The 2026 Iran-Israel-U.S. War and Strategic Chokepoints

2.1 Timeline of the Conflict

On 28 February 2026, the United States and Israel launched large-scale strikes against Iranian military and leadership targets, marking the start of the conflict [9]-[11]. In official statements, the U.S. referred to the operation as Operation Epic Fury, while Israel called it Operation Roaring Lion [9], [11]. According to both governments, the strikes were intended to weaken Iran's ballistic missile capabilities and stop further progress in its nuclear programme [12], [11]. Reports published on 1 and 2 March 2026 stated that Iran's Supreme Leader, Ali Khamenei, had been killed in the opening strikes, together with several senior Iranian officials. Among those reportedly killed were Defence Minister Aziz Nasirzadeh and IRGC commander Mohammad Pakpour, while some reports also named Ali Shamkhani, Secretary of the Iranian Defence Council, as one of the casualties [13], [14]. In response, Iran, under the leadership of Mojtaba Khamenei, launched retaliatory missile and drone attacks against U.S. embassies and military bases across the region, including in the United Arab Emirates, Saudi Arabia, Qatar, Kuwait, Bahrain, Iraq, Oman, and Jordan [15]-[19]. Iran also attempted to strike Turkey, but these missiles and drones were intercepted by NATO forces [20]. The conflict quickly spread beyond direct military confrontation and began to affect maritime security in the Gulf. On 28 February 2026, shortly after the U.S.-Israeli strikes, the Islamic Revolutionary Guard Corps began issuing VHF radio warnings that ships were no longer allowed to pass through the Strait of Hormuz [21]. The first commercial vessel attacked during this period was the Marshall Islands-flagged oil tanker MKD VYOM, which was struck by an unmanned explosive boat on 1 March, around 52 nautical miles off the coast of Muscat, Oman. The attack killed one Indian crew member [22], [23]. On 2 March, the IRGC officially confirmed that the strait was closed and warned that any vessel entering it could be set on fire [24], [25]. That same day, the U.S.-flagged tanker Stena Imperative was hit by two projectiles while docked at the Port of Bahrain, killing one shipyard worker and injuring two others [21], [26]. By 4 March, the IRGC claimed that it had established "complete control" over the waterway [27]. Iran later expanded its attacks on commercial shipping across the wider Gulf region, including strikes near Ras Tanura in Saudi Arabia and at the Port of Dubai, where the Kuwaiti tanker Al Salmi was hit by an Iranian drone on 31 March [28]. These developments caused serious disruption to transport through the Strait of Hormuz. Around 150 vessels were left stranded, while war-risk insurance was withdrawn from 5 March and insurance premiums rose sharply [29], from around 0.2% to 1% of a ship's value. In the weeks that followed, Iran moved from a policy of total closure to one of selective and tightly controlled passage. Only approved "non-hostile" vessels were allowed to transit, and even then, under Iranian supervision and along designated routes. At the same time, true alternatives to Hormuz were limited mainly to existing oil and gas pipelines, especially the Saudi East-West pipeline and the UAE ADCOP route. For cargoes that could bypass or exit the Gulf, onward routing was further constrained by insecurity in the Red Sea and the Bab el-Mandeb corridor, which represents another chokepoint risk rather than serving as a substitute for Hormuz. Although

Pakistani mediation helped secure a ceasefire framework on 7 and 8 April, commercial traffic through the strait remained well below normal levels. This suggests that negotiated military de-escalation did not immediately restore freedom of navigation or normal market conditions.

2.2 Importance of the Strait of Hormuz

In 2025, around 20 million barrels per day passed through the strait, accounting for roughly 25% of global seaborne oil trade. Of this amount, approximately 15 million barrels were crude oil, representing about 34% of the global crude oil trade. In addition, LNG shipments through the strait accounted for around 19% of global LNG trade. This route is especially important for Qatar and the United Arab Emirates, since around 93% of Qatar's LNG exports and 96% of the UAE's LNG exports pass through it. Around 80% of the oil transiting the Strait of Hormuz, including crude oil and oil products, is destined for Asia. China and India together account for 44% of the crude oil exports passing through this route, while Europe receives only around 600 thousand barrels per day (kb/d), or about 4% of these crude flows [30]. Between 45% and 50% of China's crude oil imports and around 31% of its liquefied natural gas are transported via the Strait of Hormuz from the Middle East [29], [31]. Table 1 summarizes the importance of the Strait of Hormuz for global energy flows.

Indicator	Value	Source
Oil flow (mb/d)	~20	IEA/EIA
Share of global oil trade	~25%	IEA
Share of LNG trade	~19%	IEA
Share of oil flows destined for Asia	~80%	IEA
China and India share of Hormuz crude flows	~44%	IEA
China LNG imports via Hormuz	31%	Center on Global Energy Policy at Columbia
Qatar LNG exports via Hormuz	~93%	IEA
UAE LNG exports via Hormuz	~96%	IEA

Table 1. Global Exposure to the Strait of Hormuz
Note: Figures are approximate and refer to 2025 flows where available.
Source: Center on Global Energy Policy at Columbia [31], International Energy Agency [30], and U.S. Energy Information Administration [37].

3. Conceptual Framework: Chokepoint-Driven Macroeconomic Transmission

The framework in this section is stated in general terms: conditional language describes how a chokepoint disruption can propagate through the economy, while Section 4 evaluates the 2026 episode using the available descriptive evidence.

3.1 Energy Supply Shock

Any disruption of transport through the Strait of Hormuz would be a major shock to global energy security. Fatih Birol, the Executive Director of the IEA, described it as "the greatest threat to global energy security in history" [32]. The IEA also warned that such a conflict could cause the largest disruption in the history of the global oil market. According to the IEA, in that scenario the global oil supply could fall by 8 million barrels per day, while the global liquefied natural gas (LNG) supply could decline by around 20% [32]; these are external scenario estimates rather than realized effects of the 2026 episode. What matters is not only whether oil and gas production continues, but also whether these resources can still reach international markets safely, on time, and at a reasonable cost. Even if production itself is not interrupted, restrictions in the Strait of Hormuz would still reduce the amount of energy effectively available to the market. Shipments would become slower, more expensive, and riskier. As a result, markets would likely react as if supply had already fallen, thereby pushing prices upward even before an actual physical shortage appeared [3], [5]. Companies will likely face rising production and transport costs, while households will have to pay more for utilities such as electricity and gas. This would lead to higher inflation and reductions in real income and purchasing power, which could lower overall economic growth [1], [33], [34]. Disruptions in supply chains, volatility in financial markets, changes in investor and consumer expectations, and government policy responses all shape how the crisis spreads through the global economy [34], [2]. Comparable evidence links oil and gas price shocks to inflation, output losses, and changing energy-market fundamentals [35], [36].

3.2 Maritime and Logistics Disruption

According to the U.S. Energy Information Administration [37], Saudi Arabia and the UAE have infrastructure that could partly reduce the negative effects of closing the Strait of Hormuz. However, these alternative routes do not have enough capacity to carry the full volume of crude oil that normally passes through the strait. The options for rerouting LNG are even more limited [37], [30]. Saudi Arabia's East-West pipeline and the UAE's Abu Dhabi Crude Oil Pipeline (ADCOP) could provide some additional transport capacity, but much of that capacity is already used under normal conditions [37]. Because of this, these routes could only partially offset the supply shock. A large share of oil exports will likely remain exposed to higher shipping costs, delays, and logistical bottlenecks. LNG exports from Qatar and the UAE are especially vulnerable, since there are no comparable alternative export routes available for them [30]. Any disruption

would amplify the original energy shock by raising transport costs and delaying the delivery of goods. Research on resilient operations similarly stresses alternative sourcing, supply-chain reconfiguration, and adaptive logistics capacity [38]-[40].

3.3 Financial and Sanctions Fragmentation

This conflict does not affect commodity markets alone; it also exerts its effects through the global financial system and through the established sanctions system. Two distinct mechanisms are at work. The first is a repricing of risk. As the probability and expected severity of supply disruption rise, investors demand higher compensation for holding exposed financial assets, capital retreats toward perceived safe havens, and financing costs rise for exposed firms, while sovereign borrowing costs may increase in countries most dependent on Gulf energy flows. The second mechanism is fragmentation: the use of financial sanctions, asset freezes, and restrictions on shipping, insurance, and payments to coerce or isolate a belligerent, which simultaneously fractures the channels through which energy and capital normally circulate. Measures targeting Iranian crude exports, tanker operators, and the entities that insure and finance them can reduce the volume of oil reaching the market, reinforcing rather than offsetting the supply shock. Restrictions on protection-and-indemnity (P&I) insurance are particularly consequential, because vessels cannot transit or call at major ports without adequate insurance coverage [29]; the withdrawal of war-risk cover and the sharp rise in premia documented in Section 2 illustrate how an insurance shock can curtail effective shipping capacity even where physical routes remain open. The potential scale of such a disruption is substantial: estimates cited by the World Economic Forum suggest that an extended interruption could generate global economic losses of up to USD 5.7 trillion, creating pressure for governments to act as insurers of last resort [29]. At the same time, sanctions accelerate the development of parallel arrangements, alternative payment systems, “shadow fleet” tankers operating outside mainstream insurance and flag regimes, and settlements in non-dollar currencies, fragmenting global energy trade into less transparent and less efficient segments [3], [1]. This fragmentation raises transaction costs, complicates price discovery, and can leave a durable wedge between the price that exposed buyers pay and headline benchmark prices. Financial fragmentation also has a macroprudential dimension. Energy importers running large current-account deficits face simultaneous terms-of-trade and financing shocks: import bills rise just as risk-averse capital flows toward safe assets (for example, the U.S. dollar), putting pressure on exchange rates and, for emerging markets, raising the spectre of disorderly adjustment [1], [34]. The result is that the financial channel can amplify and prolong the real shock, transmitting a Gulf disruption to economies with no direct physical exposure to the Strait of Hormuz.

3.4 Expectations and Geopolitical Risk

Much of a chokepoint shock is transmitted before, and often independently of, any change in physically delivered volumes. The relevant variable for forward-looking markets is not realized supply but the expected distribution of future supply, including the probability of escalation and the persistence of disruption. Caldara and Iacoviello [3] show that elevated geopolitical risk, measured from the intensity of news coverage of adverse geopolitical events, systematically foreshadows lower investment and employment and is associated with higher downside macroeconomic risk; both the threat and the realization of such events matter. This expectations channel explains several features of the 2026 crisis. Crude benchmarks and regional gas spreads moved sharply on news of the strikes and the closure of the Strait, before any sustained physical shortfall could be measured, consistent with markets pricing the expected probability and severity of future disruption rather than current inventories alone [3], [5]. Risk premia embedded in oil futures, freight rates, and insurance widened in tandem, reflecting not only expected supply disruption but also greater downside risk to future energy availability, including a rising probability of prolonged restrictions. Because expectations are self-referential, they can be destabilizing: precautionary inventory accumulation, hedging demand, and the rerouting of cargoes can tighten physical balances and validate the very price increases that anticipation set in motion. Crucially, the persistence of geopolitical risk matters as much as its level. A shock widely expected to be transitory leaves long-term expectations and investment plans largely undisturbed; one perceived as structural, signaling a durable elevation in the risk of operating through a chokepoint, raises required returns on exposed activity, discourages investment in affected supply chains, and can shift the economy's medium-run supply curve inward. The ceasefire framework of early April, after which traffic through the Strait remained well below normal, illustrates how geopolitical risk can stay elevated even after active hostilities subside, sustaining a wedge between de-escalation and economic normalization [2], [3].

3.5 Policy Response Channel

The final channel concerns how authorities respond, because policy is endogenous to the shock and shapes its ultimate macroeconomic footprint. Three policy levers are central to the analysis: monetary policy, strategic energy reserves, and trade policy. Each can either cushion or amplify transmission. Monetary policy faces the classic dilemma posed by an adverse supply shock. An energy-driven increase in headline inflation, occurring alongside weakening real activity, confronts central banks with a trade-off between stabilizing prices and supporting output that does not arise for demand shocks [5], [6]. Strategic petroleum reserves and coordinated stock releases offer a direct, but limited, offset to the physical shock. Finally, trade and logistics policies-including the activation of alternative pipelines, the prioritization of cargoes, diplomatic efforts to secure safe passage, and naval escort or convoy arrangements-determine how much of the latent rerouting capacity discussed in Section 3.2 is mobilized. The interaction of these levers, and the credibility of the institutions deploying them, conditions whether a chokepoint shock is contained or allowed to expand into a broader macroeconomic downturn. At the firm level, crisis resilience also depends on adaptive operating models, entrepreneurial bricolage, organizational learning, and sustainable innovation [42]-[45].

4. Stylized Evidence: Market, Shipping, and Trade Responses

4.1 Energy Markets

Energy markets registered the shock immediately and forcefully. As Table 2 shows, the Brent crude spot price rose by roughly 79% from its pre-war level, and WTI by about 70%, while natural-gas price differentials between the United States and the import-dependent Asian and European markets widened sharply. The much larger move in regional gas spreads than in the U.S. benchmark (Henry Hub) is consistent with a disruption concentrated on internationally traded, seaborne energy rather than on domestic North American supply, precisely the signature of a Strait of Hormuz chokepoint shock. Table 2 reports the daily and monthly energy-market reactions.

Panel A. Daily spot prices and daily oil spread

Variable	Pre-war (27 Feb 2026)	Peak after 28 Feb 2026	Latest available value	% change*
Brent crude oil spot price ($/bbl)	71.32	127.61	127.61	+78.9%
WTI crude oil spot price ($/bbl)	66.96	114.01	114.01	+70.3%
Henry Hub natural gas spot price ($/MMBtu)	2.99	3.27	3.04	+1.7%
Brent-WTI spread ($/bbl)	4.36	23.83	14.38**	+229.8%

Panel B. Monthly natural gas spreads

Variable	February 2026	March 2026	% change
Henry Hub-JKM spread ($/MMBtu)	7.69	15.23	+98.0%
Henry Hub-TTF spread ($/MMBtu)	8.14	14.89	+83.0%

Table 2. Oil and Gas Market Reaction

* In Panel A, percentage change is calculated as ((latest available value - pre-war) / pre-war) x 100. ** For the Brent-WTI spread, the latest available value refers to the latest common available date for both daily series, 2 April 2026, so that the spread is calculated on a consistent basis. The column should be read as an early-April value from the April 2026 EIA release, not as a real-time current price.
Source. Author's calculations based on U.S. Energy Information Administration (EIA) daily price series for Brent, WTI, and Henry Hub, and the EIA Short-Term Energy Outlook (April 2026) for the Henry Hub-JKM and Henry Hub-TTF spreads. In Panel A, the pre-war reference value is 27 February 2026, the last trading day before the outbreak of the conflict on 28 February 2026. For Brent, the EIA daily series available in the April 2026 release ends on 2 April 2026, while for WTI and Henry Hub the latest available daily value is 6 April 2026. In Panel B, monthly averages are reported; therefore, the columns are labelled "February 2026" and "March 2026." February values were reconstructed using the inverse growth-rate formula.

4.2 Financial Markets

Available market reporting points to a broad risk-off repricing around the outbreak of conflict, consistent with the expectations channel described in Section 3.4. Broad equity indices were reported to decline, while implied- volatility measures and traditional safe-haven assets moved in directions consistent with heightened uncertainty and risk aversion. Table 3 therefore summarizes qualitative financial-market patterns rather than presenting a fully quantified event study; these patterns should be read as descriptive, not causal, evidence.

Indicator	Direction after outbreak	Macroeconomic interpretation
Global equities (MSCI World)	Declined relative to the pre-war benchmark	Risk-off repricing and weaker expected earnings
Equity-market volatility (VIX)	Increased sharply	Higher uncertainty and wider risk premia
Gold	Increased	Safe-haven demand
U.S. dollar index (DXY)	Appreciated	Safe-haven flows and pressure on energy-importer currencies
10-year U.S. Treasury yield	Fell during peak stress, then partly normalized	Flight-to-safety demand despite higher inflation risk
Breakeven inflation / inflation compensation	Increased	Markets priced higher expected inflation from the energy shock
Geopolitical Risk (GPR) index	Spiked and remained elevated	Geopolitical threats and events affected expectations

Table 3. Qualitative Financial-Market Patterns

Note: The table summarizes qualitative market directions reported in market and policy sources rather than point estimates; it is not a quantified event study. The GPR concept follows Caldara and Iacoviello [3].

The configurationof these reported moves is consistent with a supply-shock interpretation: weaker risk-asset pricing, higher volatility, stronger safe-haven demand, and higher inflation compensation can coexist when markets price both weaker activity and higher input costs. Taken together, the qualitative patterns suggest that the financial channel extended the effects of the conflict beyond economies with direct physical exposure to the Strait of Hormuz [3], [1]. Cross-country evidence further indicates that oil-price movements affect stock markets differently in net oil-importing and net oil-exporting economies [46].

4.3 Shipping and Trade Disruptions

The conflict's most direct real-economy footprint, beyond energy prices, was on shipping and trade logistics. The withdrawal of war-risk insurance and the rise in premiums from roughly 0.2% to about 1% of hull value raised the marginal cost of every transit and, for many operators, made passage commercially or contractually impossible. With the Strait effectively closed and later subject to selective, supervised transit, roughly 150 vessels were stranded, and shipping that continued faced longer voyages, higher bunker consumption, and tighter vessel availability. True alternatives to Hormuz were mainly limited to bypass pipelines, especially Saudi Arabia's East- West pipeline and the UAE's ADCOP route. For cargoes that could bypass or exit the Gulf, onward routing toward Europe was further constrained by insecurity in the Red Sea and Bab el-Mandeb corridor. Table 4 collects the principal logistics indicators.

Indicator	Change
War-risk insurance premium (% of hull value)	Rose from ~0.2% to ~1% (about fivefold)
Vessels stranded / awaiting transit	~150 vessels
Strait of Hormuz transit volumes	Sharp decline; closure followed by selective, supervised passage
Tanker freight rates (VLCC, Gulf routes)	Reported increase for Gulf routes; qualitative evidence only
Average voyage length / transit time (rerouted cargo)	Increased where rerouting occurred; qualitative evidence only

Table 4. Shipping and Logistics
Note: Insurance-premium, stranded-vessel, and transit entries follow the narrative in Section 2. Freight-rate and voyage-time entries are qualitative and are not used as point estimates in the analysis.

These frictionsmatter macroeconomically because they raise the effective cost of delivered energy and goods independently of the wellhead price, lengthen supply chains, and feed into measures of global supply-chain pressure that proved inflationary during earlier disruptions. The combination of a higher commodity price and a higher cost of moving it is precisely what distinguishes a chokepoint shock from a pure production shock, and it is the mechanism through which a regional maritime conflict raises costs for firms and households far from the Gulf [34], [1].

5. Cross-Country Exposure and Heterogeneity

The macroeconomic incidence of a Strait of Hormuz shock is highly uneven, reflecting differences in energy import dependence, the commodity composition of imports, fiscal and external buffers, and exposure to the affected trade routes. Because around 80% of oil transiting the Strait is destined for Asia, and because China and India together account for a large share of those crude flows, the Asian importers are the most directly exposed on the volume dimension. Advanced importers such as Japan and South Korea, which rely heavily on seaborne crude and on Gulf LNG, face acute exposure despite their wealth, while Europe's comparatively small share of Hormuz crude flows leaves it more exposed through price and gas-market spillovers than through direct volumes. Net energy exporters present a different profile: Gulf producers face the paradox of higher prices alongside a disrupted ability to ship, while the United States, as a net exporter, is partially insulated on the external account even as domestic prices rise with the global benchmark. Table 5 sketches this heterogeneity.

Country / Group	Energy Dependence	Qualitative vulnerability assessment
China	High net dependence on oil and LNG imports; large share of Hormuz crude	High; cushioned by strategic reserves, diversified suppliers, and a managed exchange-rate regime
India	High import dependence; significant Hormuz crude	High; exposed via import bill, currency pressure, and fuel-subsidy costs
Japan	Very high import dependence; heavy Gulf crude and LNG	High; limited domestic production; reliant on reserves and diversification
South Korea	Very high import dependence	High; energy-intensive exporter exposed to margin compression
European Union	Modest direct Hormuz crude share; high gas-price sensitivity	Moderate to high; transmitted mainly through gas spreads and inflation
United States	Net energy exporter overall; still exposed to global oil-price pass-through	Moderate; domestic prices rise with global benchmarks, but the external account is cushioned
Gulf exporters (Saudi Arabia, UAE, Qatar)	Net exporters; highly shipping-exposed	Mixed; revenue gains from higher prices offset by disrupted delivery and fiscal volatility
Energy-importing emerging markets (e.g., Turkey, Pakistan)	High dependence; thin reserves and buffers	Severe; simultaneous terms-of-trade and financing shock

Table 5. Country Exposure

Note: Classification is qualitative and based on structural net-energy-import dependence and exposure to Hormuz flows [30], [37]. The vulnerability labels are analytical judgments combining exposure with available macroeconomic buffers; they are not estimated outcome effects.

Two cross-cuttingpoints sharpen the picture. First, the relevant measure of vulnerability is net energy import dependence combined with the share of those imports exposed to the chokepoint, not headline import volumes alone; a country sourcing little from the Gulf can still be badly hurt through global price transmission. Second, buffers matter: countries with strategic reserves, diversified suppliers, flexible exchange rates, and credible monetary frameworks can absorb the shock with less damage than those with thin reserves, pegged currencies, or fragile inflation expectations [1], [34]. The result is that an identical global shock produces very different national outcomes, with the heaviest burden often falling on energy-importing emerging economies that combine high exposure with limited buffers.

6. From Oil Shock to Stagflationary Risk: Mechanism Integration

6.1 Why This Is Not a Standard Oil Shock

The macroeconomic literature distinguishes oil-price increases by their source, because the source determines their effects. Kilian [5] shows that price rises driven by global demand have very different macroeconomic consequences from those driven by supply disruption, and that conflating the two yields unstable and misleading relationships between oil prices and activity. A chokepoint shock is predominantly an adverse supply and logistics shock, amplified by precautionary demand and geopolitical-risk expectations: prices rise not because the world economy is booming but because the secure delivery of energy is in doubt. This places the 2026 episode closer to the classic contractionary supply shocks studied by Hamilton [4] than to the benign, demand-driven increases of the mid-2000s analyzed by Blanchard and Galí [6]. Three features set it apart. First, it is simultaneously a price shock and a cost-of-transport shock: the same event that raises the commodity price raises the cost and risk of moving it, so the effective shock to delivered energy costs exceeds the move in the benchmark price (Section 4.3). Second, it bundles oil with LNG and refined products, hitting multiple energy markets and their downstream users at once, as the divergence in the Henry Hub-JKM and Henry Hub-TTF gas spreads in Table 2 makes clear. Third, it embeds a geopolitical risk premium that persists as long as the chokepoint is perceived as insecure, decoupling economic normalization from military de-escalation [3]. For these reasons, the shock is broader, more persistent, and more likely to generate stagflationary pressures than a crude-oil price movement considered in isolation.

6.2 Persistent Inflation Dynamics

The stagflationary risk of the shock arises from the potential combination of rising costs and falling real activity. Higher delivered energy costs raise firms' marginal costs and households' cost of living, compressing real incomes and demand even as they push up prices-the textbook supply-shock configuration in which inflation and unemployment can rise together. Whether the inflationary impulse is transitory or persistent depends on two factors: the durability of the

chokepoint risk premium and the anchoring of inflation expectations. If the risk premium is durable, because markets judge the Strait to remain insecure, then energy costs stay elevated, and the first-round price increase has more time to pass into core prices through transport, manufacturing, and services. If, in addition, inflation expectations de-anchor, second-round effects through wages and price-setting can entrench inflation well after the initial energy impulse fades [6], [41]. The persistence of war-risk premia and below-normal transit after the April ceasefire (Section 2) is precisely the kind of durable wedge that keeps delivered costs high; the breadth of the shock across oil, gas, and shipping spreads it across the consumption basket; and the elevation of geopolitical risk sustains precautionary behavior. The policy challenge, taken up next, is to prevent a persistent but ultimately supply-side shock from becoming embedded in expectations.

7. Policy Implications

7.1 Central Banks

The central-bank problem is the supply-shock dilemma in acute form: an energy-driven rise in inflation coincides with weakening activity, so stabilizing prices and supporting output pull in opposite directions [5], [6]. The evidence that much of the historical damage from oil shocks ran through the policy response itself [41] carries a clear lesson: the appropriate reaction depends on the credibility of the nominal anchor. Where expectations are well anchored, central banks can look through first-round energy effects and avoid overtightening into a supply- induced slowdown; where anchoring is fragile, they must lean against second-round effects even at a larger output cost. Clear communication that distinguishes the temporary, relative-price component of the shock from any persistent, broad-based component is therefore a central policy instrument.

7.2 Strategic Reserves

Coordinated releases from strategic petroleum reserves, organized through the International Energy Agency, can smooth the price path and counter precautionary hoarding, dampening the expectations channel [30]. Their effectiveness is bounded: reserves are finite and best suited to bridging temporary disruptions, so against a persistent chokepoint shock they can buy adjustment time but cannot substitute for restoring secure and reliable supply. The episode argues for maintaining adequate reserve coverage, pre-agreed coordination mechanisms, and, because the shock spans gas as well as oil, consideration of strategic gas storage and demand-management measures for LNG-dependent economies.

7.3 Supply Chain Diversification

Reliance on a single maritime passage for roughly one-quarter of global seaborne oil trade and about one-fifth of global LNG trade is a systemic vulnerability that no amount of ex-post policy can fully offset once a shock hits. Mitigation lies in diversification before the fact: expanding bypass pipeline capacity (such as the Saudi East-West and UAE ADCOP routes), diversifying suppliers and import routes, investing in flexible regasification and storage, and accelerating the deployment of domestic and low-carbon energy that reduces aggregate exposure to chokepoint risk [37], [30], [34]. Because alternative routes are currently insufficient to offset a full closure, the policy payoff to building redundancy ahead of the next shock is large.

8. Conclusion

This paper has treated the 2026 Iran-Israel-U.S. war as a global macroeconomic shock rather than a contained regional conflict, organized around the Strait of Hormuz as a critical maritime chokepoint. The central argument is that a chokepoint shock transmits through five interacting channels-energy supply; maritime and logistics frictions; financial and sanctions fragmentation; expectations and geopolitical risk; and the policy response-and that this multi-channel structure makes it broader, more persistent, and more likely to generate stagflationary pressures than an oil-price movement considered in isolation. The stylized evidence assembled here-sharp increases in crude benchmarks and regional gas spreads, qualitative risk-off patterns across financial markets, and elevated war-risk premia alongside curtailed transit-is consistent with this interpretation. Because inflation, output, and employment responses are not estimated here, the episode is characterized as a stagflationary risk rather than a documented stagflationary outcome. The chokepoint perspective also offers a generalizable framework. The same logic, in which the security and cost of moving a critical good, rather than its production alone, determine the size of the shock, applies to other strategic passages such as the Strait of Malacca, the Bab el-Mandeb, the Suez Canal, and the Panama Canal, and to non-energy goods that depend on them. Two implications follow. For analysts, the relevant exposure measure is not import volume alone but net dependence combined with chokepoint exposure and the durability of the associated risk premium. For policymakers, resilience can be built before the shock, through diversification and redundancy, far more cheaply than it can be restored afterward. The 2026 episode is best read not as a singular event but as a stress test of a global economy whose energy lifelines still pass through a small number of contested waterways.

References

[1] International Monetary Fund. (2026, April 14). World Economic Outlook, April 2026: Global Economy in the Shadow of War. Washington, DC: International Monetary Fund. https://www.imf.org/en/publications/weo/issues/2026/04/14/world-economic-outlook-april-2026

[2] Aguilar, P., Boeckelmann, L., & Kornprobst, A. (2026, July 2). Beyond energy prices: The ripple effects of Gulf supply disruptions. ECB Blog, European Central Bank. https://www.ecb.europa.eu/press/blog/date/2026/html/ecb.blog20260702~a6cf657ddc.en.html

[3] Caldara, D., & Iacoviello, M. (2022). Measuring geopolitical risk. American Economic Review, 112(4), 1194–1225.

[4] Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. Journal of Political Economy, 91(2), 228–248.

[5] Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. American Economic Review, 99(3), 1053–1069.

[6] Blanchard, O. J., & Galí, J. (2010). The macroeconomic effects of oil price shocks: Why are the 2000s so different from the 1970s? In J. Galí & M. J. Gertler (Eds.), International dimensions of monetary policy (pp. 373–421). University of Chicago Press.

[7] Khudaykulova, M., Yuanqiong, H., & Khudaykulov, A. (2022). Economic consequences and implications of the Ukraine–Russia war. International Journal of Management Science and Business Administration, 8(4), 44–52.

[8] Obrenovic, B., Oblakovic, G., & Asa, A. R. (2024). Bibliometric analysis of financial and economic implications during the COVID-19 pandemic crisis. Sustainability, 16(7), 2897.

[9] U.S. Department of Defense. (2026). Operation Epic Fury: Official statements.

[10] Reuters. (2026a, February 28). United States and Israel launch strikes against Iran. Reuters.

[11] Israel Defense Forces. (2026). Operation Roaring Lion: Official statements.

[12] The White House. (2026). Statement on military action against Iran.

[13] Reuters. (2026b, March 1). Reports indicate Iran's Supreme Leader killed in opening strikes. Reuters.

[14] Reuters. (2026c, March 2). Senior Iranian officials reported killed. Reuters.

[15] Reuters. (2026d, March 2). Iran launches retaliatory strikes on regional U.S. bases. Reuters.

[16] Reuters. (2026e, March 3). Gulf states report missile and drone attacks. Reuters.

[17] Reuters. (2026f, March 3). Attacks reported on U.S. installations across the Gulf. Reuters.

[18] Reuters. (2026g, March 4). Regional escalation continues across the Middle East. Reuters.

[19] Reuters. (2026h, March 4). Further Iranian strikes reported on Gulf targets. Reuters.

[20] Reuters. (2026i, March 5). NATO forces intercept Iranian missiles aimed at Turkey. Reuters.

[21] USNI News. (2026, March 2). Maritime security developments in the Gulf. U.S. Naval Institute News.

[22] Oman Maritime Security Centre. (2026, March 2). Statement on the attack on the tanker MKD VYOM [Press statement; reported via Anadolu Agency].

[23] V.Ships Asia. (2026, March 2). Statement on the attack on the MKD VYOM [Press statement].

[24] The Maritime Executive. (2026, March 2). Strait of Hormuz declared closed by Iranian forces. The Maritime Executive.

[25] gCaptain. (2026, March 2). IRGC confirms closure of the Strait of Hormuz. gCaptain.

[26] FreightWaves. (2026, March 2). U.S.-flagged tanker Stena Imperative struck at Port of Bahrain. FreightWaves.

[27] Financial Times. (2026, March 4). IRGC claims complete control of the Strait of Hormuz. Financial Times.

[28] Time. (2026, April 7). Iran expands attacks on Gulf shipping. Time.

[29] World Economic Forum. (2026, April 9). How war in the Middle East is turning governments into insurers of last resort.

[30] International Energy Agency. (2026, February). Strait of Hormuz [Factsheet]. https://www.iea.org/about/oil-security-and-emergency-response/strait-of-hormuz

[31] Downs, E. (2026, March 4). Implications of the Conflict in the Middle East for China's Energy Security. Center on Global Energy Policy, Columbia University SIPA. https://www.energypolicy.columbia.edu/implications-of-the-conflict-in-the-middle-east-for-chinas-energy-security/

[32] International Energy Agency. (2026, March 12). Oil Market Report - March 2026. https://www.iea.org/reports/oil-market-report-march-2026

[33] European Central Bank. (2022). Economic Bulletin (Issue 4/2022).

[34] Organisation for Economic Co-operation and Development. (2026, June). OECD Economic Outlook, Volume 2026 Issue 1. OECD Publishing. https://www.oecd.org/en/publications/oecd-economic-outlook-volume-2026-issue-1_2d1956f0-en.html

[35] Asa, A. R., & Nautwima, J. P. (2025). Factors influencing natural gas and LNG market prices in the global energy economy: A comprehensive review. International Journal of Innovation and Economic Development, 11(1), 21–33.

[36] Ikechi, K. S., & Anthony, N. (2020). Global oil price shocks and effects on economic growth: An econometric investigation of Nigeria. International Journal of Innovation and Economic Development, 6(4), 7–26.

[37] U.S. Energy Information Administration. (2025, June 16). Amid regional conflict, the Strait of Hormuz remains critical oil chokepoint. Today in Energy. https://www.eia.gov/todayinenergy/detail.php?id=65504

[38] Nautwima, J. P., & Asa, A. R. (2025). Reimagining logistics for sustainability: Green logistics practices, circular economy integration, and digital enablers in sustainable supply chain management. International Journal of Management Science and Business Administration, 11(5), 7–21.

[39] Obrenovic, B., Godinic, D., & Njavro, M. (2024). Sustaining company performance during the war-induced crisis using sourcing capability and substitute input. Environment, Development and Sustainability, 26, 30001–30026.

[40] Wairimu, D. M., Chege, D., Muli, S., & Ndolo, J. (2022). Moderating effect of supply chain configuration on the relationship between dynamic supply chain capabilities and resilience in the Kenyan retail sector in Kenya. International Journal of Management Science and Business Administration, 8(5), 66–74.

[41] Bernanke, B. S., Gertler, M., & Watson, M. W. (1997). Systematic monetary policy and the effects of oil price shocks. Brookings Papers on Economic Activity, 1997(1), 91–157.

[42] Alsharif, H. Z. H., Shu, T., Obrenovic, B., Godinic, D., Alhujailli, A., & Abdullaev, A. M. (2021). Impact of entrepreneurial leadership and bricolage on job security and sustainable economic performance: An empirical study of Croatian companies during COVID-19 pandemic. Sustainability, 13(21), 11958.

[43] Faulks, B., Song, Y., Waiganjo, M., Obrenovic, B., & Godinic, D. (2021). Impact of empowering leadership, innovative work, and organizational learning readiness on sustainable economic performance: An empirical study of companies in Russia during the COVID-19 pandemic. Sustainability, 13(22), 12465.

[44] Obrenovic, B., Du, J., Godinic, D., Tsoy, D., Khan, M. A. S., & Jakhongirov, I. (2020). Sustaining enterprise operations and productivity during the COVID-19 pandemic: "Enterprise Effectiveness and Sustainability Model." Sustainability, 12(15), 5981.

[45] Waiganjo, M., Godinic, D., & Obrenovic, B. (2021). Strategic planning and sustainable innovation during the COVID- 19 pandemic: A literature review. International Journal of Innovation and Economic Development, 7(5), 52–59.

[46] Oji-Okoro, I., Akadiri, S. S., Egila, J. C., & John, G. (2025). Modelling the relationship between oil price and stock markets in net oil-exporting and net oil-importing countries: A panel data approach. International Journal of Management Science and Business Administration, 12(1), 15–24.

THE USE OF AI TOOLS

The author used ChatGPT (OpenAI) to proofread the manuscript. All content was reviewed and verified by the author.

Index

A
Aleksić-Maslać, Karmela 18, 62, 119
Alkhaffaf, Maha 11, 135
Antoliš, Krunoslav 74
Apfenthaler, Edda 33
Auruskeviciene, Vilte 14

B
Barac, Zoran 33
Bartels, Niels 214
Biočina, Zdravka 189
Boyd, Brian 173

C
Camarero-Grados, Loreto 149
Capogna, Stefania 227

Č
Čevizović, Ivan 157
Čevizović, Tonka 157

D
Dedaj, But 50
Dillerup, Ralf 173
Dittes, Sven 173
Dogan, Ivan David 51

G
Gallucci Calabrese, Carlo Maria 8
Gorsky, Mikael 16, 19
Grinewitschus, Viktor 214

H
Hijden, Peter van der 1, 6

I
Ivaniček, Matej 270
Ivanković, Marija 120

K
Krkač, Kristijan 9

L
Laavunen, Thao 105
Langer, Yannik 255
Lanquillon, Carsten 200, 242, 255
Ledesma-Casas, Marina 149
Levin, Ilya 19
López-del-Hoyo, Yolanda 149
Löwe, Seline 255

M
Mąkosa, Maciej 90
Martinović, Maja 33
Matsekh-Ukrayinskyy, Lyubomyr 90
McAuliffe, Eileen 3, 7
Meckmann, Felix 214

N
Nakić, Matej 51
Natalini, Alessandra 227
Novak, Andrej 12

Nj
Njavro, Marin 15
Njavro, Mato 5

O
Oblaković, Goran 51, 172, 226
Ochs, Tom 255
Ottenbacher, Elias 255

P
Pottgiesser, Uta 214

R
Raueiser, Markus 10
Re, Maria Rosaria 227
Remenarić, Branka 157
Rischert, Enes 214
Roos, Johan 2, 6
Roud, Vitaliy 63
Rudko, Andrzej 32

ISSN 2991-9371

S
Seebach, Swen 13
Sońta, Monika 11, 90
Steidel, Isabell 173
Svintsitski, Yaraslau 270

T
Tomašković, Josip 282

Y
Yasin, Riem 200, 242
Yasin, Samir 242

Z
Zürn, Joshua 255

Ž
Žnidar, Sunčica 189

AACSB
Business Education.
Connected.

Join the Business Education Alliance

The world's largest network of business schools and organizations committed to advancing business education and making a lasting, positive impact on society.



aacsb.edu/membership

ISSN 2991-9371