



AI in Higher Education Summit

Paris - 17-18 March 2026

Edited by: Louis-David BENYAYER

AI in Higher Education Summit Paris - 17 - 18 March 2026

Edited by: Louis-David BENYAYER

Cite as Author(s), (2026), Paper title. In Louis-David Benyayer (Ed.), "Proceedings of the AI in Higher Education Summit 2026". Paris: ESCP Business School / ESCPTech Institute. Available on SSRN.

Copyright for the paper is held by the individual authors.



ORGANISING TEAM



Louis-David Benyayer
Conference chair



Alara Tascioglu
Parallel sessions chair

Session chairs: Lorena Blasco-Arcas, Gorgi Krlev, Erwan Lamy, Hsin-Hsuan Lee, Dan Levy, Vitor Lima, Marie-Catherine Mignault, Franjo Mlinaric, Miguel Palacios, Audrey Picard.

ABOUT **ESCP BUSINESS SCHOOL**

ESCP Business School was founded in 1819. The School has chosen to teach responsible leadership, open to the world and based on European multiculturalism. Six campuses in Berlin, London, Madrid, Paris, Turin and Warsaw are the stepping stones that allow students to experience this European approach to management. Several generations of entrepreneurs and managers were thus trained in the firm belief that the business world may feed society in a positive way.

This conviction and ESCP's values - excellence, singularity, creativity and plurality - daily guide our mission and build its pedagogical vision.

Every year, ESCP welcomes 10,000+ students and 6,000 managers from 135 different nationalities. Its strength lies in its many business training programmes, both general and specialised (Bachelor, Master, MBA, Executive MBA, PhD and Executive Education), all of which include a multi-campus experience.

It all starts here.



www.escp.eu

Follow us on X: [@ESCP_BS](https://twitter.com/ESCP_BS)

FOREWORD



Professor Léon Laulusa

Dean and Executive President, ESCP

From the Industrial Revolution to the rise of the internet, from the construction of the European Union to today's global challenges, ESCP has consistently accompanied the major transformations shaping our societies. In the 1960s, our School was among the first to launch a language laboratory using audio and video for immersive learning. As early as 1992, it was already experimenting with online courses — at a time when the “internet” still felt like science fiction.

Today, artificial intelligence is profoundly reshaping how we teach, learn, assess and work. Building on this pioneering legacy, ESCP became in 2024 the first French higher education institution to collaborate with OpenAI, and in 2025, the world's first school to partner with Hugging Face, a leading open-source AI platform. Through these partnerships, ESCP embraces generative AI as a transformative force in pedagogy, research, administration, and above all, in fostering the creation of new, sustainable business models.

In this context, ESCP is proud to publish the official proceedings of the first AI in Higher Education Summit, held on our Paris campus on 17 and 18 March 2026. This volume brings together the contributions of more than 40 experts — professors, researchers, consultants, ed tech providers, higher education institution leaders — offering complementary perspectives on the transformations reshaping higher education. Their work explores three key themes: governance, ethics and responsible AI; tools and experimentation; and new pedagogies and faculty development.

This initiative reflects ESCP's raison d'être — to inspire and serve society — and its mission to educate accountable, bold and creative leaders who drive ambitious, human-centred and harmonious transformations in business and society.

But we intend to go further. ESCP's new strategic plan for 2026–2030, Bold and United, sets a clear ambition: to become the first European University of Management by 2030. It is built on three complementary pillars — ESCP Business School, our historic core; ESCP School of Technology (from 2027), dedicated to AI, Big data, Cybersecurity and Digital transformation (ABCD), under the leadership of Professor Cédric Denis-Rémis; and ESCP School of Governance (from 2029).

ESCP's conviction is clear: in an age of profound transformation, hybrid skills at the crossroads of management, technology and governance are essential. Our objective is to position ESCP as a responsible European leader, guided by the belief that technology is neither good nor bad, but defined by how we use it. We also aim to serve as a living laboratory for AI-powered education, helping leaders and organisations harness emerging technologies for personal and corporate growth.

On behalf of the Executive Committee, I warmly thank all the professors and participants who contributed to the Summit and to these proceedings, particularly Professor Louis-David Benyayer and his team, as well as Professor Régis Coeurderoy and Dr. Sonia Ben Slimane for coordinating this publication.

I wish you all a joyful, inspirational and insightful read — and, who knows, perhaps a few new questions that AI has not yet thought to ask.

ESCPTech Institute Commentary



Frédéric Fréry
Co-director of the
ESCPTech Institute



Héctor González-Jiménez
Co-director of the
ESCPTech Institute

The ESCPTech Institute plays a central role in advancing ESCP Business School's strategy to lead in technology research and responsible digital transformation. Specifically, the Institute aims to bridge the gap between research on emerging technologies and real-world impact, empowering organisations, policymakers, and academic leaders with insights that shape business and public policy. The Institute's work spans topics ranging from artificial intelligence to quantum computing, through to digital platforms.

As a proud sponsor of the AI in Higher Education Summit 2026, we contribute to strategic dialogue and collaborative action on how AI is shaping the educational sphere. This summit convenes global academic leaders, organisations innovators, and policymakers to address how higher education must adapt to, and shape, the opportunities and challenges presented by AI. By focusing on AI-enhanced pedagogy, governance, research practices, and institutional transformation, the Summit aligns closely with ESCPTech's mission to foster responsible, human-centred technological progress.

A key outcome of the Summit is a White Paper synthesizing emerging research and strategic insights on AI in higher education considering the application of AI for educational, research and administrative purposes. The ESCPTech Institute has sponsored and supported the development of this White Paper in close collaboration with the ESCP AI initiative and affiliated researchers, ensuring it reflects rigorous scholarship and practical relevance.

Lastly, the Summit also strengthens ESCPTech's broader ecosystem by fostering connections with industry partners in higher education and AI, including leaders from technology firms, educational platforms, and policy organisations. This engagement complements the Institute's work by integrating practitioner perspectives, accelerating the translation of insights into practice, and enhancing ESCP's role as a hub for innovation in AI and education.

EDITO



Régis Coeurderoy
Associate Dean of Research



Sonia Ben Slimane
Executive director of
ESCP Research Institute
of Management (ERIM)

Artificial intelligence is no longer a peripheral technological development for universities and business schools, it's now transforming teaching practices, assessment methods, research processes, and the standards that underpin academic integrity. The rapid diffusion of generative AI places higher education at a critical juncture, raising fundamental questions about authenticity and fairness in evaluation, the integrity of research and authorship, and, more profoundly, the preservation of critical thinking and intellectual autonomy in a context where knowledge is increasingly “co-produced” with machines.

The AI in Higher Education Summit 2026 was conceived in response to these challenges, with the active involvement of the ESCPTech Institute. It brings together academic leaders, researchers, accreditation bodies, policymakers, and industry representatives to establish a rigorous state of the art and to articulate shared orientations for action. The programme reflects the multidimensional scope of this transformation, encompassing governance and institutional strategy, pedagogical redesign, research integrity, sustainability, and digital sovereignty. At the heart of these discussions lies a central institutional responsibility. Beyond efficiency gains and technological adoption, universities must define the principles and frameworks that will guide AI integration while safeguarding academic standards. Particularly, when generative systems assist scientific production or mediate learning pathways, institutions are called upon to ensure transparency, epistemic rigor, ethical accountability, and fairness. The challenge is therefore not simply to deploy AI tools, but to ensure that their integration reinforces, rather than weakens, intellectual autonomy and scholarly credibility.

At ESCP, we have chosen to engage with this transformation strategically and collectively. The “*AI in Higher Education Summit*” is part of a broader institutional and strategic initiative decided by Pr **Léon Laulusa**, our Executive President and Dean, reported to Pr **Francesco Rattalino**, the Vice-President for Academic Affairs, and coordinated by Pr **Louis-David Benyayer**. Over the past year, this initiative has mobilized faculty, students, and staff through dedicated training programmes, research workshops, pedagogical experimentation, and structured reflection on governance and impact. It demonstrates ESCP’s commitment to addressing AI not as a passing trend, but as a defining dimension of the future of higher education.

This volume gathers the contributions presented at the Summit and outlines a collective research and action agenda. It reflects the emergence of an international academic community determined to ensure that AI strengthens the core values of higher education.

We extend our sincere gratitude to all contributors, keynote speakers, session chairs, reviewers, and partners whose intellectual generosity and engagement have made this Summit possible. We also warmly thank the co-editors of this edition, for ensuring the rigor, clarity, and coherence of this collective endeavour.

The future of universities in the AI era will be shaped not by technology alone, but by the standards we uphold and the institutional choices we make. Through this Summit and this publication, ESCP reaffirms its determination to lead this transformation with responsibility, academic excellence, and lasting impact.

CONTENTS

INTRODUCTION SESSION

The ABC Framework for AI-Ready Universities.....	1
<i>Louis-David Benyayer, Alara Tascioglu</i>	

Session A1: AI Governance & Institutional Strategy

Artificial Intelligence (AI) and French higher education - benchmark study on institutional strategies and curriculum development in management sciences	5
<i>Josip Marić</i>	

AI-First for Good as Strategic Advantage - Governing AI-Driven Transformation and IP Value Creation	6
<i>Franjo Mlinarič</i>	

AI in Business Education: See the Future Study 2026.....	18
<i>Oliver Matthews, Andrew Crisp</i>	

From Compliance to Care: A Multi-Level Case Analysis of AI Governance in US Higher Education.....	23
<i>Jason Johnston</i>	

Session A2: Responsible AI for Stakeholders

AI as Institutional Infrastructure: A Maturity Framework for Governance-Ready Admissions	27
<i>Stéphane Ureña</i>	

EUonAIR: Shaping European Leadership in Responsible AI Governance and Education.....	31
<i>Barbara Domaradzka</i>	

From Anxiety to Agency: A Structured Framework for AI-Integrated Academic Writing Instruction	34
<i>Katarzyna Truszkowska</i>	

Session A3: Assessment Trust, Ethics & Accountability

Navigating the 'Wicked Problem' of AI Assessment: A Case Study of an Exam Redesign	41
<i>Ørjan Flygt Landfald</i>	

Operationalizing Responsible AI in High-Stakes Educational Assessment: The Duolingo English Test Case Study.....	45
<i>Alina A. von Davier, Jill Burstein, Geoff LaFlair, Kevin Yancey, Masha Kostromitina</i>	

Rebuilding Trust in Assessment with Generative AI: Evidence Judgment through Mode-Switching Competence.....	49
<i>Hsin-Hsuan Meg Lee</i>	

The (Un)Ethical Double Standards of the "ChatGPT's Ember AI-Enhanced Student Assessment"	54
<i>Vitor Lima</i>	

Session A4: AI Risks, Sustainability & Emerging Concerns

Aligning tech and sustainability in Higher Education: protecting the environment, safeguarding values, and upholding critical thinking.....	59
<i>Gorgi Krlev, Laetitia Langlois, Gabrielle Tremblay</i>	

GenAI Adoption in Business Schools: An Interactive Qualitative Analysis in a Developing Economy Context	63
<i>Prakul Jain, Prakash Sai Lokachari</i>	
Human-Centered AI Pedagogy: World Language Education as a Scalable Model for Higher Education.....	71
<i>Jinai Sun</i>	
Session B1: AI-Powered Assessment & Feedback	
AI Academic Navigator: A Case Study of GPT-4 Powered Student Assessment in Higher Education.....	74
<i>Jan Saro</i>	
AI as a Teaching Coach: A Design Case Study of Transcript-Based Feedback for Faculty.....	81
<i>Dan Levy</i>	
Beyond time savings: The hidden pedagogical benefits of AI-assisted grading	86
<i>Alexandre Billard, Louis Delbosc</i>	
Educational Technology: The Dream Triplet for Implementing AI Innovation	89
<i>Emmanuelle Deglaire, Emmanuelle Houet, Alexandre Billiard</i>	
Understanding Linguistic Feature Contributions in AES Systems	99
<i>Maryam Abo-Tabik, Giulio Corsi, Alexandru Marcoci, Georgiana Thorpe Apreutesei, Yael Benn, Lyba Razzaq, Isla Fay, Deborah Talmi, Roni Tibon</i>	
Session B2: AI for Research & Scholarly Discovery	
AI's Invisible Pollution: Academic Deepfakes and Epistemic Pollution	106
<i>Aurélien Acquier, Jozef Cossey</i>	
AI-Powered Research Impact Assessment through ResearchImpact.ai. From Framework to Field-Specific Implementation	111
<i>Sebastian Haan, Wilfred Mijnhardt</i>	
How Not to Perish by Publishing: Future-Oriented Strategies for AI-Assisted Scientific Writing.....	118
<i>Michael Kopp</i>	
The Architect and the Machine: Towards a New Ethics of the Author-Supervisor in the Age of Generative AI.....	125
<i>Erwan Lamy</i>	
Session B3: Institutional AI Platforms at Scale	
From Relevance-Lag to Continuous Curriculum Intelligence The Readiness Index: A Governed AI-Enabled Operating Model for Student Readiness in Business Education	129
<i>Antoine Casanova-Mazet</i>	
How Blackboard is Advancing Trustworthy AI-Enhanced Teaching and Learning: A Case Study in Ethical Implementation	137
<i>Hervé Didiot-Cook, Richard Gibbons, Gemini 3 (AI Collaborator)</i>	
Keeping The Searcher in Control: Designing AI for Responsible Academic Search.....	141
<i>Jesper Solheim Johansen, Christian Headley</i>	
Session B4: Interactive & Experiential AI Learning	
Case Study TECgpt: Empowering Faculty to Build Learning Experiences with AI. Tecnológico de Monterrey	145

Irving Hidrogo Montemayor, Edrei Robles Chávez, Úrsula Saldívar Dávila, Luis Güemes Frese, Paola Martínez Chiñas, Verónica Pérez Aguirre, Manuel Terán Melgarejo

PyxiScience 153

Joachim Lebovits

Will a chatbot, trained on course-aligned materials and embedded in the learning management system improve students learning outcome? 158

Hilde Skjerve, Martin F. Strand, Anette Sørensen, Audun Beyer, Emil Hasle-Hoff, Lars Erik Braaum

Session C1: AI Literacy & Critical Thinking

AI-Gen-Case: A Transformative AI-Driven Framework for Case-Based Learning in Higher Education 163

Diego Ravenda

AI Situations: A Practice-Intensive and Modular Framework for AI Literacy, Curriculum and Assessment in Higher Education 170

Marc Gonnet, Arnaud Gorgeon

From AI literacy to quiteracy: Unplugging the classroom to learn about and not with AI.. 173

Heid Reed

Reclaiming the value of Human Capital in the AI era 178

Marjolaine Catil

Why Knowledge Is Required to Collaborate with AI 182

Michal Serkieza

Session C2: Redesigning Pedagogy in the Age of GenAI

AI Agents for Roleplaying in the Classroom: Design Principles, Pedagogical Value, and Responsible Integration in Higher Education 195

Lorena Blasco-Arcas

Building AI Resilience in Management Education 199

Thomas Lorenz

From Burden to Creativity: A Demonstration-Based Approach to AI Integration in Higher Education 211

Guadalupe Dorna

From Reading Alone to Reading-with-AI: How Undergraduates Reconfigure Academic Reading Practices in an AI World 215

Justin Olmanson, Williams Kwabena Boakye, Iliia Shcherbakov

Session C3: Faculty Development for AI Integration

The AI-Era Engineer: From Problem Solver to Problem Framer 221

Romain Buquet

The AI Paradox in Education: When Machines Master What Learning Is Meant to Develop 226

Daria Mizza

Transformative Teaching with AI: A Faculty Development Case Study for Higher Education 233

Claudia Carrone, Emmanuelle Houet

Session C4: Course Design & Student Agency

AI Literacy in Learning - From Training to Reflective Practice	239
<i>Boris Walbaum, Liudmila Piatnitckaia</i>	
AI Policy as Pedagogy: Guiding Student Learning with Generative AI	244
<i>Justin Olmanson, Minji Jeon, Azadeh Hassani</i>	
Transforming Passive Learning into Intelligent Learning through a GenAI-Enabled Tutor-Like Scaffolding Model	250
<i>Nazim Ali, Sarah A. Aynsley</i>	



The ABC Framework for AI-Ready Universities

Louis-David Benyayer,
ESCP Business School
Alara Tascioglu, ESCP
Business School

Abstract

This article introduces the ABC Framework, which formalizes three complementary imperatives for higher education institutions navigating AI: Accountable governance, Bold capability-building through disciplined experimentation, and Creative redesign of pedagogy and academic work around human judgment. The ABC Framework pillars are identified through the analysis of the contributions to the 2026 AI in Higher Education Summit and in alignment with ESCP mission. They serve as a conceptual lens, a leadership agenda, and a basis for practical diagnostics and roadmaps. This framework emerged from the thematic clustering of 50 contributions to the 2026 AI in Higher Education Summit at ESCP into the Accountable, Bold, and Creative tracks, and draws on the forthcoming ESCPTech Institute whitepaper, itself a contribution to the summit.

Keywords: Artificial Intelligence, Higher Education, Implementation

1. Why a Framework Now: Not “If,” But “How”

Universities face a non-binary reality: AI delivers real benefits and real risks simultaneously. Productivity gains (teaching preparation, feedback, research workflows, administrative automation) coexist with threats to assessment validity, epistemic quality, equity, privacy, and institutional legitimacy. The strategic question is therefore no longer whether to adopt AI, but how to implement it responsibly, at scale, while preserving educational purpose and trust.

At the same time, AI adoption is rarely a purely technical matter. Evidence from institutional practice and emerging scholarship suggests that the binding constraints are often organizational and sociotechnical: governance structures, capability development, access architecture, procurement and compliance, and the redesign of academic practices (especially assessment). From a sociotechnical-systems perspective (Trist & Bamforth, 1951), sustainable outcomes depend on jointly optimizing social and technical subsystems (Pasmore et al., 1982): leveraging what humans and AI each do best.

The ABC Framework responds to this challenge by articulating three imperatives that any AI-ready institution must address concurrently.

2. The ABC Framework: Three Institutional Imperatives

A — Accountable: AI as a Governed Sociotechnical System

Accountable institutions treat AI adoption as a governance and responsibility problem before it is a tool problem. They establish clear baselines for acceptable use, disclosure, data handling, and high-stakes decision-making, and they allocate accountability explicitly, especially in sensitive domains such as admissions and assessment. Broad stakeholder adoption is integral: governance is credible only when those affected by AI-enabled decisions have voice in shaping them (UNESCO, 2023).

Regulation and ethics are not “checkboxes”; they are design constraints that determine what universities can responsibly do and how they must do it. Rights-anchored and risk-tier approaches (e.g., high-risk classification in certain educational uses under the EU AI Act) (European Parliament and Council of the European Union, 2024) imply lifecycle obligations: documentation, oversight, contestability, and continuous evaluation, not just policy statements.

Core requirements

- Coherence and accountability baselines: institution-wide norms that constrain local variation while allowing pedagogical contextualization.
- Risk-tier governance for high-stakes domains: explicit human accountability, documented decision rationales, and oversight mechanisms.
- Participatory AI governance: structures that protect academic freedom while allocating responsibility through transparent, inclusive decision-making.
- Access and compliance architecture as both a governance and an equity instrument: vetted tooling, secure provisioning, live inventories of systems/use cases/data flows/vendors, and procurement aligned to risk classification, recognizing that unequal access to AI tools creates structurally unequal conditions for learning and assessment.

B — Bold: AI as Capability Expansion Through Disciplined Experimentation

Bold institutions build institutional capability through bounded pilots, rapid learning cycles, and scalable infrastructure, avoiding both paralysis and uncontrolled diffusion. Boldness is not techno-optimism; it is the capacity to experiment responsibly, evaluate rigorously, and scale only what passes agreed criteria.

AI is transforming the university “as a system,” requiring alignment of four interdependent components: Governance, Capabilities, Technology, Practices. No single component works in isolation; boldness is the institutional ability to move these elements together, sequentially and coherently.

Core requirements

- Build systemness through institution-wide AI strategy tied to operating-model consolidation: common architectures, interoperable data and process standards, and clear decision rights rather than isolated pilots.
- Make AI and data literacy compulsory for staff and students, staged by role and tied to real tasks: readiness varies and interpretive responsibility is widely distributed.
- Institutionalize continuous evaluation, futures tracking, and wellbeing-centered design as core operating cycles: efficiency metrics alone are insufficient; persistence and success depend on affordability, time constraints, and mental-health pressures.

Evidence-informed action tends to follow a staged progression (EDUCAUSE, 2024):

- Prerequisites (governance + data foundations + risk mapping),
- First 12 months (bounded pilots + baseline literacy + assessment redesign streams),
- Years 1–3 (scaling and integration into curriculum governance and operating models),
- Beyond 3 years (AI-by-design: lifecycle governance as default).

C — Creative: Redesigning Teaching and Academic Work Around Human Judgment

Creative institutions redesign pedagogy, assessment, and academic work to prioritize what remains distinctly human in an AI-saturated environment, especially judgment, problem framing, epistemic vigilance, and ethical responsibility. Creativity here is not “novelty”; it is purposeful educational redesign (Bearman & Ajjawi, 2023).

AI does not merely disrupt teaching; it reveals weaknesses in long-standing educational proxies (Bearman et al., 2023), particularly take-home, product-based written assessments that often measure fluency and format more than understanding, reasoning, or durable learning (Xia et al., 2024). When “answer production” becomes trivial, education must shift from evaluating outputs to cultivating and assessing process, reasoning, and judgment (Xia et al., 2024).

This is a return to fundamentals: teaching is not content delivery; it is the design of learning situations that develop competencies. AI can support personalization and feedback, but it can also encourage delegation, skill atrophy, and overconfidence in fluent but unreliable content. The key determinant is not the tool; it is the conditions of use, the design of learning activities, and the cultivation of intellectual virtues.

Core requirements

- Assessment redesign as integrity strategy: reduce dependence on detection; increase emphasis on process, oral defenses, iterative work, and transparent norms. This imperative extends to research integrity and scholarly communication, which require similar redesign to preserve epistemic trust.
- Capability formation for students: balance (i) disciplinary depth, (ii) technical literacy, (iii) critical/creative thinking, and (iv) socio-emotional skills.
- Faculty development for intentional integration: support educators to design AI-inclusive learning situations and to decide when AI supports learning vs. when it short-circuits it.

3. Using the Framework

The ABC Framework is intentionally non-substitutable: each pillar is necessary, none is sufficient. Accountable without Bold risks bureaucratic paralysis and shadow AI. Bold without Accountable risks uncontrolled diffusion, compliance failures, and legitimacy loss. Creative without Accountable and Bold risks local innovation that cannot scale or be trusted.

In practice, leadership is the capacity to hold all three mandates at once: standardize what must be standardized, experiment and scale with discipline, preserve educational purpose by redesigning what counts as learning and evidence of learning.

Pillar	Level	Core Question	Failure Mode if Missing
Accountable	Institutional design	How do we govern AI responsibly?	Bureaucratic paralysis, shadow AI, compliance exposure
Bold	Organizational capability	How do we build AI capacity at scale?	Uncontrolled diffusion, pilot fatigue, legitimacy loss
Creative	Pedagogical purpose	What does learning mean in an AI world?	Creativity theatre: local innovation that cannot scale or be trusted

The ABC Framework serves three functions:

1. A conceptual lens across governance, tooling, pedagogy, and institutional transformation.
2. A leadership agenda that translates evidence into stable institutional commitments (coherence, access architecture, capability and assessment reform).
3. A basis for action and diagnostics, enabling institutions to locate themselves, prioritize interventions, and avoid common failure modes (over-reliance on policy, tool-first rollout, or “creativity theater” without governance).

AI does not merely introduce new tools into higher education; it forces institutions to clarify what they stand for, how they govern, and what they seek to cultivate in students and scholars. The ABC Framework, Accountable, Bold, Creative, proposes that AI readiness is not a matter of technological sophistication alone, but of institutional coherence, disciplined capability-building, and a renewed commitment to forming human judgment in a world of machine-generated fluency. The work ahead is therefore not only operational but intellectual and ethical: to build governance systems that learn, infrastructures that enable equitable access, and pedagogies that prepare graduates to think with, question, and transcend the systems they increasingly rely upon.

References

- Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, 54(5), 1160–1173. <https://doi.org/10.1111/bjet.13337>
- Bearman, M., Nieminen, J. H., & Ajjawi, R. (2023). Designing assessment in a digital world: An organising framework. *Assessment & Evaluation in Higher Education*, 48(3), 291–304. <https://doi.org/10.1080/02602938.2022.2069674>
- EDUCAUSE. (2024). 2024 AI landscape study. EDUCAUSE.
- ESCP Tech Institute. (2026). AI in higher education: State of the art – Building AI ready universities: Evidence based priorities for 2025–2030. ESCP Business School.
- European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council (Artificial Intelligence Act). Official Journal of the European Union.
- Pasmore, W., Francis, C., Haldeman, J., & Shani, A. (1982). Sociotechnical systems: A North American reflection on empirical studies of the seventies. *Human Relations*, 35(12), 1179–1204. <https://doi.org/10.1177/001872678203501207>
- Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations*, 4(1), 3–38. <https://doi.org/10.1177/001872675100400101>
- UNESCO. (2023). Guidance for generative AI in education and research. UNESCO Publishing.
- Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21(1), Article 40. <https://doi.org/10.1186/s41239-024-00468-z>

Artificial Intelligence (AI) and French higher education - benchmark study on institutional strategies and curriculum development in management sciences

Josip Marić*

Abstract

Purpose: This study, through a benchmarking approach to systematically cross-compare institutional strategies, curriculum development, and contributions to management research, examines how Higher Education Institutions (HEIs) in France are integrating Artificial Intelligence (AI) into management education.

Design/methodology/approach: This research adopts a netnographic method, leveraging digital ethnography to observe and analyse online communications and digital artifacts across HEIs in France. Supplementary benchmarking draws on structured criteria - including AI policy articulation, pedagogical adoption, and curriculum development - to facilitate a systematic cross-comparison among leading HEIs in France. Data sources include official institutional communication, web information, published strategy documents, and academic outputs.

Findings: In France, the integration of AI into management education and institutional governance has incited academic and policy discourse, primarily regarding the pedagogical models, learning approaches, administrative operations, and research paradigms. French HEIs, renowned for their academic rigor and innovation, are actively experimenting with AI-driven solutions, however, our insights showcase there are numerous opportunities for further improvements in these experimentations. Particularly, the study proposes alignment of institutional strategies with actionable insights for decision-makers, educators, and researchers.

Originality/value: To author's best knowledge, this is one of the rare studies covering AI & French HEIs in management sciences through a systematic and holistic manner. As AI continues to disrupt and redefine the landscape of higher education, such systematic analyses are essential to foster innovation, ensure ethical governance of AI, and make advancements in the field of management research.

Keywords: AI, Higher Education Institutions (HEI), France, benchmark, management sciences.

* Associate Professor at EM Normandie Business School, e-mail: jmaric@em-normandie.fr

AI-First for Good as Strategic Advantage - Governing AI-Driven Transformation and IP Value Creation

Prof. KU Franjo Mlinarič,
Kozminski University

Abstract

Artificial intelligence is rapidly reshaping how universities create knowledge, manage intellectual assets, and translate academic excellence into societal and economic value. For higher education leaders, the strategic challenge is no longer whether to adopt AI, but how to govern it in ways that protect intellectual capital and strengthen institutional relevance.

This paper argues that AI should be treated as institutional infrastructure rather than a collection of isolated tools or experiments. Effective adoption requires integration with governance systems for knowledge assets, intellectual property, and research-to-impact pathways. Fragmented approaches risk IP leakage, loss of academic authority, and weakened differentiation.

Drawing on my experience at Kozminski University and the Kozminski Business Hub, the paper introduces the G.A.I.N.+IP (Governance of AI for Networks and Intellectual Property) framework. The model provides practical guidance for senior administrators navigating AI transformation while addressing regulatory pressures such as the European AI Act and evolving accreditation expectations.

Keywords: (max 5)

AI governance, IP stewardship, economic value, research-to-impact translation

Author note. The author used the AI tools to assist with drafting, language refinement, and citation formatting. The final content has been reviewed and validated by the author. The views expressed in this paper are solely those of the author and do not necessarily reflect the official position of Kozminski University or its authorities.

The Strategic Imperative: From Experimentation to Infrastructure

The End of Fragmented AI Adoption

Higher education is entering a phase where the traditional separation between teaching, research, and commercialization is increasingly untenable. AI accelerates this convergence by lowering the cost and effort required for routine content creation and standardized instruction, while simultaneously increasing the strategic importance of mentorship, synthesis, judgment, and original intellectual contribution (Shimabukuro, 2025).

As AI tools diffuse across teaching, research, and student support, institutions now face strong pressure to move beyond ad hoc experimentation and develop comprehensive, faculty-focused AI policies that balance innovation, academic integrity, and equitable access (Azevedo et al., 2025). This transition fundamentally alters the competitive logic of universities. As content becomes easier to replicate through generative AI, differentiation depends on the capacity to generate original knowledge, govern it responsibly, and translate it into real-world impact (Legatt, 2025).

Recent data confirms this strategic inflection point. The AACSB's January 2026 framework report documents evidence from 48 business schools across their global network, revealing that AI integration has moved from individual faculty initiatives to leadership-driven institutional change (AACSB, 2026a). Deans and senior leaders now play visible roles in setting priorities, allocating resources, and managing institutional risk—a clear signal that AI has become a strategic factor shaping educational quality, relevance, and competitiveness.

The Governance Mandate

For administration and strategy leaders, this shift implies a clear mandate: AI governance must be embedded in core institutional strategy, not delegated solely to IT departments, teaching support units, or individual faculty initiatives (EdTech Magazine, 2026). Institutions that treat AI merely as an efficiency tool risk becoming passive contributors of data and insight to external platforms, losing control over their most valuable asset—knowledge creation.

The European University Association's 2026 report *Adopting AI that Serves the Needs and Values of Universities* emphasizes that universities should approach AI through reflection and careful institutional adaptation rather than prohibition, guided by their core missions and values (Jørgensen & Phelan, 2026). In this context, the governance challenge extends beyond risk mitigation to strategic value creation. Universities must answer a fundamental question: **Who owns, governs, and benefits from AI-augmented knowledge creation?**

In environments where AI systems continuously improve through interaction, the absence of deliberate governance risks positioning universities as knowledge donors rather than knowledge owners. For business schools operating in competitive, knowledge-intensive ecosystems, this represents a structural disadvantage that directly impacts institutional sustainability, accreditation standing, and market positioning.

Regulatory and Accreditation Landscape: From Compliance to Competitive Advantage

The European AI Act and High-Risk Classification

The European AI Act, fully effective in 2026, introduces a risk-based approach to AI classification that directly impacts higher education institutions. Universities within the EU

are classified as high-risk AI system providers, particularly for AI applications in student assessment, admissions processes, and educational decision-making systems (KPMG Law, 2025).

This classification triggers substantial compliance obligations, including (EC, 2024):

- Development of comprehensive internal AI governance frameworks.
- Mandatory training programs for all staff and users.
- Establishment of interdisciplinary expert committees for responsible AI implementation.
- Documentation of decision-making processes, risk assessments, and bias mitigation strategies.
- Human oversight mechanisms for all significant AI-driven decisions.
- Transparency and accountability structures subject to external review.

However, forward-thinking institutions are reframing these regulatory requirements not as compliance burdens, but as opportunities to build strategic governance capabilities that strengthen institutional positioning. As governance expert frameworks emphasize, institutions that can provide comprehensive evidence trails—documenting who makes AI decisions, what uses are sanctioned, how risks are evaluated, and how policies adapt—will distinguish themselves in accreditation reviews and competitive assessments.

Triple-Crown Accreditation and AI Maturity Indicators

Triple-Crown accreditation standards (AACSB, AMBA, EQUIS) and global rankings increasingly reflect expectations related to AI across multiple dimensions: strategy, governance, ethics, innovation, and societal impact (Khan, 2024). While explicit AI metrics may still be evolving, peer-review processes already assess how institutions demonstrate:

- Strategic intentionality in AI adoption.
- Coherent governance and accountability structures.
- Protection and valorization of intellectual capital.
- Responsible and ethical innovation practices.
- Durable engagement with business and society.

In this context, AI governance has become a proxy for leadership maturity. The AACSB's 2026 framework explicitly states that "quality assurance and innovation are increasingly intertwined," with AI integration reshaping expectations related to curriculum relevance, faculty engagement, learner outcomes, and ethical responsibility (AACSB, 2026). Institutions that can articulate clear ownership models for AI-assisted outputs, align AI adoption with long-term knowledge-asset strategy, and demonstrate structured research-to-impact pathways signal readiness for sustained relevance rather than short-term experimentation.

Business schools using generative AI to streamline accreditation processes—including automated report drafting, data synthesis, and KPI dashboards—are already demonstrating practical integration while reducing manual effort and enabling faculty to focus on higher-value strategic work (Benrqa, Goel, & Fagnot, 2025). However, the true competitive

advantage lies not in operational efficiency, but in the capacity to govern AI as institutional infrastructure that amplifies knowledge creation and value capture.

Structural Limits of Prevailing AI Adoption Models

Three Dominant but Insufficient Approaches

Despite widespread experimentation, most universities continue to rely on fragmented AI adoption models that fail to address fundamental governance questions:

Faculty-Driven Initiatives: Individual professors and departments experiment with AI tools for teaching enhancement, research acceleration, or administrative efficiency. While these initiatives demonstrate innovation, they often lack institutional coordination, exposing universities to unprotected IP creation, inconsistent ethical practices, and missed opportunities for strategic knowledge asset development.

Compliance-Oriented Approaches: Many institutions respond to the AI Act and accreditation pressures by establishing policies focused on academic integrity, plagiarism prevention, and risk mitigation. These approaches address symptoms rather than strategy, positioning AI primarily as a threat to be managed rather than an asset to be leveraged.

Technology-Vendor-Led Deployments: Procurement-driven AI adoption, where universities license platforms for learning management, student services, or research support, risks externalizing value creation. Vendors capture data, learning loops, and emerging intellectual capital, while universities gain operational efficiency but lose strategic control over knowledge assets.

The Fundamental Governance Gap

These models share a critical limitation: they fail to answer who owns, governs, and benefits from AI-augmented knowledge creation. Recent analysis of university AI commercialization reveals a fundamental misalignment between AI's unique requirements and traditional software-based commercialization models.

When university AI research is forced into outdated commercialization frameworks, promising innovations rarely translate into real-world impact. The commercialization gap manifests through (Westgarth et al., 2022):

- Ambiguity around intellectual property rights for AI-assisted outputs.
- Unclear asset definitions (algorithm, model, data, or process).
- Reproducibility gaps between research environments and commercial applications.
- Licensing negotiations that stall due to insufficient governance clarity.
- Loss of institutional knowledge to external platforms and vendors.

Legal precedents confirm that clear assignment agreements are essential for university ownership of AI inventions. Cases like *Stanford University v. Roche Molecular Systems* demonstrate that without explicit IP assignment protocols, institutions risk losing valuable intellectual property even when innovations emerge from university-supported research (Holland & Knight, 2011). For business schools seeking to commercialize AI-enabled research and pedagogical innovations, this governance gap represents an existential strategic risk.

The G.A.I.N.+IP Framework: AI as Institutional Infrastructure

Framework Architecture

To overcome these structural limitations, we propose the G.A.I.N.+IP framework (Governance of AI for Networks and Intellectual Property), which integrates AI governance with institutional networks and IP stewardship. The framework treats AI, knowledge assets, and research-to-impact pathways as a single governance system rather than separate domains.

The G.A.I.N.+IP framework consists of four interconnected pillars:

Pillar 1: Strategic Governance

AI and IP governance are positioned at board and senior-leadership level, not delegated to middle management or technical staff. This pillar establishes:

- Clear ownership models for AI-assisted outputs across teaching, research, and innovation activities.
- Alignment with European regulatory frameworks, particularly the AI Act.
- Defined ethical boundaries consistent with institutional values.
- Integration with accreditation requirements and quality assurance processes.

By elevating AI from an operational issue to a strategic institutional asset, this approach directly supports accreditation expectations related to accountability, intentionality, and long-term value creation.

Pillar 2: Hub-Driven Execution

The framework ensures that AI-enabled ideas move systematically from experimentation to protected value through structured institutional mechanisms. At Kozminski, early-stage ideation is supported across the academic community through:

- Open innovation platforms accessible to faculty, students, and research staff.
- Structured triage conducted jointly by Career Office, Commercialization Office, and Kozminski Business Hub.
- Assessment of protection potential, data governance requirements, and research-to-impact pathways before public disclosure.
- IP awareness and market logic embedded from project outset, not retrofitted later.

This approach addresses the university AI commercialization gap by creating clear pathways from research to market that protect institutional knowledge assets while enabling societal impact.

Pillar 3: Faculty and Cultural Alignment

AI shifts academic roles away from repetitive content delivery toward mentorship, synthesis, and knowledge authorship. While incentive systems are still evolving, the institutional direction must be explicit: faculty are positioned as creators and stewards of high-value knowledge assets, not merely as content providers.

According to IMD's 2025 playbook (IMD, 2025), effective cultural transformation requires:

- Professional development programs that build AI literacy and governance awareness.
- Recognition and reward systems that value knowledge creation and IP development.
- Clear communication about how AI amplifies rather than replaces academic expertise.
- Explicit recognition of participation in commercialization activities as a legitimate component of academic contribution.

Over time, this alignment strengthens both academic legitimacy and the university's strategic asset base, creating a virtuous cycle of innovation and value creation.

Pillar 4: Global Networks and European Strategic Resources

The framework extends AI-enabled knowledge creation beyond campus boundaries by anchoring coordination locally while connecting to innovation ecosystems across Europe and the world. The European Commission's 2026 European Innovation Council work programme allocates €1.4 billion to support deep tech entrepreneurs and researchers, with simplified access to funding, investment channels, and commercial partners (EC, 2025).

By leveraging European research frameworks, regulatory instruments, and collaborative alliances, universities can:

- Access substantial funding for high-risk, high-reward AI innovation projects.
- Connect with the Scaleup Europe Fund for late-stage growth capital.
- Participate in Advanced Innovation Challenges inspired by the US ARPA model.
- Build partnerships through the European Corporate Network and EIC corporate days.

These resources provide a values-based, legally robust environment for protecting IP and supporting credible research-to-impact pathway from early stages, positioning European universities to compete globally while maintaining ethical standards.

Implementation Principles

The G.A.I.N.+IP framework operates according to three core implementation principles:

Integration over Fragmentation: Rather than establishing separate committees for AI strategy, IP management, and innovation commercialization, the framework creates unified governance structures that address these domains holistically. Cross-functional teams bring together academic leadership, legal expertise, technology transfer professionals, and business development specialists to make coordinated decisions.

Proactive over Reactive: Traditional IP management waits until after research is published or teaching materials are created to assess protection opportunities. The G.A.I.N.+IP approach embeds IP considerations into project planning, ensuring that governance decisions precede rather than follow knowledge creation.

Value Creation over Risk Mitigation: While compliance with the AI Act and accreditation standards is non-negotiable, the framework positions governance primarily as an enabler of strategic value rather than a constraint on innovation. By clarifying ownership,

establishing research-to-impact pathways, and building trust with external partners, governance becomes a competitive advantage.

Ethical AI, IP Protection, and Institutional Trust

Ethics as Governance Infrastructure

Ethical AI governance is inseparable from effective IP stewardship. Kozminski University was among the first universities in Poland to issue institutional recommendations on the use of AI-based generators such as ChatGPT, emphasizing honesty, transparency, independent student work, and critical awareness of the limitations and risks of these tools across teaching and research. These early guidelines function as living governance instruments, intended to evolve alongside technology and regulation and to support responsible, values-aligned AI use within the academic community (Kozminski University, 2023).

Building on this, ethical AI frameworks should also confront key questions at the intersection of AI and intellectual property:

- *Attribution:* How do we acknowledge human and AI contributions to knowledge creation?
- *Ownership:* What rights do faculty, students, and the institution hold over AI-assisted outputs?
- *Data Governance:* How do we protect privacy while enabling AI learning and improvement?
- *Bias Mitigation:* What mechanisms ensure AI systems do not perpetuate discrimination or inequity?
- *Transparency:* How do we document AI decision-making processes for accountability?

Alignment with European AI governance principles ensures that innovation is pursued with human oversight and respect for fundamental rights (Futurium European AI Alliance, 2025). For business schools, this ethical foundation is particularly critical given the potential for AI systems to influence high-stakes decisions about student admissions, academic progression, and career outcomes.

IP Protection in the AI Era

Traditional intellectual property frameworks were designed for human creativity and face significant challenges in addressing AI-generated outputs. Key legal principles that universities must navigate include:

Human Authorship Requirement: Most jurisdictions require human inventive contribution for IP protection. Fully autonomous AI inventions may not be patentable, making it essential to document human involvement in AI-assisted research and development.

Work-for-Hire Doctrine: University employment contracts typically assign IP rights to the institution for work performed using university resources. However, AI platforms introduce complexity—if faculty use external generative AI tools, who owns the resulting intellectual property?

Trade Secret Protection: AI models, training datasets, and proprietary algorithms may be protected as trade secrets rather than through patents or copyrights. This requires

universities to maintain confidentiality protocols before licensing or transferring technology.

By addressing these IP complexities proactively, the G.A.I.N.+IP framework enables universities to protect knowledge assets while maintaining the transparency and openness essential to academic culture.

Trust as Competitive Advantage

Commercialization is often viewed with suspicion in academic contexts, perceived as a deviation from scholarly values. The G.A.I.N.+IP framework explicitly reframes commercialization not as a departure from academic mission, but as its extension—transforming knowledge into societal benefit while preserving institutional trust.

For leading accreditation bodies, societal impact and stakeholder engagement have become core expectations rather than optional add-ons. Universities that can demonstrate ethical AI governance, robust IP protection, and responsible research-to-impact pathways position themselves as trustworthy partners for industry, government, and civil society—strengthening their competitive position in an increasingly interconnected innovation ecosystem.

Business Partners and Alumni as Value Multipliers

Collaborative Innovation Models

Systematic collaboration with business partners and alumni amplifies AI-driven value creation while distributing risks and accelerating validation cycles. The G.A.I.N.+IP framework incorporates external stakeholders early in defining problem statements, data access conditions, and success metrics for AI-related initiatives.

Effective collaboration requires structured governance that addresses:

- *IP Ownership*: Clear pre-project agreements defining who owns research outputs, AI models, and derivative innovations.
- *Data Governance*: Protocols for sharing, using, and protecting proprietary and personal data.
- *Publication Rights*: Balanced agreements that enable academic dissemination while protecting commercial interests.
- *Revenue Sharing*: Transparent models for distributing commercialization proceeds among university, faculty, and partners.

Corporate partners benefit from access to cutting-edge research, specialized talent, and academic credibility. Universities gain real-world problem access, validation environments, and potential research-to-impact pathways. When governed effectively, these relationships reinforce institutional knowledge positions rather than diluting them.

Alumni as Boundary Spanners

Alumni play a distinctive role in the G.A.I.N.+IP ecosystem, functioning as boundary spanners who facilitate access to real-world test environments and strengthen feedback loops between academia and practice. Business school alumni, particularly those in entrepreneurial or corporate innovation roles, provide:

- Market intelligence about emerging AI applications and business needs.
- Pilot environments for testing AI innovations in authentic commercial contexts.
- Mentorship for faculty and students navigating research-to-impact pathways.
- Investment capital and strategic partnerships for university spinoffs and ventures.

In many universities, alumni engagement is structured through formal advisory councils, investor networks, and corporate partnership programs. This systematic approach transforms alumni from passive supporters into active participants in knowledge creation and commercialization, creating mutually beneficial value exchanges that strengthen institutional competitiveness.

Transferability and Implementation Roadmap

Context-Specific Adaptation

While the G.A.I.N.+IP framework reflects a specific institutional context, its architecture is intentionally designed for transferability across diverse higher education settings. The framework can be adapted to variations in institutional size and resources, regulatory environments, academic profiles, and levels of commercialization maturity:

- Institutional size and resources (from small specialized schools to large comprehensive universities).
- Geographic location and regulatory environment (European AI Act compliance vs. other jurisdictions).
- Academic focus (business schools, technical universities, liberal arts institutions).
- Commercialization maturity (nascent innovation ecosystems vs. established technology transfer operations).

The core principles (strategic governance, hub-driven execution, cultural alignment, and network leverage) remain constant, while implementation mechanisms flex to local circumstances.

Phased Implementation Approach

A phased implementation helps institutions embed integrated AI governance with manageable risk and visible early wins.

In Phase 1 (months 1–6), institutions should map existing AI uses, IP policies, and governance structures, secure explicit senior leadership sponsorship, and align internal ethics policies with key regulatory and accreditation expectations.

In Phase 2 (months 6–12), the focus shifts to designing core governance instruments: clear IP ownership rules for AI-assisted outputs, a simple but robust triage process for AI-related initiatives, a small portfolio of pilot projects, and initial faculty development on AI literacy and governance.

In Phase 3 (months 12–24), institutions scale what works by extending the governance framework across teaching, research, and commercialization, formalizing partnerships with corporate and alumni networks, and systematically linking AI initiatives to European funding and deep tech ecosystems.

In Phase 4 (months 24+), governance structures are refined based on experience and regulatory evolution, institutions share their models with peers, and they begin to document measurable impacts on knowledge creation, IP protection, and research-to-impact translation as evidence of AI governance leadership.

Success Metrics and Evidence Building

To demonstrate governance effectiveness for accreditation and strategic assessment, institutions should establish evidence trails documenting:

- Number and value of AI-related IP assets protected (patents, copyrights, trade secrets).
- Research-to-impact outcomes (licensing agreements, spinoff ventures, revenue generation).
- Faculty and student engagement in AI governance processes.
- Compliance with regulatory requirements and absence of governance failures.
- External partnerships and network connections established.
- Recognition by accreditation bodies and peer institutions for AI governance leadership.

These metrics transform governance from an abstract policy commitment into tangible evidence of institutional capacity and strategic maturity.

Conclusion: Governing the Intersection for Strategic Advantage

AI is rapidly commoditizing standardized content while amplifying the value of original thinking, synthesis, and intersectional knowledge creation. The strategic challenge for higher education leadership lies not in technological adoption, but in governing the intersection of AI, knowledge assets, and societal impact.

As we gather at the AI in Higher Education Summit 2026, the evidence is clear: fragmented approaches to AI adoption create structural vulnerabilities that undermine institutional competitiveness, expose universities to IP leakage, and position academia as knowledge donors rather than knowledge owners. In contrast, integrated governance frameworks that treat AI as institutional infrastructure—deeply connected to IP stewardship, ethical oversight, and research-to-impact pathways—enable universities to evolve from content providers into orchestrators of innovation.

The G.A.I.N.+IP framework illustrates how universities can create governed environments where ideas mature into protected, trusted, and scalable knowledge assets. While context-specific, the model offers administration and strategy leaders a pathway beyond fragmented AI experimentation toward AI-first, values-driven institutional leadership.

Three strategic imperatives emerge for higher education leaders:

First, elevate AI governance to board and senior leadership level. AI is not a technical implementation challenge—it is a strategic governance imperative that determines institutional relevance, competitive positioning, and long-term sustainability.

Second, integrate AI governance with IP stewardship and research-to-impact pathways. Treating these domains separately creates structural risks and missed

opportunities. Unified governance enables systematic value creation while maintaining ethical standards and academic legitimacy.

Third, leverage European strategic resources and global networks. The European Innovation Council work programme, Horizon Europe, Erasmus+ alliances, and national recovery and resilience plans, together with emerging deep tech ecosystems provide unprecedented opportunities for universities willing to engage strategically.

By aligning AI governance with IP stewardship, ethical oversight, global networks, and European strategic resources, universities can secure sustainable relevance in the AI era—competing on knowledge creation and impact, rather than on content alone. ***The question facing higher education leaders is not whether this transformation will occur, but whether their institutions will lead or follow.***

For business schools specifically, the imperative is particularly acute. As AACSB emphasizes, AI integration now shapes expectations related to curriculum relevance, faculty engagement, learner outcomes, and ethical responsibility. Institutions that can demonstrate mature AI governance—evidenced through clear policies, systematic implementation, measurable outcomes, and alignment with accreditation standards—will distinguish themselves in an increasingly competitive global landscape.

The AI-first university of 2026 and beyond is not defined by technological sophistication, but by governance sophistication: the capacity to orchestrate knowledge creation, protect intellectual capital, maintain ethical integrity, and translate academic excellence into lasting societal and economic value. This is the strategic advantage that integrated AI governance frameworks like G.A.I.N.+IP enable—positioning universities not merely as adopters of AI, but as architects of the AI-augmented knowledge economy.

References

AACSB. (2026a). A framework for artificial intelligence in business education. AACSB Insights. <https://www.aacsb.edu/insights/reports/2026/a-framework-for-artificial-intelligence-in-business-education>

Azevedo, R., & Robles, L. M. (2025). Institutional policies on artificial intelligence in higher education: Frameworks and best practices for faculty. *New Directions for Adult and Continuing Education*, 2025. <https://doi.org/10.1002/ace.70013>

Benrqa, Y., Goel, L., & Fagnot, I. (2025). Six ways to integrate AI for business school accreditation. *Times Higher Education*. <https://www.timeshighereducation.com/campus/six-ways-integrate-ai-business-school-accreditation>

EC-European Commission. (2024). Regulation (EU) 2024/1689 on artificial intelligence (AI Act). *Official Journal of the European Union*.

EC-European Commission. (2025). EU €1.4 billion to support deep tech innovators (European Innovation Council work programme 2026). https://eic.ec.europa.eu/news/eu14-billion-support-deep-tech-innovators-2025-11-06_en

EdTech Magazine. (2026). An overview of AI governance in education. <https://edtechmagazine.com/higher/article/2026/02/overview-ai-governance-education-perfcon>

Futurium European AI Alliance. (2025). AI in higher education: Mapping key guidelines & recommendations. <https://futurium.ec.europa.eu/en/european-ai-alliance/community-content/ai-higher-education-mapping-key-guidelines-recommendations>

Holland & Knight. (2011). Universities and other federal contractors need assignments to comply with Bayh–Dole Act after Supreme Court ruling in Stanford v. Roche. Holland & Knight LLP. <https://www.hklaw.com/en/insights/publications/2011/06/universities-and-other-federal-contractors-need-as>

IMD. (2025). Winning with AI: The business leader’s guide to AI from strategy to execution – An IMD playbook. IMD Business School.

Jørgensen, T. E., & Phelan, C. (2026). Adopting AI that serves the needs and values of universities: Final report of the EUA Task-and-Finish Group on Artificial Intelligence. European University Association.

Khan, S. (2024). Brave new words: How AI will revolutionize education (and why that’s a good thing). Viking Press.

Kozminski University. (2023). Kozminski University recommendations regarding the use of AI-based generators such as ChatGPT. Kozminski University. <https://www.kozminski.edu.pl/en/news/kozminski-university-recommendations-regarding-use-ai-based-generators-such-chatgpt>

KPMG Law. (2025). AI Act: This applies to AI in universities and research. <https://kpmg-law.de/en/ai-act-this-applies-to-ai-in-universities-and-research/>

Legatt, A. (2025). 7 AI decisions that will define higher education in 2026. Forbes. <https://www.forbes.com/sites/avivalegatt/2025/12/26/7-ai-decisions-that-will-define-higher-education-in-2026/>

Shimabukuro, J. (2025). Ed tech in higher ed – Three issues for Jan. 2026: AI as a pillar of institutional strategy. Ed Tech Journal. <https://etcjournal.com/2026/01/01/ed-tech-in-higher-ed-three-issues-for-jan-2026-ai-as-a-pillar-of-institutional-strategy/>

Westgarth, T., Chen, W., Hay, G., & Heath, R. (2022). Understanding UK artificial intelligence R&D commercialisation and the role of standards. Department for Digital, Culture, Media and Sport; Office for Artificial Intelligence; Cambridge Econometrics.

AI in Business Education: See the Future Study 2026

Oliver Matthews, 4Uni
Solutions
Andrew Crisp,
CarringtonCrisp

Abstract

This paper draws on findings from the See the Future study (2026), a survey of 1,863 respondents across 40 countries including current students, teaching and research faculty, management and professional services staff, employers, and alumni. The study was conducted by CarringtonCrisp and 4Uni Solutions, and sponsored by EFMD and Full Fabric. The paper examines the current state of AI adoption in business education, focusing on three areas of concern: the gap in AI knowledge and confidence between stakeholder groups, the limited scope of institutional AI policy, and the disconnect between employer expectations and what universities are currently delivering. The findings show that while AI use is near-universal across all groups surveyed, expertise remains low and structured support — through training, dedicated time, and access to tools — is consistently identified as insufficient. Faculty report genuine benefits to their teaching and research, yet hold significant reservations about how students are engaging with AI. Employers are the most experienced users and the most critical of universities' response, with 58% believing institutions are not doing enough to develop AI skills in graduates and 77% expecting new recruits to arrive with practical AI experience. The paper argues that business schools are past the point of early adoption but have yet to establish the coherent frameworks needed to move forward effectively.

Keywords: Faculty AI readiness, institutional AI policy, employer-graduate skills gap, AI adoption drivers, student AI engagement

Introduction

The See the Future study, published in February 2026, draws on responses from 1,863 participants across 40 countries, collected through an online survey during December 2025 and January 2026. Respondents include current students, faculty engaged in teaching and research, faculty in management and professional services roles, employers, and alumni. The study was conducted by EFMD, CarringtonCrisp, 4uni-solutions, and Full Fabric. This paper focuses on the AI-related findings and their implications for business schools and the academics working within them.

The headline is straightforward: AI is now in use across every stakeholder group surveyed, yet the confidence to use it well is in short supply. At least 89% of respondents in every group report using AI tools in some capacity — in their studies, their work, or their private lives. Despite this near-universal adoption, only 6% of faculty and professional staff currently describe themselves as AI experts. Among students, that figure drops to 4% prior to starting their studies. Use is widespread; expertise is not.

The Knowledge Gap

One of the more striking findings concerns the gap in AI knowledge between different groups. Employers report the highest levels of confidence, with 66% claiming substantial or expert knowledge of AI. Among teaching and research faculty, that figure falls to 43%, and among management and administrative staff to 38%. Students sit at the bottom of this ranking, with only 31% reporting substantial knowledge.

This matters because the groups with the least confidence — students and faculty — are the ones most directly involved in the learning process. When 43% of students report believing they know more about AI than those teaching them, it points to a credibility problem that institutions cannot afford to ignore. Faculty are not unaware of this: the data shows that 44% of teaching staff agree, at least in part, that students have outpaced them in AI knowledge. Whether or not that perception is accurate, it shapes the classroom dynamic.

According to respondents, the remedy is not complicated. Faculty and staff consistently identify training and dedicated time to experiment as the most important factors in improving their use of AI, followed by access to tools. The ask is not unreasonable. It is also not being met at scale.

A divide within institutions

The study allows for a separation of two groups that are often discussed together: deans/faculty with significant management or operational responsibility on one hand, and university professional services staff on the other. When examined separately, some meaningful differences emerge that have practical implications for how institutions approach AI development internally.

On training, 72% of deans and faculty in management roles report having received AI training provided by their institution, compared to 63% of professional services staff. That gap of nearly ten percentage points is notable because professional services staff are frequently the people responsible for deploying AI in areas such as admissions, student communications, and marketing — precisely the functions where deployment is already most advanced. The staff doing the most AI-dependent work are the least likely to have been formally trained to do it.

Both groups report a high awareness of their own limitations. Across management faculty and professional services staff combined, 60% say they feel they are making reasonable use

of AI tools but could do considerably more, and only 14% describe themselves as expert users. That shared sense of unrealised potential is consistent across the data, but the training gap suggests that professional services staff are being asked to increase usage with less institutional support.

There is also a difference in how policy is perceived. Deans and faculty report higher coverage across almost every policy category, which likely reflects their proximity to where policy is written rather than evidence that the policies themselves treat the two groups equitably. Professional services staff, by contrast, appear to be operating in a space where policy has not fully caught up with practice.

AI Policy: Present but Incomplete

Seventy-three percent of university-based respondents report that their institution has an official AI policy in place. However, the coverage of those policies is uneven. Among management and professional services staff, 66% say their policy addresses student use of AI, 57% say it covers research, 55% teaching preparation, 51% administrative use, and 46% classroom involvement. Among teaching and research faculty, the numbers are consistently lower across every category.

The picture in the workplace is not much better. Only 52% of alumni are employed in organisations with formal rules about AI use at work, and just 53% have received any training from their employer on AI tools. The majority — 70% — report having taught themselves simply by downloading tools and experimenting. Only 3% took a course with a university or business school. That figure alone should give institutions cause for reflection.

Why Institutions Are Adopting AI

The motivations driving AI adoption at universities are worth examining, not least because they push back against some common assumptions. Among management and professional services staff, innovation and teaching are the most frequently cited drivers of adoption, each identified by 47% of respondents. Administrative productivity follows at 45%, with research at 42%. Cost savings rank seventh, cited by 23%, and staff shortages come last at 17%.

This counters a key fear that AI adoption in higher education is primarily a cost-cutting exercise, or a response to staffing pressures. It is being driven by a genuine desire to improve how institutions operate. That framing has implications for how AI initiatives are designed and communicated internally.

In student recruitment and admissions specifically, deployment is already well advanced. Sixty-three percent of institutions report using AI to automate communication with prospective students, 62% to personalise marketing, 54% to manage enquiries, and 45% to analyse application materials. These are not pilot projects. They are operational realities, and they are being managed largely by professional services staff who, as noted above, are less likely than their management colleagues to have received formal training.

The Impact on Teaching and Research

Faculty views on AI's impact on their own work are notably more positive than their views on student use. Sixty-two percent of teaching and research faculty agree that AI has improved their teaching, and 66% say it has improved their research. A further 64% believe AI can help them improve still further. These are significant majorities, and they suggest that

when faculty engage with AI on their own terms — in research, in preparing materials, in managing workloads — the experience tends to be constructive.

The picture shifts when the focus turns to students. Fifty-three percent of faculty believe students only use AI superficially, and 44% agree that students use it primarily to cheat. Forty-three percent say AI is dumbing down university education. The tool usage data adds a layer of texture here: ZeroGPT is used by 18% of management and professional services staff, and Turnitin by 23%. The coexistence of generative AI tools and AI detection tools within the same institutions points to a genuine tension that policy has not yet resolved. Institutions are simultaneously enabling AI use and attempting to police it, often without clear boundaries between the two.

Students themselves are not entirely dismissive of these concerns — 41% agree that AI is 'dumbing down' university study, and 43% think it is mostly used to cheat on assignments, though 30% disagree with that characterisation. At the same time, 74% of students say AI tools have improved their learning experience, and 69% agree it is essential they learn about AI to support their future careers. These are not contradictory positions. Students can simultaneously recognise that AI is being misused in some quarters and find genuine value in using it themselves. What the data reflects is an environment in which norms are still being established and the educational contract around AI is unresolved.

The Employer Perspective

Employers are the most confident users of AI in the study. Ninety-three percent report using AI tools, and 25% describe themselves as experts. AI is applied across business functions, most frequently in data analysis, human resources, and business development. Eighty-two percent of alumni agree that AI tools have improved their productivity at work.

Yet employers are also the group most critical of universities' response to AI. Fifty-eight percent believe universities are not doing enough to develop AI skills in graduates, and 77% expect new recruits to have experience with AI tools when they join. In practical terms, employers want universities to provide opportunities for students to learn common workplace AI tools — cited by 76% — followed by developing an understanding of ethical AI use at 60%, and basic AI literacy at 53%.

These expectations are not new. Employers have long asked universities to better align graduates with workplace realities. What is different here is the pace. The gap between what employers expect and what universities are currently delivering is visible in the data.

Technology Partnerships and the Road Ahead

Many institutions are pursuing collaborations with technology providers. Half of those in management and professional roles report working with Microsoft Copilot, while 46% each are working with Google and OpenAI. Anthropic is involved with 8% of institutions. More than one in five institutions report having deployed or piloted a custom-built AI tool, most commonly in student support services and research.

The direction of travel is clear even if the destination is not. Thirty-seven percent of management and professional services staff say it is too early to tell what the full impact of AI will be on their institution. That honest uncertainty runs through the study. Policies are being written that will need to be rewritten in the near future. Tools are being adopted without full understanding of their implications. Training is being requested but not always provided — and, as the data shows, not being provided equally across the different staff groups who need it.

What the data does not show is paralysis. Across every group, the dominant response to AI is engagement — sometimes cautious, sometimes enthusiastic, rarely indifferent. For academics in business schools, the findings present both a challenge and an opening. The demand for better AI education is real, the institutional appetite for progress is there, and the question of who leads that work and who gets supported in doing so is still very much open.

References

Crisp, A., Matthews, O. (2026). See the Future: The end of the beginning, <https://www.efmdglobal.org/resources/report-see-the-future/>

From Compliance to Care: A Multi-Level Case Analysis of AI Governance in US Higher Education

Jason Johnston, PhD
Executive Director of Online Learning & Course Production
University of Tennessee, Knoxville
jjp@utk.edu

Abstract

The rise of generative AI has pushed US universities to act quickly toward restrictive policies focused primarily on academic integrity and controlling risk. However, these compliance-heavy frameworks often ignore the human element, undermining equity and trust. This submission presents a multi-level case analysis contrasting typical university AI policies found across the US with a divergent, Ethics of Care based approach at the University of Tennessee, Knoxville. Utilizing a hybrid case study design, this paper examines the collaborative creation of UT's eight "human-centered" AI principles and their operationalization within a Digital Learning department. Findings suggest that governance based on an Ethics of Care not only reduces anxiety but fosters greater adoption and adaptation than standard top-down mandates, transforming compliance into a living, relational, and even innovative practice.

Keywords: Ethics of Care, AI Policy Creation, United States, Universities

1. The Human AI Governance Gap

The rise of generative AI (GenAI) use at universities has pushed administrators to act quickly in response. GenAI poses both risks and opportunities for higher education, but most initial policies in US universities aim toward control and restrictions, primarily addressing the risks (Oh & Sanfilippo, 2025). Many institutions focus their policies on academic integrity and misconduct rather than on support. However, GenAI governance should be built with adaptive approaches to ensure that policies keep pace with the rapid advances (Taeihagh, 2025). In addition, restrictive-heavy policies may not consider the human in the mix, resulting in policies that diminish academic freedom and student progress.

This case study will examine typical AI policies across US universities and compare them to a collaborative guideline approach at the University of Tennessee, which was operationalized within the department of Digital Learning inspired by the Ethics of Care framework (Gilligan, 1977; Monchinski, 2009; Noddings, 1988; Tronto, 1993). Ethics of Care is

primarily a relational way of moral thinking, considering how connectedness shapes ethics. Too often, policies fail to address the human element, ignoring equity concerns and undermining trust and transparency. AI ethical guidelines based on caring can be both adaptive to our faculty, staff, and students as well as effective for digital transformation. Instead of dogmatic rules enforced from upper administration, they can reorient our AI use “concerned with the relations in which all must live” (Noddings, 1988, p. 219).

2. Methodology and Framework

This submission utilizes a hybrid case study design starting with a broad landscape analysis of US higher education AI policies to establish a cross-case synthesis, followed by an in-depth single-case study of the University of Tennessee’s collaborative creation of ethical AI guidelines, and a department-level application of those guidelines using an Ethics of Care framework.

This hybridized design adopts Yin’s (2018) structural framework to define the case boundaries and notes replication across the selected institutions. However, the data analysis relies on Merriam’s (2009) heuristic approach to interpret the findings. This allows us to rigorously categorize the US policy landscape while interpreting the meaning in comparison to the specific case application using the Ethics of Care framework.

While this case study will be mostly descriptive, Yin (2018, p. 6) states that “*some of the best and most famous case studies have been explanatory.*” The final discussion of the case will explore how the final guidelines were formed and why they were strengthened by the Ethics of Care Framework.

3. Phase 1: The US Landscape Analysis

While a growing number of US universities have AI policies in place, they tend to be ambiguous, without practical applications, and lack centralized support (Oh & Sanfilippo, 2025). Policies tend to emphasize prohibitions on academic misconduct, without clearly identifying what that means. As student concerns grow, centralized services are faced with developing their own determinations. Adaptive policies and frameworks are needed as GenAI rapidly develops and increases in function (Humble, 2025), but flexible structures may communicate a lack of coherence or accountability for compliance, resulting in fragmented institutional environments. US policies are often written at faculty and staff, rather than with them, fostering resistance rather than change.

4. Phase 2: The University of Tennessee Case

The University of Tennessee (UT), Knoxville, is a public, land-grant university established in 1794. It has over 40,000 students, 900+ programs of study, and over 2000 instructional faculty (University of Tennessee, Knoxville, n.d.). UT has a shared governance model, giving faculty primary responsibility over the academic core of the institution through the faculty council (University of Tennessee, Knoxville Faculty Senate, n.d.).

Institutional Strategy: The Collaborative Creation of AI Guidelines

When the public version of the Generative AI (GenAI) ChatGPT 3.5 model was released in the fall of 2022, universities initially panicked over the fear of widespread cheating and academic dishonesty. As faculty voiced their concerns to the provost at the University of Tennessee, Knoxville, there was a moment of gathering resources and direction on what to do next. In the Spring of 2023, rather than issuing a top-down mandate from the General

Counsel or IT, the provost convened a number of task forces, including a “*Philosophy Committee*” composed of 13 faculty and staff from diverse disciplines. This charge came as a strategic shift away from strict policies to develop common principles that could address the current threat of GenAI, but might survive ongoing technological advancements with AI.

The adaptive and collaborative approach of the philosophy committee rejected static rules and instead co-authored eight principles that could be applied across departments. In short, the headings for these principles were: Intentional, Human-Centered, Inclusive, Open & Transparent, Critically Engaged, Adaptive, Contextual, and Prepared. Rather than a rigid checklist, these were intended to serve as foundations for further discussion in the various colleges and departments. In this way, the guidelines were shaped by the community, who were then actively taking part in the digital transformation, rather than just having the guidelines delivered to them.

Operational Application: The Digital Learning Dev Team

In November of 2023, the author led his Digital Learning “*Dev Team*” through a process of developing guidelines for AI in our day-to-day work, guided by the eight principles communicated by the Philosophy committee. The dev team consisted of a variety of staff who support faculty course development through instructional design, media development, and professional development offerings.

Our first collaborative “*town-hall*” style meeting discussed the principles and then brainstormed through potential guardrails, challenges, ideas, and next steps. Functional teams clashed during this meeting, citing various fears and needs. It was clear that the only way forward was to use an Ethics of Care framework toward a common language and guiding policy. One top-down mandate was clear from the provost’s office: The University will not ban GenAI, so we need to agree on how we will use it. This created a safe zone for discussion and experimentation. One key outcome was a discussion and guideline around “*work creep*,” a deep concern that staff would step outside of their expertise and use AI to do the jobs of other people on the team. A welcome by-product of this discussion was that our communication and transparency about the current use of AI increased and anxiety about navigating this new world of artificial intelligence decreased.

Asynchronous edits and comments on the common document allowed staff to think through their concerns over the following weeks. In February of 2024, the author led a second synchronous meeting to solidify the guidelines, and a consensus was established. We agreed to reconvene in the fall of 2024 to revisit, which we did for some additions and edits. We then set a date and revisited and revised in the fall of 2025 to accommodate the expanding technology and to include new staff in the process.

5. Discussion: From Compliance to Care

As we compare the more directive “*compliance*” AI policies found in many US universities to the more Ethics of Care model at UT, there are some differences in terms of content. However, beyond the main talking points, the process of collaborative development created a context that fostered more trust across campus and principles that were more easily applied to a variety of departmental contexts, like the Digital Learning Dev Team.

The collaborative development of ethical guidelines was not only more adaptive and effective than standard top-down policies, but adoption increased, and anxiety decreased because the team felt heard and cared for in the process. This supports the Ethics of Care

idea that “in a basic and crucial sense, each of us is a relationally defined entity and not a totally autonomous agent” (Noddings, 1988, p. 222). Instead of automatons crafting static rules for compliant university robots to follow, GenAI guidelines became living, breathing codes for our working relationships with each other. So, through this approach, creating guidelines for using AI helped the Digital Learning team at the University of Tennessee become more human in the process.

References

- Gilligan, C. (1977). In a different voice: Women's conceptions of self and of morality. *Harvard Educational Review*, 47(4), 481–517.
<https://doi.org/10.17763/haer.47.4.g6167429416hg5l6>
- Humble, N. (2025). Higher education AI policies: A document analysis of university guidelines. *European Journal of Education*, 60(3) 1-13.
<https://doi.org/10.1111/ejed.70214>
- Monchinski, T. (2009). *Education in hope: Critical pedagogies and the ethic of care* [Doctoral dissertation, City University of New York]. ProQuest Dissertations and Theses. (UMI No. 3378603)
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. Jossey-Bass.
- Noddings, N. (1988). An ethic of caring and its implications for instructional arrangements. *American Journal of Education*, 96(2), 215–230. <https://doi.org/10.1086/443894>
- Oh, S. H., & Sanfilippo, M. R. (2025). Responsible AI in academia: Policies and guidelines in US universities. *Information and Learning Sciences*, 126(9–10), 561–587.
<https://doi.org/10.1108/ILS-03-2025-0042>
- Taeihagh, A. (2025). Governance of generative AI. *Policy and Society*, 44(1), 1–22.
<https://doi.org/10.1093/polsoc/puaf001>
- Tronto, J. C. (1993). *Moral boundaries: A political argument for an ethic of care*. Routledge.
- University of Tennessee, Knoxville. (n.d.). About. <https://www.utk.edu/about>
- University of Tennessee, Knoxville Faculty Senate. (n.d.). Faculty council.
<https://senate.utk.edu/faculty-council/>
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.

AI as Institutional Infrastructure: A Maturity Framework for Governance-Ready Admissions

Stéphane Ureña

Founder & CEO, Semantikmatch

stephane@semantikmatch.com

Abstract

University admissions are among the most consequential AI-mediated decisions in public life, yet most institutions deploy AI at the level of isolated task automation- plagiarism checkers, scoring tools, chatbots- without addressing the governance, auditability, and regulatory obligations that such high-stakes processes demand. This paper proposes a three-level AI Maturity Framework for higher education administration, distinguishing Task Automation, Workflow Augmentation, and Institutional Infrastructure. Drawing on a deployment at École des Ponts Business School (500 applications, 98-99% criteria reliability, 10× speed improvement), we argue that only Level 3- governance-ready, regulation-compliant AI infrastructure- can meet the requirements of the EU AI Act while preserving institutional sovereignty. The paper contends that AI governance in admissions is not merely a compliance matter but a question of democratic legitimacy in access to opportunity.

Keywords: AI maturity framework, EU AI Act, algorithmic governance, admissions infrastructure, institutional sovereignty

1. Introduction

University admissions decisions shape life trajectories, institutional reputations, and societal equity. They are, by any measure, high-stakes. Yet across European higher education, AI adoption in admissions has proceeded largely without a coherent governance architecture- deployed piecemeal, evaluated informally, and rarely designed with regulatory foresight.

Three converging pressures make this unsustainable. Application volumes have grown dramatically as internationalisation accelerates, each file comprising transcripts, letters, certifications, and interviews spanning dozens of countries and languages. Generative AI has created new fraud vectors- from synthetic essays to deepfake video submissions- that conventional review cannot reliably detect. And the EU AI Act (Regulation 2024/1689) now classifies AI systems used to determine access to education as high-risk under Annex III, imposing binding obligations on providers and deployers for risk management, transparency, human oversight, and auditability.

This paper argues that responding to these pressures requires more than incremental automation. It requires a conceptual and architectural shift toward AI as institutional infrastructure- governed, auditable, and compliant by design. We propose a maturity framework, illustrate it through a deployment case study, and discuss the strategic and political implications for European universities.

2. A Three-Level AI Maturity Framework

To move beyond ad hoc AI adoption, we propose distinguishing three levels of AI maturity in university administration, each characterised by different architectural assumptions, governance capabilities, and regulatory readiness.

Level	Scope	Governance	Regulatory Readiness
Level 1 - Task Automation	Isolated tools (plagiarism detection, chatbots, scoring)	None or minimal; black-box outputs	Non-compliant under EU AI Act
Level 2 - Workflow Augmentation	Decision-support tools integrated into existing workflows	Partial; logs and reporting but no audit trail	Partially compliant; gaps in transparency
Level 3 - Institutional Infrastructure	End-to-end governance-ready AI across the decision chain	Full auditability, explainability, human oversight, bias monitoring	Designed for EU AI Act Compliance

Most AI currently deployed in European universities operates at Level 1 or early Level 2. These tools improve specific tasks but do not address the systemic governance questions that regulators, students, and society are now raising. Level 3 represents a fundamentally different proposition: AI embedded in institutional processes with the same rigour applied to financial systems or data protection- auditable, accountable, and designed for scrutiny.

This framework draws on institutional theory (DiMaggio & Powell, 1983) and the emerging literature on algorithmic governance (Yeung, 2018; Binns, 2018), which argues that governing algorithmic decision-making requires not only technical safeguards but institutional structures that embed accountability into organisational practice.

3. Semantikmatch: A Level 3 Architecture

Semantikmatch is designed as end-to-end AI infrastructure for admissions and academic decision-making. The platform combines large language models with domain-specific control layers across five modules: eligibility verification, document validation, criteria evaluation, video and language assessment, and progression tracking. Each module operates under a governance-by-design architecture comprising four pillars:

- **Auditability:** Every AI-assisted decision is logged with a complete trace of inputs, model outputs, confidence scores, and human review actions, satisfying Article 12 of the EU AI Act.
- **Explainability:** Automated assessments are accompanied by structured justifications aligned with institution-defined criteria- not opaque scores- addressing Article 13’s transparency requirements.
- **Human oversight:** The platform operates as a decision-support system. Admissions officers retain authority over final decisions, with configurable override mechanisms per Article 14.
- **Bias monitoring:** Systematic testing across demographic dimensions with dashboards that flag disparate impact patterns, supporting non-discrimination principles throughout the regulation.

4. Case Study: École des Ponts Business School

École des Ponts Business School partnered with Semantikmatch to pilot an AI-augmented admissions process for its graduate programmes, processing 500 applications across a full cycle. The platform was integrated alongside the institution’s CRM, with all AI outputs surfaced through a dedicated review interface preserving human decision authority.

4.1 Results

Metric	Result
Applications processed	500
Overall accuracy	98-99%
Processing time	Minutes → seconds per application

4.2 Qualitative Impact

Beyond quantitative outcomes, admissions officers reported greater confidence in evaluation consistency and noted that AI-structured assessments made implicit criteria explicit, reducing subjective variability. Critically, the platform was used as an analytical co-pilot: officers reviewed, validated, and occasionally overrode AI recommendations, confirming that governance-ready AI augments rather than displaces professional judgment.

5. The Strategic Imperative: Sovereignty, Liability, and Legitimacy

The regulatory landscape created by the EU AI Act poses a stark strategic question for university leaders: who governs the AI that governs access to your institution?

Universities that rely on opaque, third-party AI tools for admissions screening- tools whose decision logic they cannot audit, whose training data they do not control, and whose compliance posture they cannot verify- are exposed on three fronts. First, regulatory liability: under the AI Act, deployers bear direct obligations regardless of provider compliance. An institution cannot outsource accountability. Second, reputational risk: a single publicised case of algorithmic bias or undetected fraud can undermine years of institutional credibility. Third, and most fundamentally, sovereignty: when critical academic decisions depend on infrastructure controlled by external technology vendors, institutions forfeit a degree of self-governance that is constitutive of the university’s social mandate.

Our legal analysis- conducted with specialised counsel- reveals that the scope of Annex III extends well beyond admissions. Virtually any AI-powered process affecting student access, progression, or outcomes falls under high-risk classification: mobility assessments, academic equivalence evaluations, performance monitoring. The compliance imperative is institution-wide.

The response cannot be reactive. Institutions that delay building governance-ready AI capacity will face a compounding disadvantage: regulatory exposure, vendor lock-in, and inability to demonstrate the transparency that students, regulators, and society increasingly demand.

6. Conclusion

This paper has proposed a three-level maturity framework for AI in higher education administration and argued that only Level 3- institutional infrastructure- meets the governance, regulatory, and ethical requirements of the emerging European landscape. The Semantikmatch deployment at École des Ponts Business School demonstrates

that this is not a theoretical aspiration but an operational reality: institutions can achieve transformative gains in efficiency, accuracy, and fairness while maintaining full human oversight and regulatory compliance.

For university leaders, the strategic imperative extends beyond procurement decisions. It concerns the institutional capacity to govern the algorithms that increasingly mediate access to education. In a democratic society, admissions processes carry a public trust mandate: they determine who gains access to knowledge, credentials, and social mobility. Delegating these decisions to ungoverned AI systems- or failing to govern them adequately- is not merely a regulatory risk. It is a failure of institutional responsibility.

References

- [1] Binns, R. (2018). Algorithmic Accountability and Public Reason. *Philosophy & Technology*, 31(4), 543–556.
- [2] DiMaggio, P.J. & Powell, W.W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2), 147–160.
- [3] EU AI Act: <https://artificialintelligenceact.eu/annex/3/>
- [4] Yeung, K. (2018). Algorithmic Regulation: A Critical Interrogation. *Regulation & Governance*, 12(4), 505–523.
- [5] Case study: <https://semantikmatch.com/resources>

EUonAIR: Shaping European Leadership in Responsible AI Governance and Education

Barbara Domaradzka,
Kozminski University

Abstract

The rapid integration of Artificial Intelligence into higher education presents institutions with a dual challenge: harnessing the transformative potential of AI while mitigating its ethical risks. While many universities are adopting AI tools, few have established a holistic strategy that encompasses governance, ethics, and cross-border collaboration.

This presentation introduces EUonAIR, a pioneering Erasmus+ European University alliance co-led by Kozminski University, dedicated to embedding "Responsible AI" into the very fabric of academic structures. The project moves beyond isolated AI experiments to build a sustainable, European ecosystem for ethical AI adoption.

Our strategy rests on three interconnected pillars designed to reshape university administration and education:

1. **Institutional Governance & Frameworks:** Developing a robust AI governance framework that guides universities in the ethical deployment of AI tools. This involves moving from ad-hoc usage to strategic management of AI resources, ensuring data privacy, fairness, and transparency in administrative and academic processes.
2. **Faculty Empowerment & Training:** Recognizing that technology is only as good as its users, EUonAIR prioritizes comprehensive training for faculty. We focus on "Training the Trainers" in AI ethics, ensuring that educators are equipped to guide students through the moral complexities of algorithmic decision-making.
3. **Cross-Institutional Curriculum Co-creation:** Breaking down silos to develop transnational AI modules and joint micro-credentials. By co-creating curricula across borders, the alliance fosters a unified European approach to AI literacy, bridging the gap between technical skills and business ethics.

The presentation will outline the roadmap for these initiatives and demonstrate how the EUonAIR alliance serves as a model for modernizing higher education. We will discuss the practical drivers for implementation and the cultural changes required to transition from traditional academic models to an AI-enhanced, ethically grounded university environment.

Keywords: Responsible AI, Higher Education Governance, Erasmus+ Alliance, AI Ethics, Cross-institutional Curriculum, Faculty Training.

References

Articles written by prof. Aleksandra Przegalińska, Ph.D. Vicerector for Innovation and AI, Kozminski University

- "Converging Minds: The Creative Potential of Collaborative AI" (book, 2024, coauthor: Tamilla Triantoro).
- "Collaborative AI in the workplace: Enhancing organizational performance through resource-based and task-technology fit perspectives" (Article, International Journal of Information Management, 2024).
- "Cobots in knowledge work: Human – AI collaboration in managerial professions" (Article Journal of Business Research, 2021).
- "Strategizing AI in business and education: Emerging technologies and business strategy" (2023, współautor: D. Jemielniak).
- "Is AI a Good Fit? The Impact of Personality on Generative AI Collaboration and Enjoyment" (Artykuł, 2024)
- Collaborative AI in the workplace: Enhancing organizational performance through resource-based and task-technology fit perspectives" (International Journal of Information Management, 2024)
- "Cobots in knowledge work: Human – AI collaboration in managerial professions" (Artykuł, Journal of Business Research, 2021)

And some others :

Is AI a Good Fit? The Impact of Personality on Generative AI Collaboration and Enjoyment KOVBASIUK ANNA, Triantoro Tamilla, CIECHANOWSKI LEON [i in.], International Journal of Human-Computer Studies, 2026, vol. 210, No.:103747.

The personality profile of early generative AI adopters: a big five perspective

KOVBASIUK ANNA, Triantoro Tamilla, PRZEGALIŃSKA-SKIERKOWSKA ALEKSANDRA [i in.], Central European Management Journal, 2025, vol. 33, nr 2, s.252–264.

Collaborative AI in the workplace: Enhancing organizational performance through resource-based and task-technology fit perspectives

PRZEGALIŃSKA-SKIERKOWSKA ALEKSANDRA, Triantoro Tamilla, KOVBASIUK ANNA [i in.], International Journal of Information Management, 2025, vol. 81, s.1-23,

From Expert Systems to Generative Artificial Experts

SOWA KONRAD, PRZEGALIŃSKA-SKIERKOWSKA ALEKSANDRA, Journal of Artificial Intelligence Research, 2025, nr 82, s.2100-2124.

Cobots in knowledge work: Human – AI collaboration in managerial professions

SOWA KONRAD, PRZEGALIŃSKA-SKIERKOWSKA ALEKSANDRA, CIECHANOWSKI LEON, Journal of Business Research, 2021, vol. 125, s.135-142.

Digital Coworker: Human-AI Collaboration in Work Environment, on the Example of Virtual Assistants for Management Professions

SOWA KONRAD, PRZEGALIŃSKA-SKIERKOWSKA ALEKSANDRA, W: Digital Transformation of Collaboration. Proceedings of the 9th International COINs Conference. / PRZEGALIŃSKA-

SKIERKOWSKA ALEKSANDRA, Grippa Francesca, Gloor Peter A. (red.), 2020, Springer, s.179-201, ISBN 978-3-030-48992-2.

From Anxiety to Agency: A Structured Framework for AI-Integrated Academic Writing Instruction

Katarzyna Truszkowska,
Oxford Academy of
English

Abstract

The rapid adoption of generative AI has created significant anxiety among students and educators regarding academic writing instruction. This case study presents a three-day intensive workshop at Oxford University that transformed AI-related anxiety into student agency through a structured four-stage pedagogical framework: Understanding, Applying, Evaluating, and Creating. Unlike approaches that prohibit AI use, this intervention positions AI literacy as a core learning objective while maintaining academic integrity through constraint-based AI tool design. The proprietary pedagogical AI employed refuses to generate content, instead providing Socratic questioning, diagnostic feedback, and logical gap analysis that preserves student authorship. Analysis of one participant's reflective essay demonstrates transformation from writing avoidance to confident engagement, with self-reported skill progression from 10-15% to 40-50% competency. The student's systematic use of AI support across all framework stages illustrates how structured guidance enables strategic tool use while maintaining authorial control. Findings suggest that effective AI integration in academic writing curricula requires explicit AI literacy outcomes, constraint-based tool design, process-based assessment models, and parallel faculty development. This approach addresses both cognitive and affective dimensions of learning, preparing students for AI-augmented academic and professional contexts while reducing anxiety through concrete guidance rather than vague prohibitions. The four-stage framework offers institutions a principled pathway for developing new curricula that embrace technological reality while maintaining academic rigour.

Keywords: generative AI, academic writing, pedagogical framework, AI literacy, higher education

Introduction

The rapid adoption of generative AI has created a crisis of confidence in academic writing instruction. Students face competing pressures, such as institutions warning against AI use while peers openly discuss using ChatGPT, which creates uncertainty about appropriate use and fear of unintended plagiarism. Educators simultaneously struggle with assessment validity, preparing students for AI-augmented workplaces, and maintaining academic standards. This case study presents a curricular intervention designed to transform anxiety into agency. Rather than prohibiting AI use, it positions AI literacy as a core learning objective with structured pedagogical scaffolding guiding students toward thoughtful and ethical AI integration. This case study documents a three-day intensive workshop at Oxford University that shifted AI-integrated academic writing instruction from reliance on commercial large language models (LLMs) to a proprietary pedagogical AI tool designed to prevent academic dishonesty and overreliance.

Curriculum Design: A Four-Stage Framework

The workshop curriculum employed a four-stage framework originally presented at the TESOL Conference at the University of Sharjah in April 2025, reimagining Bloom's Taxonomy for AI-augmented education (Truszkowska, 2025). Each stage builds from basic AI awareness to sophisticated AI-integrated creative work.

Stage 1: Understanding

Students explore AI capabilities and limitations through hands-on analysis of commercial LLM outputs, learning to recognise hallucinations, bias, and reliability issues. This demystifies AI, reducing both over-reliance and fear (Quelle & Bovet, 2024).

Stage 2: Applying

Clear policies guide appropriate AI use through explicit discussion of academic integrity standards, modelling of appropriate interactions, and practice exercises with bounded tasks. This replaces vague prohibitions with concrete guidance (University of Sheffield Library, 2023).

Stage 3: Evaluating

Students develop metacognitive awareness through comparative analysis of student-generated versus AI-suggested text, fact-checking exercises, and decision-making practice about when to accept, modify, or reject AI guidance (Lee et al., 2025).

Stage 4: Creating

Students complete full writing projects with documented AI interactions, developing portfolios that demonstrate ownership while leveraging AI as a sophisticated tool that enhances rather than replaces their capabilities.

Implementation Context: Oxford Summer Workshop

The curriculum was implemented during a three-day intensive workshop at Oxford University. Participants were students aged 18 and above preparing to study at English-speaking universities - a population facing anxiety about meeting academic writing standards in their second language while navigating unfamiliar AI use policies.

Pedagogical AI Tool: Constraint-Based Design

The workshop employed a proprietary AI tool designed specifically for pedagogical purposes. Unlike commercial LLMs, this tool embodies several deliberate constraints:

1. **No Content Generation:** The tool cannot write sentences, paragraphs, or essays, eliminating the temptation to outsource authorship.
2. **Socratic Questioning:** The system guides through questions rather than answers, prompting reflection on the student's own ideas.
3. **Diagnostic Feedback:** When students submit draft text, the tool identifies strengths and weaknesses but never rewrites content.
4. **Logical Gap Analysis:** The system highlights missing connections, unsupported claims, and structural weaknesses without fixing them.
5. **Phase-Appropriate Prompts:** The tool adapts its questioning to the writing stage (planning, drafting, refinement).

These constraints were explicitly discussed with students as pedagogical choices designed to maintain their agency and authorship while providing sophisticated support.

Evidence of Transformation

Case Analysis: One Student's Journey

One participant's reflective essay provides compelling evidence of the anxiety-to-agency transformation. Her initial relationship with writing was characterised by doubt and avoidance: "I've never been a big fan of writing because I never thought I could 'do a good job at it' and didn't come as natural to me as public speaking." This anxiety extended to the writing process itself, where solitary engagement with "various articles and resources" led to disengagement and distraction. The introduction of structured AI support transformed this experience. She documented systematic engagement across all framework stages while writing on the topic of increased life expectancy and intergenerational pressure.

Stage 1 & 2 (Understanding and Applying)

She began independently developing her argument structure, then strategically engaged the AI tool: "I gave it my ideas and asked it if I had missed anything, if it thought the arguments are strong enough, or if I should find others. This step really helped me enhance my plan for the essay." This demonstrates confidence in her own thinking (she began with her ideas) combined with strategic tool use (she sought feedback to strengthen her approach).

Stage 3 (Evaluating)

During drafting, she returned to the tool for targeted support: "When I was writing the essay and needed to add a bit more content I again turned to the AI tool and asked it to give me ideas for more interesting people or research, which I could include to support my argument and make it stronger." She recognised gaps in her argument independently and used AI to identify potential evidence sources. This involved exercising judgment about what her argument needed rather than accepting AI's determination of completeness.

Stage 4 (Creating)

In the refinement phase, she maintained authorial control while leveraging AI feedback: "I gave each paragraph separately to the AI tool and asked it to help me with clarity. What I most needed help with was cohesion between sentences, so that the connections between the ideas are more explicit and to enhance the overall flow. I didn't blindly copy paste what I got, but put my essay and AI's version side by side and took all the suggestions I thought truly perfected my paragraphs." This represents sophisticated agency. She identified her specific weakness (cohesion), sought targeted feedback, and selectively implemented suggestions based on her own judgment.

Active Transformation

The student's concluding reflections reveal the depth of transformation: "Having the freedom to use AI, not just to do the work for you, but actually help you think about your writing, help you enhance it and finally give you a reflection of the strengths and weaknesses was surprisingly even more enjoyable than I predicted." Anxiety transformed into enjoyment. The writing process, previously avoided, became "stimulating, pleasant and interesting." She rated her experience 7 out of 10 and estimated her skill progression from "about 10/15% to hopefully about 40/50%." Most significantly, she reported changed expectations for university study: "Now I also have a bit of an anticipation for university, since I will be able to further develop them there." Anxiety about academic writing had transformed into anticipation. It shows a fundamental shift in self-efficacy and future orientation.

Engagement and Persistence

The structured AI support addressed a persistent challenge in writing instruction: maintaining engagement during the lengthy, often frustrating process of composition and revision. She says: "Academic writing could get boring when you get caught up in various articles and resources and having the opportunity to use a tool which can give you a 'teacher-like' reflection on your work while you are still in the progress keeps you engaged with what you are doing and not get distracted as easily." This suggests that pedagogical AI can reduce the anxiety-driven avoidance that often derails student writing projects.

Curriculum Implications for Higher Education

New Curricula Require New Outcomes

In an AI-augmented world, academic writing curricula must explicitly teach AI literacy as a core competency alongside traditional writing skills. This includes:

- Critical evaluation of AI outputs: treating AI suggestions as drafts requiring human judgment

- Strategic tool selection: knowing when and how to use AI effectively
- Transparent documentation: making AI interactions visible and ethically defensible
- Metacognitive awareness: understanding one's own writing process and AI's role within it

These outcomes represent a fundamental curriculum shift from teaching writing as if AI does not exist to teaching writing as an AI-augmented practice.

Constraint-Based AI Design as Pedagogical Principle

Not all AI tools are pedagogically equivalent. Commercial LLMs designed to maximise user satisfaction often undermine learning by doing too much. Curriculum design should prioritise or develop AI tools that:

- Refuse to generate complete work products
- Provide feedback rather than corrections
- Ask questions rather than provide answers
- Make students work harder, not less

This principle applies across disciplines as pedagogical AI should increase cognitive load in productive ways rather than reducing it.

Process-Based Assessment Models

Traditional product-only assessment becomes problematic when AI can generate sophisticated outputs. New curricular models should:

- Require documentation of process (portfolios, screen captures, decision logs)
- Include reflective components that demonstrate metacognitive awareness
- Evaluate strategic AI use as a competency
- Make AI interactions transparent rather than hidden

This shifts the assessment from detecting AI use to evaluating the quality of AI integration.

Explicit Anxiety Reduction Strategies

Curriculum design should acknowledge and address student and faculty anxiety about AI:

- Replace vague warnings with concrete guidance
- Model appropriate AI use rather than only prohibiting misuse
- Create safe spaces for experimentation and mistake-making
- Celebrate strategic AI use rather than only punishing problematic use

The workshop's explicit framework reduced anxiety by providing structure and clarity in an otherwise ambiguous landscape.

Faculty Development as Parallel Curriculum

Students cannot develop AI agency if instructors remain anxious and uncertain. Curricular innovation requires parallel faculty development that:

- Provides hands-on experience with pedagogical AI tools
- Develops instructor AI literacy across the four-stage framework
- Creates communities of practice for sharing strategies and concerns
- Addresses faculty anxiety about changing roles and competencies

Challenges and Limitations

Several challenges emerged during implementation.

Initial Resistance

Some students initially expected the AI tool to write for them. They required explicit instructions and reframing of AI's role and the pedagogical rationale for constraints.

Variable Digital Literacy

Students entered with varying levels of comfort with technology, particularly AI, requiring differentiated support and sometimes one-on-one guidance.

Tool Dependency

The proprietary AI tool required stable internet access and technical support. Broader implementation would require institutional licensing of similar tools.

Cultural Variations

Students preparing for English-speaking universities arrived with different cultural assumptions about academic integrity, collaboration, and technology use. The curriculum needed to address these variations explicitly.

Recommendations for Curricular Adoption

Institutions developing new AI-augmented writing curricula should consider:

Implementing the Four-Stage Framework Across Courses

Rather than treating AI integration as an add-on, embed the Understanding-Applying-Evaluating-Creating progression throughout writing curricula. Each course or module should explicitly indicate which stage(s) it addresses.

Developing or Adopting Pedagogical AI Tools

Invest in AI tools specifically designed for learning, not efficiency. Where commercial tools are used, provide explicit guidance on constraint-based use (e.g., "Ask ChatGPT for three counterarguments but don't use its phrasing").

Making Portfolio Assessment Standard Practice

In AI-augmented courses, process documentation should be weighted as heavily as final products. This reduces anxiety about AI detection while rewarding strategic use of AI.

Creating Institutional AI Literacy Standards

Establish clear, progressive competencies for AI use across educational levels. Students should graduate with documented AI literacy as an explicit credential.

Investing in Faculty Development

Provide sustained professional development focused on pedagogical AI integration. This should include hands-on tool exploration, curriculum design workshops, and ongoing peer support.

Building Communities of Practice

Create spaces, both physical and digital, where students and faculty can discuss AI integration challenges, share strategies, and normalise the anxieties accompanying major pedagogical transitions.

Conclusions

This case study demonstrates that AI integration in academic writing curricula need not create an integrity crisis or undermine learning outcomes. When approached through structured, developmental frameworks that prioritise student agency over efficiency, AI can transform the writing experience from anxiety-provoking to confidence-building. The four-stage framework - Understanding, Applying, Evaluating, Creating - provides a curricular roadmap for this transformation. By explicitly teaching AI literacy, modelling appropriate use, requiring critical evaluation, and maintaining authorship standards, educators can prepare students for the realities of AI-augmented academic and professional writing. Most significantly, this approach addresses the affective dimension of learning. Students move from anxiety about AI use to agency in AI integration, from doubt about their capabilities to confidence in their judgment, from avoidance of writing to anticipation of future development. As higher education institutions develop new curricula for an AI-augmented world, this model offers a principled, practical approach that maintains academic rigour while embracing technological reality. The goal is not to return to a pre-AI past but to move forward into an AI-integrated future where students possess both traditional writing competencies and the digital literacy, critical judgment, and strategic thinking required for success in an AI-augmented world.

References

- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. Microsoft Research. https://www.microsoft.com/en-us/research/wp-content/uploads/2025/01/lee_2025_ai_critical_thinking_survey.pdf
- Quelle, D., & Bovet, A. (2024). The perils and promises of fact-checking with large language models. *Frontiers in Artificial Intelligence*, 7, 1341697. <https://doi.org/10.3389/frai.2024.1341697>
- Truszkowska, K. (2025, April). *Revised Bloom's taxonomy in the age of AI* [Conference presentation]. TESOL Conference, University of Sharjah, Sharjah, United Arab Emirates.
- University of Sheffield Library Guides. (2025). Using AI ethically: Guidance for students and staff. University of Sheffield. <https://www.sheffield.ac.uk/library>

Navigating the ‘Wicked Problem’ of AI Assessment: A Case Study of an Exam Redesign

Ørjan Flygt Landfald,
BI Norwegian Business
School

Abstract

Generative artificial intelligence (GenAI) presents a "wicked problem" for higher education assessment, characterized by challenges that resist definitive formulations or simple technical fixes (Corbin et al., 2025). As AI fundamentally changes how we write, synthesize, and communicate information, the validity of traditional home exams has eroded. Our observations internally at BI Norwegian Business School and from the sector in general reveal a critical breakdown: Assessors frequently struggle to distinguish authentic work from AI outputs, and both faculty and students face uncertainty regarding the boundaries between legitimate assistance and academic dishonesty or between using AI to support learning and replace learning.

Current policies often rely on vague definitions of "originality" or ineffective detection tools, failing to address how human-AI co-creation alters the nature of student work (Luo, 2024). This case study details how BI Norwegian Business School is navigating this tension through its "Exam Redesign Project," a systemic initiative to enact necessary assessment reform (Lodge et al., 2025).

Moving beyond the "solutionism" of policing tools, BI's exam redesign is anchored by three initiatives currently being piloted in "sandbox courses," structured around the logic of testing, verifying, and assessing:

- **Testing: The Assessment Center.** A mid-term exam conducted in a controlled environment to assure basic conceptual understanding before students engage in complex tasks.
- **Verifying: Digital Oral Defense.** A scalable validation mechanism using risk-based sampling (5-10% of cohorts) to verify authorship and critical understanding.
- **Assessing: Process-Based Evaluation.** A shift from grading the final product to evaluating the learning journey, utilizing real-time AI scaffolding to measure the development of higher-order literacies.

This paper argues that by combining secure baseline testing, randomized verification, and process-oriented assessment, universities can maintain qualification integrity. We present preliminary data from the pilot phase, offering a roadmap for institutions seeking assessment models that are both pedagogically robust and financially viable.

Keywords: (max 5): Assessment Reform, Academic Integrity, AI Literacy

1. Introduction: The "Wicked Problem"

Generative artificial intelligence (GenAI) presents a "wicked problem" for assessment in higher education: challenges that resist definitive formulations and offer better or worse rather than "correct" solutions (Corbin et al., 2025).

The core of this problem is a fundamental shift in literacy. AI is about to change the way we write, read, find, synthesize, and communicate information. While academic writing is still an essential skill for university-level education, it is now very hard to assess whether students have learned *how* to write or used AI to write for them. Often, assessors simply don't know how to grade student assignments. This ambiguity creates a paralyzed environment where both faculty and students are uncertain about where to draw the boundaries between using AI for learning and misusing AI for cheating.

At BI Norwegian Business School, we observed a reactive trend: Many faculty members have transitioned back to traditional proctored exams due to a fear of AI misuse rather than for pedagogical reasons. While this secures the exam, it often comes at the cost of authentic assessment. To counter this regression, BI initiated the "Exam Redesign Project," aimed at finding a middle ground that secures integrity without abandoning modern tools.

2. Introduction: The "Wicked Problem"

The redesign is driven by the recognition that the ability to understand how, when, and when not to cooperate with AI is now an essential skill for both academia and work life.

General competencies like analytical thinking, social skills, resilience, and creativity will become increasingly important for the future (World Economic Forum, 2025). Consequently, these skills should be a part of learning and assessment, rather than being sidelined by a return to rote memorization exams.

General and subject-specific AI literacy must therefore be part of course designs and assessed in exams (Becker et al., 2024). However, current assessment formats often fail to capture these literacies. If we only grade the final output, we fail to measure the students' ability to critically evaluate and refine AI-generated content. Our project argues that we need structural innovation to assess these new competencies effectively.

3. The BI Solution: Three Forms of Redesign

BI's approach moves away from "solutionism" the attempt to "fix" AI through detection software—and toward structural navigation. The project is anchored by three initiatives currently being piloted in "sandbox courses." These pillars are designed to be financially sustainable for large-enrollment business courses while remaining pedagogically sound.

3.1. Testing: The Assessment Center ("Academic Theory Test")

The first pillar addresses the need for a secure baseline. Before students can engage in high-level applications, we must ensure they possess the necessary mental schema.

The Assessment Center is a mid-term exam conducted in a controlled environment without aids to test basic conceptual understanding.

- **Student-Centric Logistics:** Students can book timeslots themselves within a testing window.
- **Pedagogical Intent:** Students are given three attempts to pass. This structure can reduce high-stakes stress and at the same time raise the threshold for passing. The pedagogical intention is to assure basic knowledge before entering a more authentic graded exam testing learning outcomes related to application of knowledge and higher-order skills.

3.2. Verifying: Digital Oral Defense

The second pillar addresses the "who" of submission. Mastering academic writing is key to developing advanced critical thinking and analytical capabilities. However, we need methods to quality-assure students' understanding in home exams to ensure they have not simply outsourced the thinking process.

BI utilizes **10-minute Zoom-based defenses** to solve the scalability issue. Rather than interviewing every student—which would be financially impossible—we utilize a risk-based sampling model:

1. **Random Draw (5-10%):** A random selection to normalize the process.
2. **Academic Uncertainty (5-10%):** A targeted draw based on discrepancies in the submitted work.

This approach verifies the student's ability to defend their own work. The mere possibility of being called in for a defense creates a "normative pressure" that encourages integrity across the entire cohort, making the format sustainable without exceeding faculty capacity.

3.3. Assessing: Process-Based Evaluation

The third pillar represents the shift from product to process.

Building on learning analytics that previously tracked only the quantity of student activity, this pilot aims to implement continuous evaluation of the quality of learning. Utilizing a specialized LLM API integrated into the learning platform, the system provides real-time scaffolding and feedback while measuring learning outcomes during active student tasks.

This shifts the focus from the final product to the learning journey itself (Schuwirth & Van der Vleuten, 2011). It allows the institution to assess the students' resilience and analytical thinking as they interact with the material, ensuring that "AI Literacy" is treated as a functional competence to be developed, not a threat to be policed.

4. Discussion

Our goal with the "Exam Redesign Project" at BI Norwegian Business School is to avoid having to choose between security and authenticity. By separating the functions of assessment—using the Assessment Center for testing knowledge, Digital Oral Defenses for verifying authorship, and Process-Based Evaluation for assessing skills—we can move one step closer to navigate the "wicked problem" of AI.

This structural approach allows us to maintain qualification integrity while equipping students for an AI-augmented workforce. Ultimately, these initiatives adhere to our project's core principle: **AI should support learning, not replace it.**

References

- Becker, K. P., Parker, J. L., & Richter, D. (2024). Framework for the Future: Building AI Literacy in Higher Education. Moxie.
- Corbin, T., Bearman, M., Boud, D., & Dawson, P. (2025). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*. Advance online publication. <https://doi.org/10.1080/02602938.2025.2553340>
- Lodge, J. M., Bearman, M., Dawson, P., Gniel, H., Harper, R., Liu, D., McLean, J., Ucnik, L., & Associates. (2025). Enacting assessment reform in a time of artificial intelligence. Tertiary Education Quality and Standards Agency, Australian Government.
- Luo, J. (2024). A critical review of GenAI policies in higher education assessment: a call to reconsider the “originality” of students’ work. *Assessment & Evaluation in Higher Education*, 49(5), 651–664. <https://doi.org/10.1080/02602938.2024.2309963>
- Schuwirth, L. W. T., & Van der Vleuten, C. P. M. (2011). Programmatic assessment: From assessment of learning to assessment for learning. *Medical Teacher*, 33(6), 478–485. <https://doi.org/10.3109/0142159X.2011.565828>
- World Economic Forum. (2025). The Future of Jobs Report 2025. Geneva: World Economic Forum.

Operationalizing Responsible AI in High-Stakes Educational Assessment: The Duolingo English Test Case Study

Alina A. von Davier, Jill Burstein, Geoff LaFlair, Kevin Yancey, Masha Kostromitina

Duolingo, Inc.

Abstract

Artificial intelligence is increasingly embedded in high-stakes educational assessments, raising critical questions about validity, fairness, transparency, and governance. While responsible AI (RAI) frameworks articulate high-level ethical principles, guidance on their operationalization in assessment remains limited. This paper examines how RAI can be translated into concrete practices through a case study of the Duolingo English Test (DET), a digital-first English proficiency assessment designed around AI-enabled processes. We describe a set of assessment-specific RAI standards including Validity and Reliability, Fairness, Privacy and Security, and Accountability and Transparency, and illustrate how they are implemented across the assessment lifecycle. Implications for higher education institutions evaluating AI-enabled assessments are discussed.

Keywords: responsible AI, high-stakes assessment, AI-based scoring, test validity and fairness, higher education admissions

1. Background: Responsible AI and Assessment in Higher Education

Artificial intelligence (AI) is increasingly embedded in educational assessment systems, reshaping how tests are designed, administered, scored, and interpreted. In higher education, AI-enabled assessments promise substantial benefits, including scalability, efficiency, expanded access, and the ability to measure complex skills such as writing and speaking at scale. At the same time, the use of AI in high-stakes contexts (e.g., particularly admissions and placement) raises concerns about ethics and governance practices related to validity, privacy, and transparency. These concerns (e.g., algorithmic bias, test-taker privacy, data representation) have motivated the rapid development of responsible AI (RAI) principles and frameworks across sectors as well as and to human-centered AI for a clear articulation of human values that need to be reflected in AI applications (von Davier & Burstein, 2024). Domain-agnostic AI governance efforts, such as those advanced by the Organisation for Economic Co-operation and Development (OECD, 2024) and the U.S. National Institute of Standards and Technology, emphasize trustworthiness characteristics including validity, fairness, accountability, transparency, privacy, and security (National Institute of Standards and Technology [NIST], 2023). However, translating these high-level principles into concrete, operational practices remains challenging, particularly in educational assessment, where decisions can have long-term consequences for individuals'

academic and professional trajectories. Specifically, developing standards and sustainable practices to align with principles requires substantial effort and diverse expertise.

Educational assessment has a long-standing tradition of ethical and validity-focused standards. Argument-based validity theory emphasizes the need for evidence to support test score interpretations and uses, conceptualizing validity as a chain of inferences from observed performance to real-world outcomes (Kane, 2013). Similarly, the Standards for Educational and Psychological Testing foreground validity, reliability, and fairness as core principles guiding test development and use (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014). While these frameworks remain foundational, they were largely developed prior to the widespread use of modern AI techniques and therefore do not explicitly address risks introduced by automated and generative systems.

This gap has important implications for higher education institutions adopting AI-enabled assessments. Universities are often asked to make trust judgments about AI-based tools without clear guidance on what responsible implementation looks like in practice. From an institutional perspective, RAI in assessment is not only a technical issue, but also a governance issue: how AI is used across the assessment lifecycle, how risks are mitigated, and how evidence supporting score use is documented and communicated to stakeholders.

This paper addresses these challenges by examining responsible AI standards in educational assessment through a concrete use case: the Duolingo English Test (DET). Rather than proposing abstract principles, we focus on how responsible AI can be operationalized in a high-stakes, real-world assessment system that is already widely used in higher education admissions.

2. The Duolingo English Test as a Use Case for Responsible AI

Human-centered AI prioritizes human values, needs, goals, and decision power in the design of AI technologies (Auernhammer, 2020), consistent with principles like Embedding Values into Autonomous Systems (IEEE, 2017). In educational testing, this means creating AI systems that align with assessment science principles (e.g., validity), ethical considerations for all stakeholders, and test-takers' experience and access. Trust in these AI systems is essential for their widespread adoption, as users must be confident in their reliability, fairness, and accuracy. With that in mind, the Duolingo English Test (DET), a digital-first, high-stakes English language proficiency assessment used by thousands of higher education institutions worldwide, was designed from the outset to leverage AI across the assessment ecosystem. AI is used in test design (e.g., automated item generation), measurement (e.g., automated scoring of writing and speaking), and security (e.g., plagiarism detection and remote proctoring), making the DET a particularly instructive case for examining responsible AI in assessment.

To address the ethical and validity challenges associated with extensive AI use, the DET has developed a set of Responsible AI standards (Burstein, 2025) tailored specifically to educational assessment. These standards were created through a multi-stage process involving (a) review of AI ethics literature, (b) alignment with established assessment standards and guidelines, (c) cross-disciplinary expert consultation spanning assessment science, machine learning, law, and security, and (d) validation against the NIST AI Risk Management Framework. The resulting framework consists of four interrelated standards:

Validity and Reliability; Fairness; Privacy and Security; and Accountability and Transparency. We offer a high-level description of each standard and associated example practices below.

The Validity and Reliability standard focuses on ensuring that AI-enabled processes support appropriate score interpretations for their intended use. In the DET, this includes explicit construct definition by human subject-matter experts, alignment between AI features and theoretically grounded language constructs, and systematic evaluation of automated scoring models against human ratings. Importantly, AI systems themselves become objects of validation: scoring models are evaluated for accuracy, consistency, and explainability, and model features are designed to be interpretable and linked to meaningful aspects of language proficiency.

The Fairness standard explicitly targets test equity. AI-generated test content undergoes human-in-the-loop fairness and bias review to identify construct-irrelevant barriers, culturally sensitive material, and potential sources of disadvantage for specific groups. In addition, demographic representation in training and evaluation data is documented and monitored, and statistical techniques such as differential item functioning and differential rater functioning are used to detect and mitigate algorithmic bias. These practices reflect a shift from treating fairness as a post-hoc evaluation to embedding it throughout the assessment lifecycle.

The Privacy and Security standard addresses risks related to data protection, test-taker privacy, and academic integrity. While AI is used to support test security (e.g., detecting plagiarism or anomalous test-taking behavior), decisions are not fully automated. Instead, the DET employs a human-in-the-loop approach in which AI tools surface evidence and trained human reviewers make final determinations. This design reflects emerging consensus that meaningful human oversight is essential when AI systems are used in high-stakes educational decisions.

Finally, the Accountability and Transparency standard emphasizes governance, documentation, and communication. The DET documents how AI is used in test development, scoring, and security, and maintains formal processes for reviewing and approving changes to AI systems. Research and technical documentation are disseminated publicly to support institutional decision-making and stakeholder trust. Transparency, in this context, is not limited to disclosure, but includes enabling stakeholders to understand how AI-supported scores are generated, what evidence supports their use, and what limitations remain.

The DET case study presented here illustrates how ethical principles can be used to build responsible AI standards for educational measurement. It further describes the types of practices that are aligned with RAI standards to support test quality and equity. While the DET example is an illustration of standards and practices, it is important to keep in mind that RAI infrastructure is not a one-time exercise and must be sustainable long term. To that end, it must develop in tandem with AI and other technological advances, including increasingly powerful generative and multimodal models. Moreover, some emerging ethical concerns remain only partially addressed within current assessment-focused frameworks.

3. Implications for Higher Education

For higher education institutions, this case study demonstrates that evaluating AI-enabled assessments requires attention not only to outcomes, but also to processes and governance structures. Responsible AI in assessment is best understood as an ongoing institutional

commitment, supported by explicit standards, human oversight, and transparent evidence practices. By examining how responsible AI is operationalized in systems like the DET, institutions can make more informed decisions about adopting AI-enhanced assessments in ways that align with educational values and equity goals.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Psychological Association. <https://www.apa.org/science/programs/testing/standards>

Auernhammer, J. (2020). Human-centered AI: The role of Human-Centered Design research in the development of AI. In S. Boess, M. Cheung, & R. Cain (Eds.), *Synergy-DRS international conference 2020* (pp. 1315-1333). <https://doi.org/10.21606/drs.2020.282>

Burstein, J. (2025). Responsible AI standards. Duolingo Research Report DRR-25-05. <https://go.duolingo.com/ResponsibleAI>

Burstein, J., LaFlair, G. T., Yancey, K., von Davier, A. A., & Dotan, R. (2025). Responsible artificial intelligence for test equity and quality: The Duolingo English Test as a case study. In E.M. Tucker, E. Armour-Thomas, & E.W. Gordon (Eds.), *Handbook for assessment in the service of learning, Volume I* (pp. 419-454). UMass Amherst.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). Building a validity argument for the Test of English as a Foreign Language. Routledge. <https://doi.org/10.4324/9780203937891>

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1-73. <https://doi.org/10.1111/jedm.12000>

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems (Version 2). IEEE. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). <https://doi.org/10.6028/NIST.AI.100-1>

OECD. (2024). AI principles. OECD. <https://www.oecd.org/en/topics/sub-issues/ai-principles.html>

von Davier, A. A. & Burstein, J. (2024). AI in the Assessment Ecosystem: Implications for Fairness, Bias, and Equity. In: Ilic, P., Casebourne, I., Wegerof, R. (eds.) *Artificial Intelligence in Education: The Intersection of Technology and Pedagogy*, Springer Nature Switzerland.

Rebuilding Trust in Assessment with Generative AI: Evidence Judgement through Mode-Switching Competence

Hsin-Hsuan Meg Lee,
ESCP Business School

Abstract

When essays, reports, and professional outputs can be produced quickly with AI, surface quality stops being reliable evidence of capability. Grades become harder to defend. The evidential chain between work and judgement has shifted, and institutions have largely responded by tightening procedure, rewriting tasks, mandating disclosure, expanding invigilated components. This paper argues that procedural responses leave the deeper problem untouched. In institutions that actively embrace AI, teaching students to work with AI as normal practice creates an interpretive problem: stakeholders still read grades as signals of individual capability that should hold across contexts and tools. Mode-switching competence, the capacity to judge when AI is warranted, regulate reliance under epistemic risk, and reason when AI is removed or unreliable, should be treated as a core assessed capability. Four micro-switch patterns show how this judgement can be made observable without surveillance or platform dependence.

Keywords: genAI, assessment design, mode-switching competence, evaluative judgement

The Evidential Problem

In the GenAI era, higher education faces a trust problem that is commonly framed as an academic integrity issue. When essays, reports, decks, and even “authentic” professional outputs can be produced cheaply and quickly, surface quality no longer reliably evidences capability. Educators struggle to stand behind what a submission indicates; students doubt whether their capabilities can be seen and distinguished; employers question whether credentials still carry meaningful information. The practical symptom is familiar: AI influence is often suspected but rarely demonstrable to an evidential standard, leaving judgement trapped between intuition and proof (Corbin et al., 2025a; Corbin et al., 2025b). This paper does not argue for ever-tighter attribution or the “prevention” of AI involvement. It argues that the core challenge is evidential: what a grade can credibly be interpreted to mean when AI is embedded in ordinary student workflows (Digital Education Council & Pearson, 2025). Institutional responses have nevertheless leaned toward control, rewriting tasks, tightening rules, mandating AI-use statements, and expanding invigilated and oral components (Krause et al., 2025). Such measures may stabilise procedure; they do not, on their own, restore interpretable evidence of capability.

The problem becomes especially acute in institutions that actively embrace AI, because adoption changes what is treated as competent practice. When students are

taught to draft, edit, and analyse with AI as legitimate learning practice, they reasonably infer that competent work includes tool-supported work. Assessment then inherits an interpretive mismatch: institutions teach AI-enabled workflows as normal practice, while stakeholders continue to read grades as signals of individual capability that should travel across contexts and tools. Students experience moving goalposts, while staff face a double demand: to encourage AI use for learning while defending outcomes to employers and accrediting bodies who require interpretable evidence of capability. Under these conditions, “acceptable use” policies struggle to stabilise practice because the boundary of use is context-dependent and difficult to operationalise consistently across tasks and markers (Corbin et al., 2025b).

In AI-embracing institutions, the implication is not merely that some tasks will be “AI-off” while others are “AI-enabled,” but that programmes must be designed so evidence remains interpretable across both conditions. AI-off assessment provides a diagnostic baseline of what students can carry unaided when tools are removed. AI-enabled assessment is equally necessary because institutions increasingly teach, and implicitly certify, readiness to incorporate AI into professional workflows. The trust problem therefore sits inside AI-enabled assessment: when AI use is legitimate, attribution becomes uncertain and scepticism becomes the default, so a finished artefact cannot carry the evidential burden on its own. What a grade needs to certify, then, is not the quality of the output but the quality of the judgement behind it: what was decided, why, what was verified, and what would change that judgement, especially given evidence that AI-based learning tools can improve some outcomes while producing uneven effects on skills and higher-order processes (Luo et al., 2025). The remaining design challenge is how to make such judgement observable without turning assessment into surveillance or a narrow test of platform-specific proficiency.

From calibrated mode choice to metacognitive regulation

If assessment in AI-enabled contexts is to carry credible evidential weight, it must demonstrate how students decide when and how to work with AI. The capability at stake is calibrated mode choice: judging whether a task warrants AI support, regulating reliance under epistemic risk, and sustaining reasoning when support is withdrawn or unreliable. Without evidence of this calibration, assessment outcomes remain ambiguous, because it is unclear whether apparent performance reflects student judgement or tool dependence.

This calibrated mode choice, mode-switching competence, can be understood as metacognitive regulation in collaborative human–AI work. Research on collaborative AI frames effective use as regulating joint cognitive labour: planning task allocation, monitoring the relevance and reliability of AI contributions, and evaluating when independent reasoning must take precedence (Sidra & Mason, 2025). The assessed object is therefore not reflective commentary, but the quality of regulation: recognising when AI adds epistemic value, when it introduces risk, and when reasoning must proceed without it.

Mode-switching competence is a distinct construct. It is visible in what students delegate to AI, what they retain as human judgement, what they interrogate or verify, and how they respond when outputs conflict with domain constraints or new evidence (e.g., Gonsalves, 2024). It consists of *ex ante* judgement about whether AI is warranted; calibrated reliance under uncertainty or constraint; the ability to continue reasoning when AI is removed; and the capacity to revise decisions when conditions change. This distinguishes it from AI literacy, which concerns knowing tools and prompting; from technical fluency; from post-hoc reflection; and from generic critical thinking; none of which require the student to regulate reliance in real time under epistemic risk.

Recognising mode-switching competence as the assessed capability also foregrounds a validity threat: platform capture. When institutions standardise on a single enterprise AI system, repeated exposure to its interaction norms and defaults can train students to optimise for that system’s affordances. Assessment may then reward platform familiarity rather than transferable judgement about technology use itself. From an assessment perspective, this is both an institutional risk and a construct validity problem: platform-bound competence inflates apparent performance on skills that do not travel across tools or contexts, while grades continue to be interpreted as evidence of general capability.

Designing observable judgement through predictable micro-switches

Mode-switching competence is only assessable when assessment design creates situations in which students must calibrate reliance in real time. Because most AI-enabled work occurs outside observed environments, mode choice and regulation cannot be inferred reliably from a finished artefact or post-hoc declarations. The design challenge is therefore to embed predictable, low-burden micro-switches into assessment: short checkpoints where students demonstrate how they allocate cognitive labour, manage epistemic risk, and adapt when conditions change. These checkpoints must remain feasible at scale, so they should be brief, standardised, and easy to mark, with heavier oral components used sparingly. Critically, their evidential value depends on timing: when a switch occurs determines whether assessment reveals unassisted carryover of understanding or merely rehearsed compliance. This aligns with a broader shift toward redesign that sustains interpretability under routine GenAI use, rather than relying on attribution claims that are difficult to substantiate consistently (Bearman et al., 2024; Oc et al., 2024).

A first pattern is AI-on build, AI-off defend, suited to modules where the primary deliverable is a report, proposal, analysis memo, or strategy recommendation. Students complete the core work with AI permitted, then complete a short defence without AI in a fixed window. The defence targets one high-stakes judgement call and requires the student to state the decision and its assumptions, identify a plausible failure condition, and specify what would change their mind. This “dual condition” logic, combining extended coursework with a bounded independent demonstration, also appears in institutional guidance seeking a workable balance between tool-augmented work and evidence that learners can meet outcomes on their own (University of Reading, 2025). The assessed object is calibration logic: whether the student can articulate and stress-test the reasoning that governs their AI-enabled output.

A second pattern is AI-on submit, AI-off stress-test, which differs from the defence by perturbing the environment rather than probing the original reasoning. A stress test is a short, timed disruption, e.g., new constraint, contradictory data, scenario twist, or trade-off that makes the original solution partially invalid. Students respond without AI by stating what remains stable, what breaks, what assumption was exposed, and what they would verify next. The purpose is not to punish incomplete anticipation, but to observe whether students can adapt reasoning when conditions change, the circumstance in which mode-switching becomes visible.

A third pattern targets verification without relying on written appendices that can themselves be outsourced. Instead of a “verification exhibit,” require verification under constraint: students select one consequential claim and, in a time-bounded setting, demonstrate how they would check it and what the check implies, confirmed, corrected, or uncertain. This operationalises a core principle in evaluative judgement: assessment should keep students as arbiters of quality and make the basis of judgement visible (Bearman et al., 2024). The assessed object is observable judgement: what they choose to verify, how they do it, and how they respond when evidence remains incomplete.

A fourth pattern targets platform capture by assessing mode choice itself rather than platform performance. Students receive a short prompt and must declare, before producing any output, whether they will work in AI-on or AI-off mode, with a one-sentence justification tied to epistemic risk (e.g., originality requirements, sensitivity to hallucination, reliance on domain constraints, cost of error). They then complete a brief response under the chosen mode, after which the mode is switched and they produce a short revision note explaining what changed and why. The assessed object is the calibration rule and the capacity to recalibrate under altered conditions, not the elegance of the final answer.

These patterns work only when switching is expected rather than exceptional. If observation occurs through rare audits, students will treat checkpoints as risks to manage rather than competences to develop. A more defensible institutional logic is to bake in predictable micro-switches across a module or programme: repeated, low-burden checkpoints, with higher-stakes defences used selectively. Sequencing becomes part of the evidential design. Early AI-enabled work can support exploration, but should be followed by AI-off checkpoints that reveal what students can carry unaided; early AI-off baselining should be followed by AI-enabled tasks that test whether students can integrate tools without surrendering responsibility for reasoning. Placed at informative moments, after exploration, under changed constraints, and before final consolidation, these checkpoints sample judgement when it is most diagnostic. In this way, assessment can remain interpretable as evidence of capability even as GenAI becomes a normal component of student work, because what is being certified is not tool output but the student's capacity to regulate reliance, verify under uncertainty, and revise judgement when conditions change.

Conclusion

Rebuilding trust in assessment under GenAI is less a matter of detecting misuse than of restoring interpretable evidence of capability. In AI-embracing institutions, credibility depends on making student judgement visible when tool use is legitimate and often encouraged. This paper has argued that mode-switching competence, calibrated decisions about when and how to rely on AI, when to verify, and when to proceed unaided, should be treated as a core assessed capability. Predictable micro-switches offer a practical way to surface this regulation without resorting to continuous surveillance or reducing assessment to platform proficiency. The resulting evidence travels more defensibly across contexts because it certifies judgement, not polish.

References

- Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 49(6), 893-905.
- Corbin, T., Bearman, M., Boud, D., & Dawson, P. (2025a). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*, 1-17.
- Corbin, T., Dawson, P., Nicola-Richmond, K., & Partridge, H. (2025b). 'Where's the line? It's an absurd line': towards a framework for acceptable uses of AI in assessment. *Assessment & Evaluation in Higher Education*, 1-13.
- Digital Education Council, & Pearson. (2025). *The Next Era of Assessment: A global review of AI in assessment design*. Digital Education Council.
- Gonsalves, C. (2024). Generative AI's impact on critical thinking: Revisiting Bloom's taxonomy. *Journal of Marketing Education*, 02734753241305980.
- Krause, S., Panchal, B. H., & Ubhe, N. (2025). Evolution of learning: Assessing the transformative impact of generative AI on higher education. *Frontiers of Digital Education*, 2(2), 21.
- Luo, J., Zheng, C., Yin, J., & Teo, H. H. (2025). Design and assessment of AI-based learning tools in higher education: A systematic review. *International Journal of Educational Technology in Higher Education*, 22(1), 42.
- Oc, Y., Gonsalves, C., & Quamina, L. T. (2025). Generative AI in higher education assessments: Examining risk and tech-savviness on student's adoption. *Journal of Marketing Education*, 47(2), 138-155.
- Sidra, S., & Mason, C. (2025). Generative AI in human-AI collaboration: validation of the collaborative AI literacy and collaborative AI metacognition scales for effective use. *International Journal of Human-Computer Interaction*, 1-25.
- University of Reading, Centre for Quality Support and Development (CQSD). (2025, June). *Generative AI tools and assessment (Version 1.4)*. University of Reading.

The (Un)Ethical Double Standards of the “ChatGPT’s Ember AI-Enhanced Student Assessment”

Vitor Lima
ESCP Business School

Abstract

This paper explores the ethical double standards of generative AI in education through "ChatGPT's Ember," an AI evaluator for business students. By reversing roles, positioning decision-makers as subjects of algorithmic judgment, the study exposes tensions regarding fairness, autonomy, and accountability, reframing AI ethics as a leadership responsibility rather than mere technical compliance.

Keywords: Generative AI (GenAI); Algorithmic Fairness; AI Ethics; Executive Education; Automated Assessment

Extended Abstract

The growing adoption of generative artificial intelligence (GenAI), which is capable of creating new multimodal content (Huang & Rust, 2025; Lima, 2025), has changed how people are selected, sorted, and chosen in organizations for various reasons (Shelby et al., 2023). For example, in many human resources (HR) processes, the use of Automated Tracking Systems (ATS) serves as a means to screen pipelines and provide performance reviews based on predictive scores, all in the name of efficiency and scalability (Susarla et al., 2023). However, such a phenomenon is not exclusive to businesses, as it may not be apparent at first glance and takes many forms and shades in educational contexts as well (Epp & Humphreys, 2025; Grewal et al., 2024). The undeniable, and perhaps understandable, hype behind GenAI runs high when the tools sit on one specific side of the table, the executives' (or students'), to be precise (Tully et al., 2025; Valenzuela et al., 2024). But what happens when the same technology these executives praise turns toward them? This role reversal constitutes the ethical hinge of the “ChatGPT’s Ember AI-Enhanced Student Assessment,” reframed here as an experiential ethics exercise designed for executive education.

At ESCP Business School, the Executive MBA (EMBA) and the Master’s in Management (MiM) classrooms hold special relevance as a research context, because participants are simultaneously decision-makers and current subjects of institutional judgment. By putting

them in front of a GenAI evaluator, such as ChatGPT's Ember, coded as my teaching assistant, the exercise clearly highlights the paradoxical nature of technology that Mick and Fournier (1998) talk about. Despite following internal ethics protocols in the development of the GenAI-powered exercise, including informed consent, compliance with the General Data Protection Regulation (GDPR), and alignment with the EU AI Act, tensions between control and chaos, freedom and enslavement, new and obsolete, competence and incompetence, efficiency and inefficiency are all part of the in-class GenAI-driven experience. Figure 1 features a QR code that can be scanned to direct the reader to a YouTube video detailing the case.

FIGURE 1
ChatGPT's Ember AI-Enhanced Student Assessment



A key phenomenon emerging here is the presence of double standards that shape how students experience moral distance, as questions of merit, fairness, and legitimacy no longer exist as abstractions. Across informal reports and classroom discussions, students repeatedly articulate that using GenAI as part of decision-making processes is perceived as good and morally acceptable as long as the advantages exist for those controlling the system and, to some extent, become empowered by it. Similar asymmetries have been documented elsewhere (e.g., Celiktutan et al., 2024).

One key objective of ChatGPT's Ember as an evaluator is thus ambiguously provocative by design. Students engage with a GenAI teaching assistant that does not replace the human professor but contributes to scoring presentations, cases, or structured memos against visible rubrics, all in accordance with the ethical protocols outlined in the syllabus. The code is simple; the experience is not. Students ask whether the system grasps intent, context, or nuance. After a few iterations, the classroom atmosphere begins to resemble the anxieties of job seekers navigating automated evaluation filters, such as those embedded in ATS. And that is precisely the point. The ethical discussions sparked by the Ember evaluator can be framed within the applied AI ethics tradition (e.g., Dabis & Csáki, 2024; Jobin et al., 2019), which builds upon the foundations of bioethics and information ethics. In particular, Beauchamp and Childress' (2019) principlist approach provides a first analytical layer, as it conceptualizes ethics through a small set of interrelated principles rather than abstract moral theories. This approach was later adapted to AI studies by Floridi et al. (2018) through the AI4People framework. Justice and fairness emerge as primary concerns when EMBA and MiM students face algorithmic evaluations that diverge from their self-assessments. This perception aligns with Jobin et al.'s (2019) large-scale review of AI ethics guidelines, which identifies justice and fairness among the most frequently cited principles, reflecting

widespread concern about discrimination, bias, and unequal treatment in automated systems.

A first ethical dimension emerging from these discussions concerns the interdependence of justice, access, beneficence, and non-maleficence as principles rather than isolated entities. Access is closely intertwined with justice, as Hermann (2022) explains. Here, algorithmic systems tend to privilege people with greater digital, linguistic, or contextual capital, a pattern that becomes visible when students compare their ability to “perform well” under GenAI evaluation. Beneficence, often assumed in managerial discourse as a pathway to efficiency and scalability gains, becomes ethically fragile once students recognize that what benefits organizations does not necessarily benefit individuals equally. Non-maleficence reframes harm not as an exceptional failure but as a systemic risk (e.g., Fleisig et al., 2024; Scott et al., 2024; Scott & Orlikowski, 2025), echoing Floridi et al.’s (2018) argument that avoiding harm requires attention to cumulative and indirect effects rather than isolated outcomes. Together, these principles collide rather than align, exposing ethical tensions that cannot be resolved through procedural compliance alone.

The second ethical dimension centers on autonomy and explicability, which function as enabling conditions for ethical evaluation rather than standalone virtues (Valenzuela et al., 2024). Autonomy, rooted in bioethical traditions of self-determination (Beauchamp & Childress, 2019), becomes contested when students realize that consent to GenAI-mediated assessment is often procedural rather than substantive. Across sessions, a recurring hype-driven dynamic emerges, resembling a “Tulip Mania” logic in which participation is framed as inevitable or necessary to avoid falling behind, regardless of whether the value of GenAI is clearly understood. This phenomenon connects to Floridi et al.’s (2018) notion of meta-autonomy, which emphasizes the importance of retaining the ability to decide when and how decision-making authority is delegated to AI systems. Within this context, a burning question remains unanswered: why should GenAI be embedded into evaluative ecosystems if the benefits ultimately are limited to a set of already privileged actors? While no definitive answer is offered, the question itself becomes pedagogically productive.

One potential path explored during these sessions, and one that has proven particularly promising, involves returning to foundational concepts. Aligning expectations about what AI systems, and GenAI models in particular, are and are not becomes a necessary step toward meaningful AI literacy. Explainable AI, understood as the intelligibility and interpretability of how AI systems function and generate outputs (Li et al., 2025), is positioned not as a technical solution but as a pedagogical starting point. From there, the same evaluative situations can be revisited through different ethical traditions, such as consequentialism or virtue ethics (Dabis & Csáki, 2024; Shelby et al., 2023; Thomsen, 2023), transforming ethics into a strategic lens for reflection and decision-making rather than a checklist for compliance. Without intelligibility, students cannot judge whether an assessment is fair or harmful. Without accountability, responsibility for outcomes becomes diffused across models, designers, and institutions, particularly given the variety of modes of interaction with GenAI systems (Lima, 2025). The Ember exercise renders this diffusion experientially visible. By confronting executives with the same level of opacity faced by job applicants and employees, the exercise exposes how ethical commitments weaken with moral distance, reframing AI ethics as a leadership responsibility grounded in power, agency, and institutional accountability, rather than technical compliance alone.

References

- Celiktutan, B., Klesse, A.-K., & Tuk, M. A. (2024). Acceptability lies in the eye of the beholder: Self-other biases in GenAI collaborations. *International Journal of Research in Marketing*, 41(3), 496-512.10.1016/j.ijresmar.2024.05.006
- Dabis, A., & Csáki, C. (2024). AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI. *Humanities and Social Sciences Communications*, 11(1)10.1057/s41599-024-03526-z
- Epp, A. M., & Humphreys, A. (2025). Collaborating with Generative AI in Consumer Culture Research. *Journal of Consumer Research*, 52(1), 32-48.https://doi.org/10.1093/jcr/ucaf014
- Fleisig, E., Smith, G., Bossi, M., Rustagi, I., Yin, X., & Klein, D. (2024). Linguistic Bias in ChatGPT: Language Models Reinforce Dialect Discrimination. *arXiv preprint arXiv:2406.08818*, 1-24.
- Grewal, D., Guha, A., Beccacece Saturnino, C., & Becker, M. (2024). The Future of Marketing and Marketing Education. *Journal of Marketing Education*, 47(1), 61-77.10.1177/02734753241269838
- Hermann, E. (2022). Leveraging Artificial Intelligence in Marketing for Social Good-An Ethical Perspective. *Journal of Business Ethics*, 179(1), 43-61.10.1007/s10551-021-04843-y
- Huang, M. H., & Rust, R. (2025). The GenAI Future of Consumer Research. *Journal of Consumer Research*, 52(1), 4-17.https://doi.org/10.1093/jcr/ucaf013
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.10.1038/s42256-019-0088-2
- Li, B., Lai, E. Y., & Wang, X. S. (2025). From Tools to Agents: Meta-Analytic Insights into Human Acceptance of AI. *Journal of Marketing*, https://doi.org/10.1177/00222429251355266https://doi.org/10.1177/00222429251355266
- Lima, V. M. (2025). A sociosemiotics mapping of consumer-AI agent interactions in consumption journeys. *ESCP Impact Papers*, 2025-29-EN, 1-7.https://dx.doi.org/10.2139/ssrn.5467387
- Mick, D. G., & Fournier, S. (1998). Paradoxes of technology: Consumer cognizance, emotions, and coping strategies. *Journal of Consumer Research*, 25(September), 123-143.
- Scott, M. L., Bone, S. A., Christensen, G. L., Lederer, A., Mende, M., Christensen, B. G., & Cozac, M. (2024). Revealing and Mitigating Racial Bias and Discrimination in Financial Services. *Journal of Marketing Research*, 61(4), 598-618.10.1177/00222437231176470
- Scott, S. V., & Orlikowski, W. J. (2025). Exploring AI-in-the-making: Sociomaterial genealogies of AI performativity. *Information and Organization*, 35(1)https://doi.org/10.1016/j.infoandorg.2025.100558
- Shelby, R., Rismani, S., Henne, K., Moon, A., Rostamzadeh, N., Nicholas, P., Yilla-Akbari, N. M., Gallegos, J., Smart, A., Garcia, E., & Virk, G. (2023). Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society,
- Susarla, A., Gopal, R., Thatcher, J. B., & Sarker, S. (2023). The Janus Effect of Generative AI: Charting the Path for Responsible Conduct of Scholarly Activities in Information Systems. *Information Systems Research*, 34(2), 399-408.https://doi.org/10.1287/isre.2023.ed.v34.n2
- Thomsen, F. K. (2023). Algorithmic indirect discrimination, fairness and harm. *AI and Ethics*, 10.1007/s43681-023-00326-010.1007/s43681-023-00326-0

Tully, S., Longoni, C., & Appel, G. (2025). Lower Artificial Intelligence Literacy Predicts Greater AI Receptivity. *Journal of Marketing*, 10.1177/0022242925131449110.1177/00222429251314491

Valenzuela, A., Puntoni, S., Hoffman, D., Castelo, N., De Freitas, J., Dietvorst, B., Hildebrand, C., Huh, Y. E., Meyer, R., Sweeney, M. E., Talaifar, S., Tomaino, G., & Wertenbroch, K. (2024). How Artificial Intelligence Constrains the Human Experience. *Journal of the Association for Consumer Research*, 9(3), 241-256.10.1086/730709242

Aligning tech and sustainability in Higher Education: protecting the environment, safeguarding values, and upholding critical thinking

Gorgi Krlev, Associate Dean and Professor of Sustainability at ESCP Business School

Laetitia Langlois, Sustainability Project Manager (ESCP)

Gabrielle Tremblay, Ecological Transition Project Manager (ESCP)

In close collaboration with ESCP AI Team and Alara Tascioglu.

Abstract

Debates on artificial intelligence (AI) and sustainability often oscillate between two polarized narratives: AI as a solution to environmental and social crises, or AI as a threat to planetary boundaries and human well-being. While both perspectives highlight important concerns, their opposition oversimplifies AI as a socio-technical system embedded in material infrastructures and institutional practices.

This interactive workshop proposes a pedagogical reframing: teaching AI and sustainability not as a binary moral judgment of “good” versus “bad” technology, but as a field of accountability, trade-offs, and creative agency. Drawing on the ESCP Sustainability Fingerprint and existing AI training initiatives, the session explores how to integrate environmental, social, and ethical dimensions of AI into higher education.

By emphasizing systems thinking, contextual evaluation, and learning when not to use AI, the workshop aims to equip educators with tools to cultivate responsible, critical, and creative AI literacy in higher education.

Keywords: Artificial Intelligence; Sustainability Education; Responsible Innovation; AI Literacy; Systems Thinking

1. Motivation

Concerns about AI’s environmental footprint—energy use, water consumption, resource extraction—are increasingly documented in scientific and policy literature (e.g., IEA; Nature Energy; OECD). At the same time, AI is actively mobilized for climate modeling, biodiversity monitoring, energy optimization, and adaptation strategies (IPCC AR6).

In educational settings, this tension often translates into **fragmented teaching approaches**:

- technical courses detached from environmental consequences,
- sustainability courses treating AI as a black box,
- ethical discussions reduced to abstract principles or alarmist narratives.

The result is not necessarily ignorance, but pedagogical simplification. Students may learn what to think about AI (enthusiasm or rejection) rather than how to think with and about AI in relation to sustainability.

This workshop responds to a growing need for **synergetic AI literacy**, where students are equipped to:

- understand AI's material and social embeddedness,
- evaluate its impacts contextually,
- decide *when, how, and whether* to use AI responsibly.

2. Goals: Teaching accountability for users while enabling creativity and positive impact

A. Teaching the environmental impact of AI

A first step toward accountability is to challenge the widespread perception of AI as a purely digital or immaterial layer. AI is a **physical technology**, deeply embedded in infrastructures:

- data centers and energy grids,
- cooling systems and water resources,
- extractive supply chains for hardware and rare minerals.

Teaching AI through a **systems-thinking lens** invites students to see it as part of interconnected socio-technical and ecological systems rather than as a standalone tool. The objective is not to moralize AI use, but to ground discussions in science-based understanding of resource constraints and environmental externalities. Key pedagogical elements include:

- AI and energy: understanding energy intensity, carbon intensity of electricity mixes, and efficiency trade-offs (IEA; Nature Energy).
- AI and water: cooling requirements of data centers and regional water stress (OECD; academic case studies).
- AI and material resources: hardware lifecycles, e-waste, and extraction impacts (UNEP).

B. Teaching the social and ethical implications of AI

Environmental impacts cannot be separated from social and ethical dimensions. Relevant teaching angles include:

- Social issues in extraction regions, where mineral sourcing for AI hardware intersects with labor conditions, geopolitical dependency, and environmental injustice.
- Data usage and creative labor, including the appropriation of creators' content and unresolved questions of consent, compensation, and intellectual property.
- Mental health and cognitive impacts, particularly in relation to automation of cognitive tasks, dependency risks, and changes in learning practices.

Here again, the goal is not to condemn AI, but to **situate responsibility**: who benefits, who bears the costs, and how these costs are distributed across societies and generations.

C. Teaching AI as an enabler of sustainability

Avoiding binary narratives also means recognizing AI's potential contributions—when used deliberately and appropriately. Teaching AI as an enabler requires emphasizing conditions of use, not abstract promises. Pedagogical examples include:

- **AI as a tool for scientific research**, accelerating climate modeling, material discovery, or ecosystem monitoring (IPCC; Nature Climate Change).
- **AI for mitigation**, such as energy efficiency, grid optimization, or industrial process improvement.
- **AI for adaptation**, including risk forecasting, disaster preparedness, and urban resilience.
- **AI for cultural and narrative transformation**, supporting new ways of communicating sustainability challenges.

D. Teaching how—and when—to use AI

A crucial but often neglected dimension is **frugality**: learning when *not* to use AI. This reframes AI from a productivity shortcut to a pedagogical object—something students learn with, about, and against when necessary. Key pedagogical components include:

- **AI literacy**: understanding that AI systems can be wrong, biased, opaque, or misleading.
- **Judgment and proportionality**: matching the tool to the task rather than defaulting to AI use.
- **Creativity and experimentation**: treating AI as a sandbox or lab for exploration, not as an authority.
- **Knowledge creation and sharing**: using AI to co-create, adapt, and disseminate knowledge responsibly (e.g. custom GPTs for curricula, with transparency and critical oversight).

3. Questions for collective discussion

This workshop is designed as a space for dialogue rather than prescription. Questions include:

- How can we teach AI's environmental impacts without slipping into technological determinism or moral panic?
- What does “responsible AI use” mean in practice for students and educators?
- How do we reconcile creativity, experimentation, and restraint in AI pedagogy?
- Should AI literacy include learning to refuse, limit, or redesign AI usage?
- How can sustainability education benefit from AI without reinforcing extractive or inequitable models?

4. Pedagogical materials to enrich demonstration and discussion

1. Progress on **custom-GPT project** to enhance curriculum revolution (embed sustainability in all disciplines, from accounting to sales)
2. ESCP-wide **policies and indicators** to boost responsible AI use.
3. Best-practice cases of **AI for sustainability in research**.
4. Pedagogical **training videos** for responsible use by students.

Closing note

Teaching AI and sustainability is not about choosing sides in a technological debate. It is about cultivating discernment, responsibility, and imagination in a context of planetary constraints and rapid technological change. By moving beyond binary narratives, educators can help students navigate complexity—and act within it.

GenAI Adoption in Business Schools: An Interactive Qualitative Analysis in a Developing Economy Context

Prakul Jain, Research Scholar, Department of Management Studies, Indian Institute of Technology Madras

Prakash Sai Lokachari, Professor, Department of Management Studies, Indian Institute of Technology Madras

Abstract

The study examines factors influencing GenAI use for learning and academic work among business school (B-School) students. Taking the case of a premier B-School in a developing economy, it utilises the systems method of Interactive Qualitative Analysis (IQA) to model causal interlinkages among the factors at cognitive, affective and structural levels. Findings are relevant to policymakers and educators, and scholars interested in GenAI in higher education.

Keywords: GenAI Adoption, Higher Education, Business School, Interactive Qualitative Analysis (IQA)

Introduction

Generative AI (GenAI) tools such as ChatGPT, Claude, Gemini, and Copilot have entered academic life, promising personalised learning and productivity enhancement. As GenAI rapidly reshapes how institutions learn and create value, understanding its adoption in knowledge-intensive sectors becomes essential. Higher education institutions (HEI), at the confluence of academia and research, emerge both as centres of innovation and experimentation. With students worldwide keen to improve their learning outcomes, the potential of GenAI to reshape knowledge creation, assessment, and cognitive engagement has been widely recognised (Popenici & Kerr, 2017; Dhakal, 2025). In this context, HEI have the potential to be a living “laboratory” for understanding how complex knowledge work evolves when GenAI becomes integrated into daily practice. As education often shapes the formative as well as the occupational years of a human, the impact of GenAI among learners at all levels is particularly decisive.

Extant literature reveals an evolving understanding of GenAI in higher education, with key research areas revolving around pedagogical augmentation, ethical concerns, and institutional readiness analyses (Sahar & Munjiati, 2025; Belkina et al., 2025). However, most studies rely on quantitative methods such as Structural Equation Modelling (SEM) and Regression Analysis, which, while statistically powerful, lack the interpretive depth needed to capture the adoption process of a rapidly evolving technology such as GenAI. This leaves mutual causality observed in real learner experiences, for example, how trust influences ethics or how self-efficacy shapes psychological comfort, sorely understudied. Further, context-specific research that incorporates the human side of adoption is less understood from a student perspective (Belkina et al., 2025; Ivanov et al., 2024; Sharma et al., 2024). A paradox particularly noticeable in the ethics-trust-efficacy nexus, despite the area being one of the most dominant themes (Sahar & Munjiati, 2025; Belkina et al., 2025). This is further exacerbated by the lack of studies in Global South contexts, which are fast emerging as hotbeds of GenAI adoption, integration and investment (Zhao et al., 2024; Henadirage & Gunarathne, 2025). Hence, there is a need for scholarship on GenAI in HEI to move beyond isolative approaches and contribute to the explainability of the interconnectedness and complexity of the human-GenAI engagement system (Hughes et al., 2025; Dhakal, 2025).

In India's HEI, the integration of AI is emphasised under the National Education Policy (NEP) 2020. However, adoption research remains limited. Jasrai (2025) and Sharma et al. (2024) identify digital self-efficacy and institutional facilitation as drivers of GenAI use, whereas ethical ambiguity and plagiarism fears remain inhibitors. The role of individual learners' confidence, awareness, and psychological comfort is also an area that is not sufficiently discussed (Sharma et al., 2024). Similarly, quantitative studies model adoption behaviour through the Technology Acceptance Model (TAM), Unified Theory of Acceptance and Use of Technology (UTAUT), and Theory of Planned Behaviour (TPB) (Ivanov et al., 2024; Jasrai, 2025). These frameworks quantify predictors like performance expectancy, ease of use, and social norms, but often ignore interdependencies among cognitive, emotional, and contextual factors (Kapur et al., 2025; Wang et al., 2023).

The current study aims to address these gaps by utilising the systems method of Interactive Qualitative Analysis (IQA) to understand the causal interlinkages among factors underlying the adoption of GenAI for learning and academic work. IQA captures not only the presence of constructs but also their causal relationships, offering a holistic understanding that addresses the isolative approaches in the extant literature (Hughes et al., 2025; Dhakal, 2025; Ivanov et al., 2024). While the IQA exercise models the interlinkages of GenAI adoption among students, the causal interpretations deepen understanding at the cognitive, affective and structural (CAS) layers. Against this backdrop, the study proposes the following objectives:

1. Analyse the factors that influence GenAI adoption for learning and academic work, taking the case of business school students.
2. Evaluate the interrelationships between the factors that influence GenAI adoption for learning and academic work, taking the case of business school students.
3. Develop a conceptual framework for maximising learning outcomes among business school students through GenAI adoption at cognitive, affective and structural levels.

The study's results are relevant to policymakers and educators working in HEIs, as well as scholars seeking applied knowledge of GenAI in business school settings. In terms of novelty, the study is among the first to combine a systems approach (through IQA) to develop a multi-layered (CAS) framework in a world that is evolving and co-creating with GenAI.

Literature Background

Extant literature (Kapur et al., 2025; Wang et al., 2023) emphasises that the psychological impact is a new frontier variable encompassing motivation, curiosity, and anxiety in the context of GenAI research. This is confirmed by systematic reviews and bibliometric analyses (Sahar & Munjiati, 2025; Belkina et al., 2025), which, across 30+ studies, identify self-efficacy, trust, and ethics as dominant themes. Studies from Sri Lanka and Turkey (Henadirage & Gunarathne, 2025; Ursavaş et al., 2025) also emphasise structural factors, such as policy enablement and personalisation, as cultural adaptations in the context of GenAI adoption. Similarly, policy enablement, through institutional governance and regulatory support, directly influences the intention to adopt AI technologies (Hughes et al., 2025; Dhakal, 2025). Self-efficacy has been validated as a foundational driver of technology engagement and creative confidence through 14 studies, including Wang et al. (2023), Shahzad et al. (2025), and Singh et al. (2025), and, along with perceived usefulness, was also found to be a strong predictor of AI use (Chatterjee & Bhattacharjee, 2020; Sharma et al., 2024; Ivanov et al., 2024). Trust in AI tools has been explored longitudinally by Song (2025), demonstrating that trust evolves with increasing performance satisfaction and mediates perceptions of ethics and continued use, despite ongoing ethical challenges such as plagiarism, fairness, and privacy highlighted by Dhakal (2025) and Kapur et al. (2025). Similarly, ethical anxiety and trust deficits consistently moderate adoption intention (Hughes et al., 2025; Hsiao et al., 2025). The diversity among AI tools has been recognised in cross-tool comparison studies by Jasrai (2025) and Belkina et al. (2025) as shaping user preferences, with customised usage reflecting the personal adaptation of AI features for individualised learning (Henadirage, 2025; Ouyang, 2022). Lastly, cross-country comparisons (Zhao et al., 2024) suggest contextual variability in emerging economies, where adoption depends more on policy clarity and literacy support than on technological sophistication.

Methodology

This study utilises the systems method of IQA developed by Northcutt and McCoy (2004), grounded in participant insights and the grounded theory approach, to achieve its research objectives (refer to Figure 1 for a complete overview). The study sample consisted of frequent GenAI users across disciplines such as marketing, strategy, operations and finance. Taking the case of students at a premier business school in India, the analysis consisted of a focus group discussion (n = 12) and semi-structured interviews (n=8) on the brainstorming question: **“What factors govern your adoption of Generative AI for learning and academic work?”**.

Result and Analysis

A summary of the results of the IQA process, with the identified affinities and their interlinkages, is given in Figure 1 below:

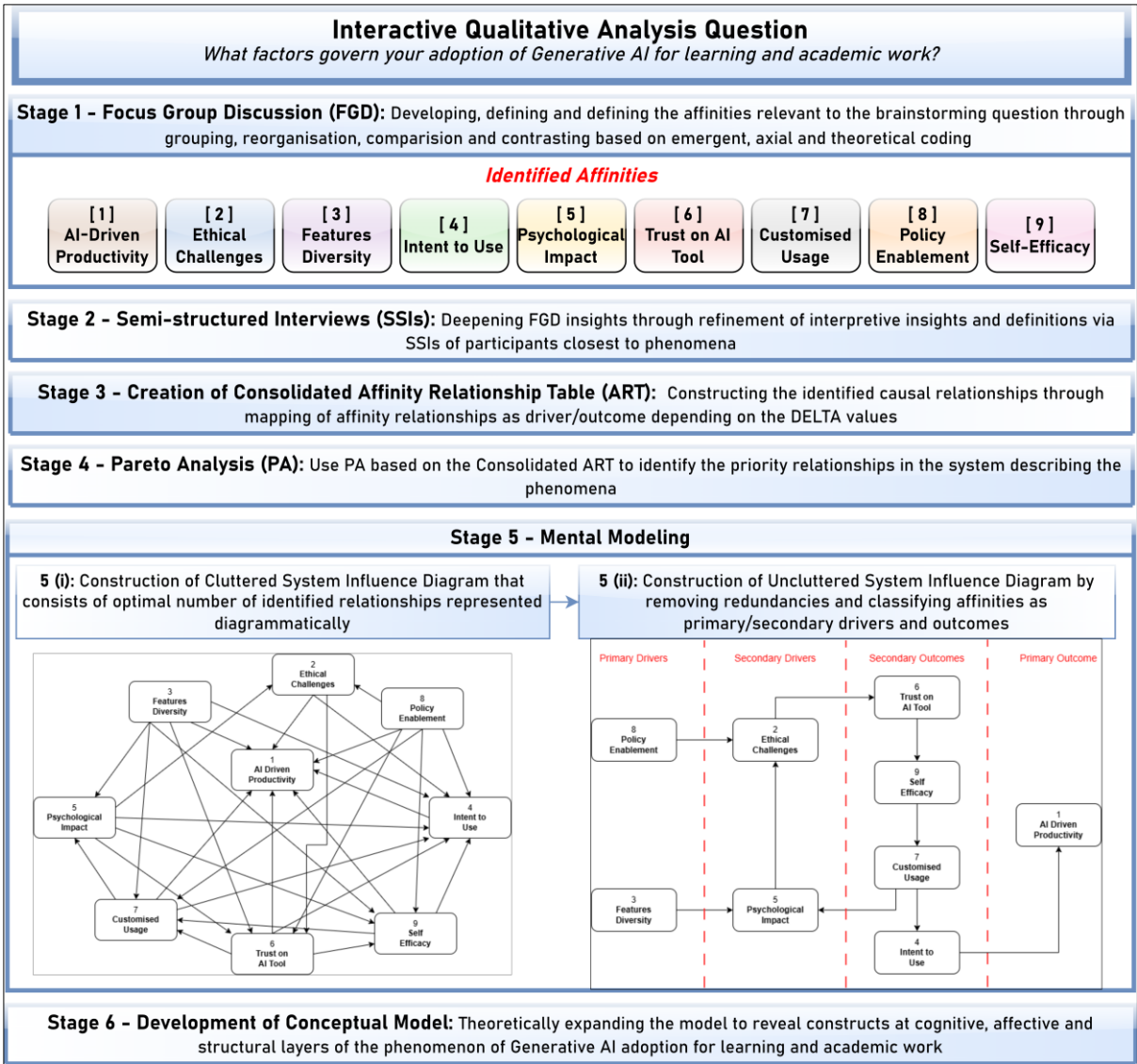


Figure 1: Overview of the IQA process and results

Source: Authors' own compilation

The study finds that self-efficacy, trust, and ethics are core to GenAI adoption for learning. When supported by policy and feature diversity, these lead to adaptive behaviours like customised use and productive outcomes. While TAM sees 'perceived usefulness' as a simple predictor, in India, it results from psychological and ethical balance. Social norms, important in Western contexts, are negligible here (Ursavaş et al., 2025), showing Indian learners' cultural independence and self-regulation. This pattern extends models like UTAUT and TPB. Instead of linear cause-and-effect, IQA shows recursive loops: high trust boosts productivity, which reinforces trust. Adoption depends on psychological stability, morality, and readiness, not just perceived benefits. As GenAI evolves, its adoption resembles a multi-layered ecosystem of cognitive, emotional, and structural factors (see Figure 2).

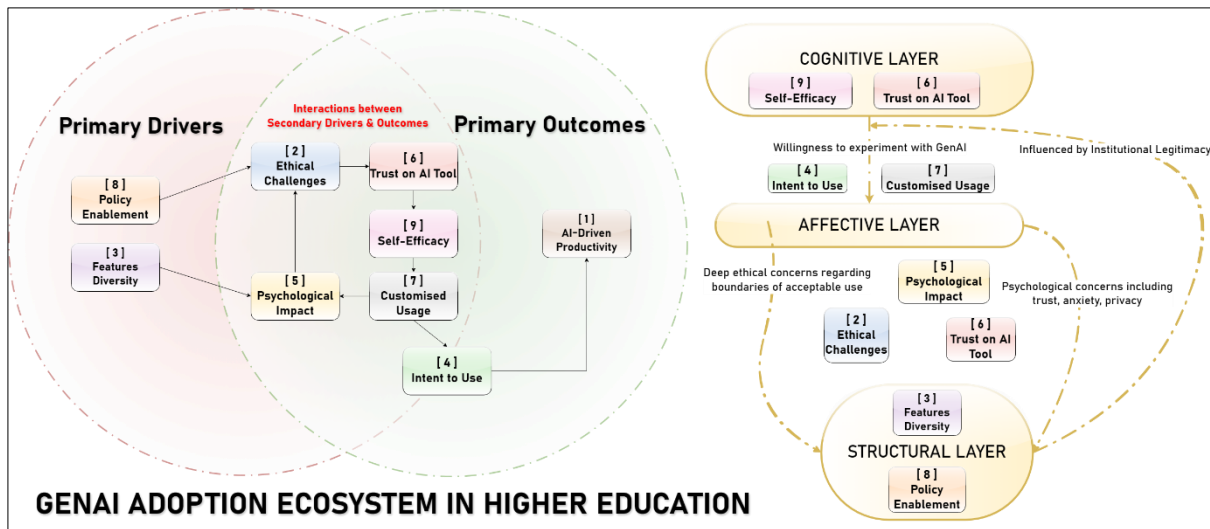


Figure 2: Proposed conceptual framework

Source: Authors' own compilation

Implications

Given the rapidly evolving field of GenAI, scholarship on its impactful adoption is both timely and meritorious. Accordingly, the current study has implications across theoretical, practical, and policy domains. Theoretically, the study integrates UTAUT, TPB, and SCT constructs into a unified systems model, enriching the quantitative-dominant GenAI adoption discourse with qualitative causality (Chatterjee & Bhattacharjee, 2020; Sharma et al., 2024; Ivanov et al., 2024). Compared to extant research, this IQA approach advances understanding by integrating these layers into a single dynamic system. Practically, the insights offered can guide universities in strategically planning for AI literacy, ethical transparency, and emotional resilience, rather than focusing solely on technical training. It also complements the literature on AI-driven productivity, which has documented improvements in efficiency and assessment quality, as reported by Ouyang and Zheng (2022), Aithal (2023), and Singh et al. (2025). Finally, policy implications include insights for effectively embedding student voice in AI governance and identifying intervention-lever points for HEIs to enhance learning outcomes and improve organisational performance.

Conclusion

The study examined factors influencing GenAI use for learning and academic work among business school (B-School) students. The use of IQA based on organic participant understanding (i.e., students closest to the phenomenon of actual GenAI use for education and learning) enabled a learner-centred, systems-based understanding of GenAI adoption that includes CAS layers. However, the study has limitations that future scholarship can

address. First, participant choice was limited to management students, which future studies should expand for broader applicability. Second, the evolving nature of GenAI tools requires ongoing revalidation, and future work could explore this longitudinally. Lastly, the IQA model could be empirically validated with SEM or PLS techniques to confirm directional relationships.

Disclosure Statement

During the preparation of this work, the authors used Grammarly v1.2.223.1804 to improve the language of the manuscript. The authors have reviewed and edited the manuscript content as needed after using the tools. The authors take full responsibility for the contents of the article.

References

Aithal, P. S., & Aithal, S. (2023). Application of ChatGPT in higher education and research–A futuristic analysis. *International Journal of Applied Engineering and Management Letters (IJAEML)*, 7(3), 168-194. <http://dx.doi.org/10.2139/ssrn.4674364>

Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., Grundy, S. & Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. *Computers and Education: Artificial Intelligence*, 100407. <https://doi.org/10.1016/j.caeai.2025.100407>

Chatterjee, S., & Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling. *Education and Information Technologies*, 25(5), 3443-3463. <https://doi.org/10.1007/s10639-020-10159-7>

Dhakal, S. P. (2025). A Scoping Review of Generative Artificial Intelligence (GenAI) and Pedagogy Nexus: Implications for the Higher Education Sector. *Metrics*, 2(3), 17. <https://doi.org/10.3390/metrics2030017>

Henadirage, A., & Gunarathne, N. (2025). Barriers to and opportunities for the adoption of generative artificial intelligence in higher education in the global south: Insights from Sri Lanka. *International Journal of Artificial Intelligence in Education*, 35(1), 245-281. <https://doi.org/10.1007/s40593-024-00439-5>

Hsiao, C. H., & Tang, K. Y. (2025). Beyond acceptance: an empirical investigation of technological, ethical, social, and individual determinants of GenAI-supported learning in higher education. *Education and Information Technologies*, 30(8), 10725-10750. <https://doi.org/10.1007/s10639-024-13263-0>

Hughes, L., Malik, T., Dettmer, S., Al-Busaidi, A. S., & Dwivedi, Y. K. (2025). Reimagining higher education: Navigating the challenges of generative AI adoption. *Information Systems Frontiers*, 1-23. <https://doi.org/10.1007/s10796-025-10582-6>

- Ivanov, S., Soliman, M., Tuomi, A., Alkathiri, N. A., & Al-Alawi, A. N. (2024). Drivers of generative AI adoption in higher education through the lens of the Theory of Planned Behaviour. *Technology in Society*, 77, 102521. <https://doi.org/10.1016/j.techsoc.2024.102521>
- Jasrai, L. (2025). Extending UTAUT model to examine the usages of ChatGPT among Indian students in higher education: a structural equation modelling approach. *The TQM Journal*. <https://doi.org/10.1108/TQM-12-2024-0498>
- Kapur, I., Sharkey, C. N., & Ren, C. (2025). Student Perspectives on Artificial Intelligence: Challenges, Opportunities, and Societal Implications. *Journal of Evidence-Based Social Work*, 1-27. <https://doi.org/10.1080/26408066.2025.2517073>
- Northcutt, N., & McCoy, D. (2004). *Interactive qualitative analysis: A systems method for qualitative research*. Sage.
- Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, 27(6), 7893-7925. <https://doi.org/10.1007/s10639-022-10925-9>
- Popenici, S. A., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and practice in technology enhanced learning*, 12(1), 22. <https://doi.org/10.1186/s41039-017-0062-8>
- Sahar, R., & Munawaroh, M. (2025). Artificial intelligence in higher education with bibliometric and content analysis for future research agenda. *Discover Sustainability*, 6(1), 401. <https://doi.org/10.1007/s43621-025-01086-z>
- Shahzad, M. F., Xu, S., & Zahid, H. (2025). Exploring the impact of generative AI-based technologies on learning performance through self-efficacy, fairness & ethics, creativity, and trust in higher education. *Education and Information Technologies*, 30(3), 3691-3716. <https://doi.org/10.1007/s10639-024-12949-9>
- Sharma, S., Singh, G., Sharma, C. S., & Kapoor, S. (2024). Artificial intelligence in Indian higher education institutions: a quantitative study on adoption and perceptions. *International Journal of System Assurance Engineering and Management*, 1-17. <https://doi.org/10.1007/s13198-023-02193-8>
- Singh, R., Singh, S. K., & Mishra, N. (2025). Factors influencing student learning performance and continuous use of artificial intelligence in online higher education. *Discover Education*, 4(1), 292. <https://doi.org/10.1007/s44217-025-00728-8>
- Song, D. (2024). How Learners' Trust Changes in Generative AI over a Semester of Undergraduate Courses. *International Journal of Artificial Intelligence in Education*, 1-20. <https://doi.org/10.1007/s40593-024-00446-6>
- Ursavaş, Ö. F., Yalçın, Y., İslamoğlu, H., Bakır-Yalçın, E., & Cukurova, M. (2025). Rethinking the importance of social norms in generative AI adoption: investigating the acceptance and use of generative AI among higher education students. *International Journal of Educational Technology in Higher Education*, 22(1), 38. <https://doi.org/10.1186/s41239-025-00535-z>
- Wang, S., Sun, Z., & Chen, Y. (2023). Effects of higher education institutes' artificial intelligence capability on students' self-efficacy, creativity and learning performance. *Education and Information Technologies*, 28(5), 4919-4939. <https://doi.org/10.1007/s10639-022-11338-4>

Zhao, L., Rahman, M. H., Yeoh, W., Wang, S., & Ooi, K. B. (2025). Examining factors influencing university students' adoption of generative artificial intelligence: a cross-country study. *Studies in Higher Education*, 50(12), 2646-2668. <https://doi.org/10.1080/03075079.2024.2427786>

Human-Centered AI Pedagogy: World Language Education as a Scalable Model for Higher Education

Jinai Sun
Professor of Chinese,
Department of World
Languages & Cultures
North Central College,
U.S.

Abstract

The rapid development of generative artificial intelligence (AI) is transforming not only industries but also the fundamental structure of higher education. As automation increasingly replaces repetitive cognitive tasks, the value of human contribution is shifting toward interpretation, ethical reasoning, intercultural communication, and creative problem-solving. This paper argues that world language education provides a scalable model for human-centered AI pedagogy across higher education. Drawing on classroom innovation, AI-integrated language instruction, and interdisciplinary curriculum design, this study proposes a framework that positions AI as a cognitive partner rather than a substitute for human agency. By integrating intercultural competence, ethical reflection, and AI literacy, world language education demonstrates how institutions can redesign learning for the AI era. The future of higher education depends not on competing with AI, but on cultivating distinctly human capacities that technology cannot replicate.

Keywords: Human-centered AI; Liberal arts education; World Language education; AI pedagogy; Intercultural competence

Reimagining Higher Education in the Age of Automation

The emergence of generative AI has accelerated a structural shift in both the global workforce and higher education. Tasks that rely on pattern recognition, language production, and data synthesis can now be partially or fully automated. However, automation does not eliminate work; it redistributes value. As prediction becomes cheaper, judgment becomes more valuable (Agrawal et al., 2018). In this evolving landscape, educational institutions must prepare students not merely to perform tasks, but to exercise informed, ethical, and context-sensitive judgment.

World language education offers a compelling model for this transformation. Language learning inherently integrates communication, cultural analysis, perspective-taking, and reflective interpretation. As Byram (1997) argues, intercultural communicative competence requires not only linguistic proficiency but also attitudes of openness, interpretive skills,

relational understanding, and critical cultural awareness. In an AI-mediated world where information is abundant but contextual understanding is fragile, these intercultural capacities become essential forms of human expertise.

Human-centered AI frameworks further emphasize that technology should augment rather than replace human agency (Li, 2023). When designed responsibly, AI systems expand human creativity and analytical reach while preserving ethical accountability. In language classrooms, AI tools can automate vocabulary drills or grammar feedback, enabling instructors to focus on dialogue, debate, intercultural analysis, and applied communication. As Danesi (2024) suggests, generative AI functions most productively as a cognitive partner—supporting metalinguistic awareness and reflective thinking rather than bypassing intellectual effort. The central question for higher education is therefore not whether AI will replace educators, but how institutions can redesign pedagogy to elevate uniquely human capacities.

A Scalable Human-Centered AI Pedagogical Framework

To prepare students for AI-driven transformation, curriculum design must shift from content transmission to capability development. World language education provides a scalable framework for implementing this shift across disciplines through four interrelated pillars:

People (Stakeholders). Prioritize empathy, collaboration, intercultural competence, and ethical responsibility. These relational capacities ground technological fluency in human values.

Curriculum. Integrate AI literacy across programs while preserving humanistic inquiry. Students must understand how AI systems function, where bias emerges, and how technology intersects with culture, language, and power (Holmes & Bialik, 2019).

Instruction. Use AI tools strategically to enhance—not substitute—critical engagement. Assignments should require students to analyze AI-generated outputs, question assumptions, and contextualize information within social and cultural frameworks.

Assessment. Evaluate synthesis, ethical reasoning, intercultural perspective-taking, and reflective judgment rather than procedural recall.

In practice, students may use AI to generate preliminary drafts or simulate intercultural scenarios, but they must critique outputs, identify bias, and situate knowledge within broader human contexts. The human contribution shifts from producing information to interpreting meaning, exercising judgment, and designing responsible applications.

Because world language education already integrates communication, culture, reflection, and applied knowledge, it offers a replicable blueprint for AI-era curriculum redesign. Rather than treating AI integration as a technical add-on, this model embeds AI literacy within relational and ethical learning structures.

Human-centered AI pedagogy therefore positions world language education not as peripheral to innovation, but as central to the future of higher education. In the AI age, competitive advantage will not derive from speed alone, but from depth, discernment, intercultural intelligence, and responsible judgment.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction machines: The simple economics of artificial intelligence. Harvard Business Review Press.

Byram, M. (1997). Teaching and assessing intercultural communicative competence. Multilingual Matters.

Danesi, M. (2024). AI in foreign language learning and teaching: Theory and practice. Nova Science Publishers.

Holmes, W., & Bialik, M. (2019). Artificial intelligence in education: Promise and implications for teaching and learning. Center for Curriculum Redesign.

Li, F.-F. (2023). The worlds I see: Curiosity, exploration, and discovery at the dawn of AI. Flatiron Books.

AI Academic Navigator: A Case Study of GPT-4 Powered Student Assessment in Higher Education

Jan Saro, Czech University of Life Sciences Prague, Faculty of Economics and Management

Abstract

This case study presents the AI Academic Navigator, a web-based platform leveraging GPT-4 for comprehensive cognitive assessment and personalized learning recommendations in higher education. Traditional psychometric assessments suffer from delayed feedback, generic interpretation, and low student engagement. Our platform addresses these limitations by integrating seven validated psychological instruments—including the Stroop Test, Mental Rotation Test, Grit Scale, and RIASEC career inventory—within an AI-guided interface that provides immediate, personalized synthesis. Deployed at the Czech University of Life Sciences Prague with 120 first-year business students, the system achieved 74% completion rate, significantly exceeding typical paper-based assessments. The privacy-first architecture stores all assessment data client-side, addressing GDPR requirements while enabling rich AI-powered personalization. We present the system architecture, user experience flow, implementation results, and key lessons learned for educators considering AI integration in student assessment.

Keywords: artificial intelligence, cognitive assessment, personalized learning, GPT-4, higher education

Context and Motivation

Higher education institutions increasingly recognize the importance of understanding individual student differences for effective teaching. Traditional psychometric assessments, while scientifically validated, present significant practical limitations in classroom settings. Students typically receive delayed feedback—often days or weeks after completion—reducing the immediate relevance of results. Interpretations tend to be generic, failing to connect findings to students' specific academic contexts or career aspirations. When multiple instruments are administered, results remain fragmented across separate reports, leaving students unable to see meaningful patterns across their cognitive profile.

Perhaps most critically, completion rates for traditional paper-based assessments rarely exceed 50% in voluntary settings, as students perceive them as bureaucratic exercises disconnected from their learning experience. The emergence of large language models, particularly GPT-4, presented an opportunity to address these limitations—not by replacing

validated instruments, but by enhancing interpretation, integration, and personalization in ways previously impossible at scale.

System Architecture

The AI Academic Navigator follows a privacy-first, client-centric architecture. Assessment data remains on the student's device, with only anonymized prompts sent to GPT-4 for synthesis. The technology stack includes: React 19 with TypeScript for the frontend, OpenAI GPT-4 API for AI integration, client-side localStorage for GDPR-compliant data storage, and a bilingual Czech/English interface.

Integrated Assessments

The platform incorporates seven validated psychological instruments: the Stroop Test (Stroop, 1935) measuring selective attention and cognitive flexibility; Mental Rotation Test (Shepard & Metzler, 1971) assessing spatial visualization ability; VAK Assessment (Fleming, 1987) identifying Visual/Auditory/Kinesthetic learning preferences; Grit Scale (Duckworth, 2007) measuring perseverance and passion; Growth Mindset inventory (Dweck, 2006) assessing beliefs about intelligence malleability; RIASEC (Holland, 1959) for career interest typology; and Emotional Intelligence measures (Goleman, 1995) evaluating self and social awareness.

Platform Interface

The dashboard presents all available assessments as interactive cards (Figure 1). Students can complete tests in any order, with visual progress indicators showing completion status.

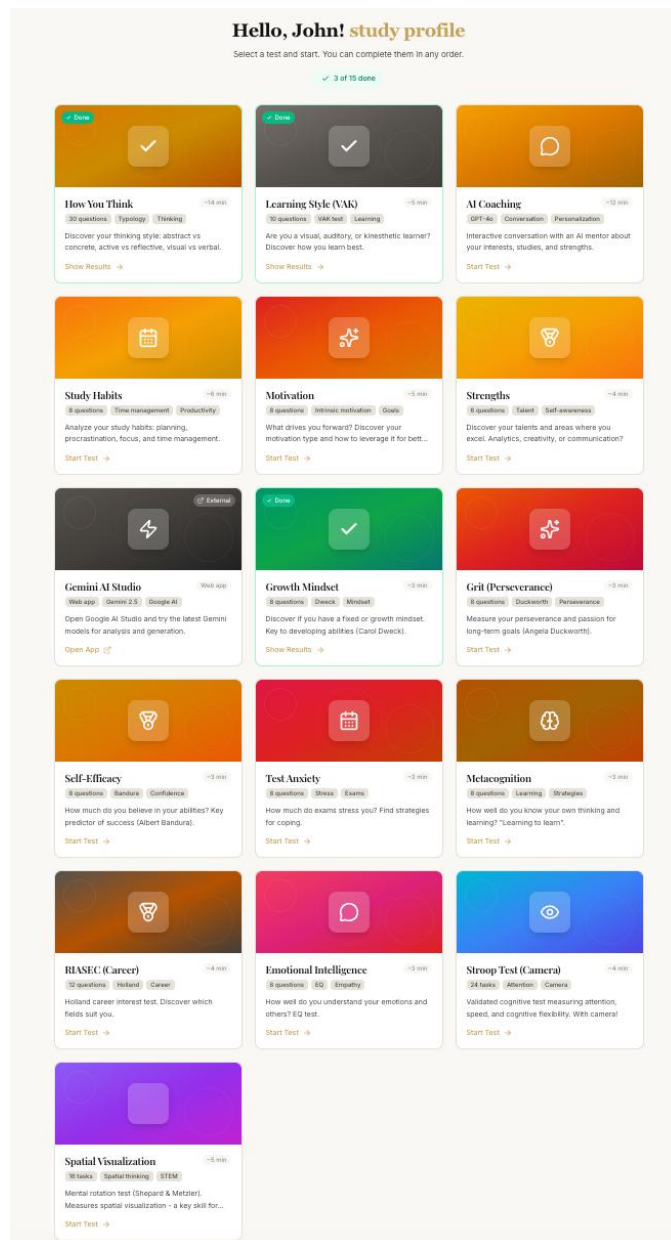


Figure 1: Assessment Dashboard with 16 cognitive and metacognitive tests

Beyond structured assessments, students engage in a GPT-4 powered conversational interview (Figure 2). The AI explores interests, study habits, and career aspirations through natural dialogue, generating a comprehensive "Student Passport" report.

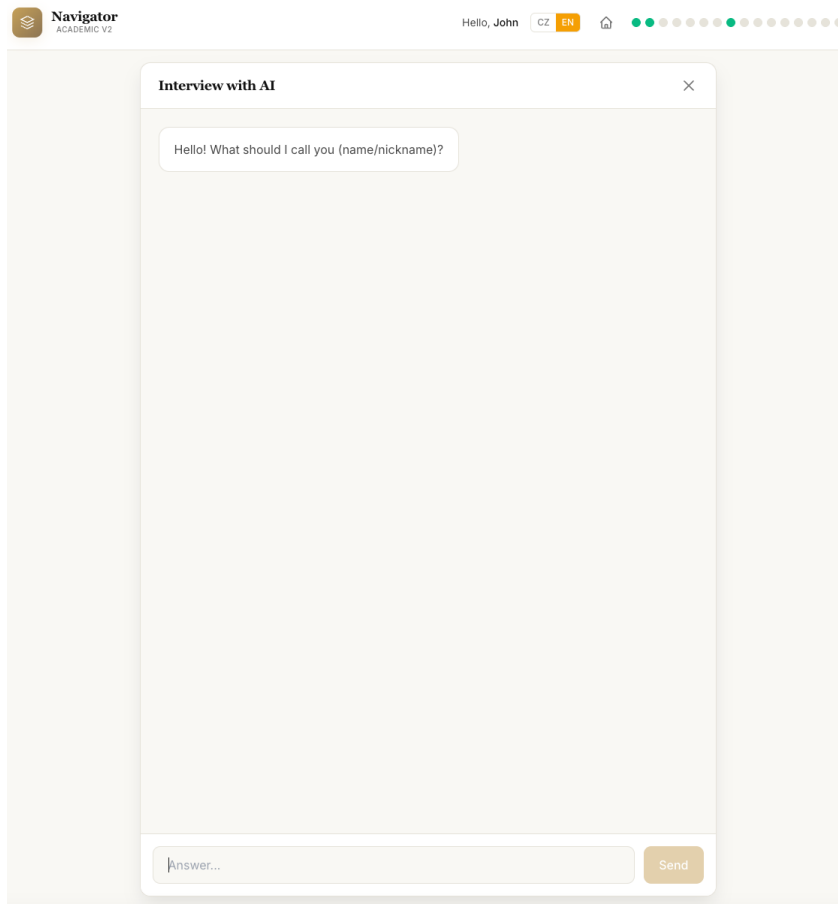


Figure 2: AI Interview - Natural conversation exploring student interests and goals

After completing assessments, GPT-4 synthesizes all results into a coherent learning profile (Figure 3), identifying patterns across instruments—for example, correlating visual learning preference with spatial ability scores and career interests.

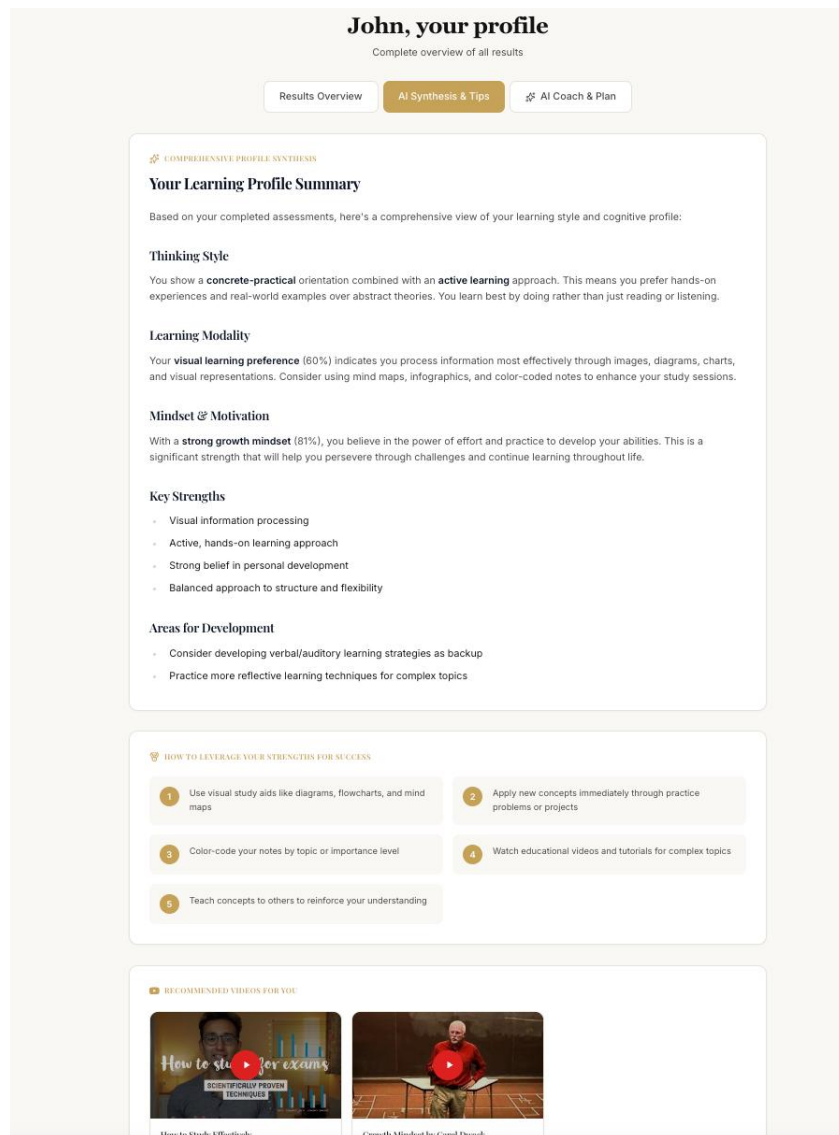


Figure 3: AI Synthesis - Integrated analysis with personalized study tips

The coaching module (Figure 4) transforms assessment data into actionable plans: weekly study schedules, habit trackers, and motivational guidance tailored to each student's cognitive profile.

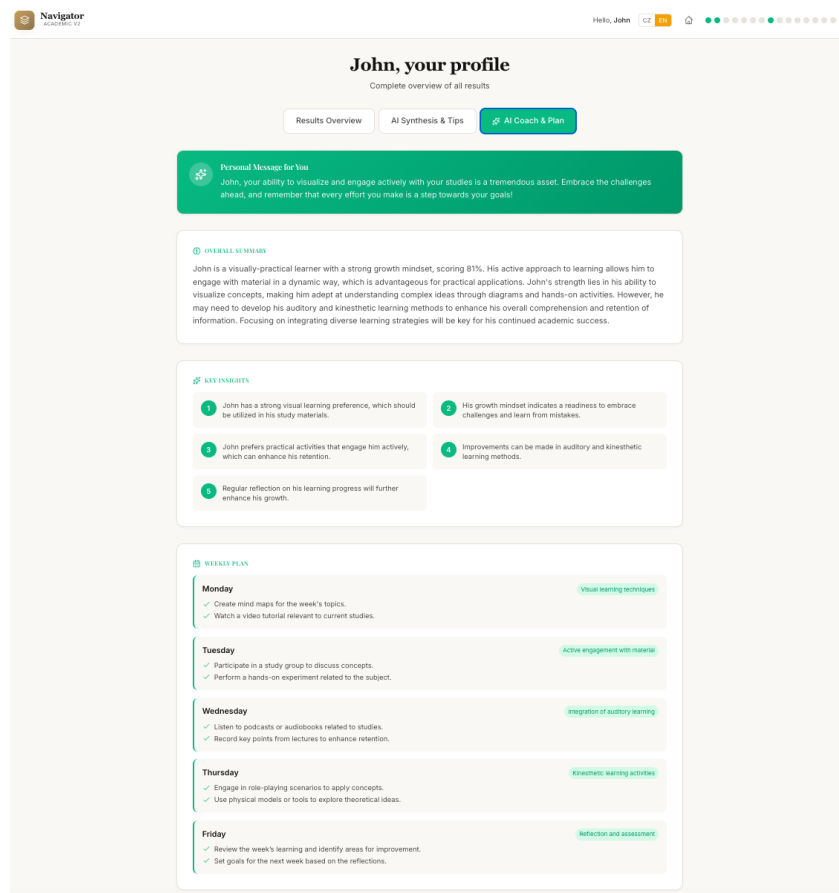


Figure 4: AI Coach - Personalized weekly plan and habit tracking

Implementation Results

The platform was piloted with 120 first-year business students at the Czech University of Life Sciences Prague during Fall 2025. Of 120 enrolled students, 89 (74%) completed the full assessment battery—a completion rate significantly exceeding the 45–50% typically observed with paper-based instruments at our institution. Analytics revealed that 34 students (38% of completers) returned to the platform multiple times, primarily to re-read AI synthesis reports or explore the coaching module. Average session duration was 42 minutes, with students completing an average of 5.2 assessments per session.

Qualitative review of AI-generated syntheses demonstrated GPT-4's ability to integrate cross-instrument patterns meaningfully. For example, one student with strong visual learning preference (VAK), high spatial ability (Mental Rotation), and investigative career interest (RIASEC) received specific recommendations for data visualization coursework and research methodology—connections that would be difficult to derive from individual test reports alone. The AI coaching module generated 89 unique weekly study plans, each reflecting the student's specific cognitive profile and self-reported goals. AI synthesis generation averaged 3.2 seconds (SD=1.1s). The bilingual Czech/English interface served both domestic and international students effectively, with 23% of completers using the English interface.

Lessons Learned

Our implementation experience yielded several insights relevant to educators considering similar AI integrations. First, AI augments rather than replaces: the most effective use of GPT-4 was enhancing validated instruments with personalized interpretation, not

generating original assessments. Students trusted results because underlying instruments had established scientific validity. Second, integration creates value: cross-assessment synthesis provided insights no single test could deliver. Connecting learning style preferences with career interests and cognitive abilities created actionable guidance that standalone reports cannot match. Third, privacy enables trust: client-side data storage addressed GDPR concerns while maintaining full personalization capability. Students expressed comfort knowing their assessment data never left their devices. Fourth, framing matters: students needed explicit guidance that AI recommendations complement, not replace, academic advising. We added disclaimers and encouraged students to discuss AI insights with human advisors. Fifth, immediate feedback drives engagement: the dramatic improvement in completion rates (74% vs. typical 50%) appears directly linked to immediate AI-generated feedback, transforming assessment from a passive data collection exercise into an interactive self-discovery experience.

Conclusion and Future Directions

The AI Academic Navigator demonstrates practical integration of large language models into higher education assessment. By combining validated psychological instruments with GPT-4's analytical and synthesis capabilities, we provide personalized, actionable insights at scale that were previously available only through individual consultations with trained counselors. Our pilot results suggest that AI-enhanced assessment addresses key limitations of traditional approaches: delayed feedback becomes immediate, generic interpretation becomes personalized, and fragmented results become integrated narratives. The 74% completion rate indicates that students perceive genuine value in AI-mediated self-assessment. Future development will focus on longitudinal tracking—connecting initial assessments with academic outcomes to validate AI recommendations—and expanding the instrument battery to include domain-specific assessments for different academic programs. The future of AI in education lies not in replacement but in augmentation: helping students understand themselves better and chart more effective learning paths.

References

- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087–1101.
- Dweck, C. S. (2006). *Mindset: The New Psychology of Success*. Random House.
- Fleming, N. D. (1987). *Teaching and Learning Styles: VARK Strategies*. Neil D. Fleming.
- Goleman, D. (1995). *Emotional Intelligence*. Bantam Books.
- Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology*, 6(1), 35–45.
- OpenAI. (2023). GPT-4 Technical Report. arXiv:2303.08774.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662.

AI as a Teaching Coach: A Design Case Study of Transcript-Based Feedback for Faculty

Dan Levy, Harvard Kennedy School

Abstract

This design case study examines an early-stage prototype that uses generative AI to provide timely, low-stakes, transcript-based feedback to higher-education instructors after a single class session. Motivated by the “faculty feedback gap,” the project treats AI as a teaching coach rather than an evaluator, prioritizing actionable guidance that can be read in three to five minutes and that transfers to future classes. The system analyzes a class transcript (optionally with slides) and focuses on teaching dimensions that transcripts can support with relative reliability: structure and signaling, questioning patterns, and cognitive demand/sense-making. The paper surfaces key design tensions, including usefulness versus reliability, inference versus epistemic humility, and automation versus faculty agency, and describes scope-limiting choices intended to protect trust in high-stakes educational contexts. It concludes with implications for building AI-supported feedback tools that emphasize constrained claims, transparent uncertainty, and instructor-controlled experimentation.

Keywords: Generative AI, Faculty feedback, Teaching coach, Transcript-based analysis, Epistemic humility

1. Introduction: The faculty feedback gap

Research on expertise development consistently emphasizes the importance of deliberate practice supported by timely, specific feedback. Yet in higher education, teaching is a professional activity in which such feedback is remarkably scarce. Most faculty teach a class, return to their offices, and rely on vague impressions of how the session went. Formal student evaluations arrive weeks or months later, often conflating multiple factors and offering little guidance for improvement. Peer observations and teaching coaching can be valuable, but they are time-intensive and rarely occur more than once or twice a year.

This gap is particularly striking given that many faculty care deeply about teaching and actively seek to improve. The problem is not motivation, but the lack of scalable mechanisms for frequent, formative feedback on teaching practice. Recent advances in generative AI raise a natural question: could AI help provide instructors with timely, low-stakes feedback after individual class sessions, supporting deliberate practice in teaching?

This case study explores an early-stage attempt to answer that question. Rather than proposing an AI system to evaluate teaching quality or student learning, we investigate a

more modest and intentionally constrained idea: AI as a teaching coach. The goal is to help faculty learn from a single class session in ways that transfer to future classes—potentially on entirely different topics—by surfacing observable patterns and suggesting small instructional experiments instructors might try next time.

2. Design goals and constraints

From the outset, the project was shaped by a set of explicit design goals and constraints informed by both the science of learning and faculty psychology.

Design goals

First, the feedback needed to be formative and low-stakes. Faculty are understandably likely wary of tools that appear evaluative or judgmental, particularly when powered by opaque algorithms. The system was therefore designed to avoid scores, rankings, or judgments of “teaching quality,” and instead to frame feedback as support for reflection and experimentation.

Second, feedback had to be actionable within real faculty time constraints. Instructors are unlikely to spend long periods analyzing feedback after each class. We set a target of feedback that could be read and processed in three to five minutes, focused on one or two concrete ideas to try in a future session.

Third, the feedback needed to emphasize transferable teaching insights rather than retrospective optimization of a specific class. Many classes are taught only once per term, or even once per year. Advice framed as “how this class could have been better” risks feeling both irrelevant and demoralizing. Instead, feedback was oriented toward habits, patterns, and instructional moves that could be applied in the instructor’s next class, regardless of topic.

Constraints

At the same time, the project faced significant constraints. The primary data source was a class transcript, typically generated from Zoom or lecture capture tools. Transcripts offer partial and imperfect visibility into classroom activity. They capture spoken words, but not tone, energy, body language, or much small-group or individual work. Any system built on transcripts must therefore grapple with substantial epistemic limits.

Privacy and ethics also loomed large. The system was designed as a personal tool run by the instructor, not as an institutional evaluation mechanism. Full transcripts could be analyzed transiently, but not stored, and the feedback needed to avoid making claims about individual students.

Finally, faculty skepticism toward AI-generated advice was treated not as a barrier to overcome, but as a design reality to respect. A single confident but incorrect claim could undermine trust entirely. Reliability and epistemic humility were therefore prioritized over breadth or sophistication.

3. The prototype: what we built

The prototype developed in this project analyzes a class transcript, optionally supplemented with instructional materials such as slides, and generates short-form feedback structured around a small number of teaching dimensions. The system is intentionally narrow in scope.

Inputs

The core input is a transcript of a single class session, typically in plain text or caption formats (e.g., .vtt). When available, instructors may also upload slides used in the session. Slides are treated as optional context that can clarify instructor intent and planned structure.

What the system does

Rather than attempting a comprehensive analysis of teaching, the system focuses on three categories that are both pedagogically meaningful and reasonably inferable from transcripts:

- Structure and signaling: openings, transitions, and closing synthesis that help students follow the arc of a class.
- Questioning patterns: instructor questions intended to elicit student thinking, excluding logistical or rhetorical questions.
- Cognitive demand and sense-making: the extent to which instructor prompts invite explanation, comparison, or application, using cautious, probabilistic language.

When transcripts include clear speaker labels, the system may also comment cautiously on participation patterns. When labels are absent, it is permitted to make only qualitative, hedged inferences about instructor-led versus student-generated discourse, and explicitly avoids quantitative claims.

What the system does not do

Equally important is what the system explicitly avoids. It does not score teaching, evaluate effectiveness, infer student engagement or learning, or compare instructors to norms or benchmarks. It does not assume that silence indicates disengagement, nor that talk equates to learning. These exclusions are not technical limitations but deliberate design choices intended to protect trust and credibility.

Output format

Feedback is delivered in a consistent, constrained format:

- A one-sentence framing statement clarifying scope and limits
- Two to three bullets on “what seems to be working”
- One or two transferable teaching experiments the instructor might try in a future class
- An optional closing question inviting the instructor to choose a focus for future feedback

This structure is designed to feel more like a brief coaching note than an evaluation report.

4. Key design tensions and decisions

The primary contribution of this case study lies in the design tensions encountered and the decisions made to resolve them.

Usefulness versus reliability

One recurring tension was between providing advice that felt useful and ensuring that claims were well supported by the available evidence. Generative AI systems are capable of producing confident, polished feedback even when the underlying data are thin. In early experimentation, this posed a clear risk.

The design response was to sharply limit scope and to prefer explicitly acknowledging insufficient evidence over offering speculative advice. In practice, this meant focusing on a small number of observable patterns and avoiding fine-grained claims. Reliability was treated as a prerequisite for usefulness, not a tradeoff against it.

Class-specific improvement versus transferable insight

Another central tension concerned the unit of improvement. Should feedback aim to help instructors teach this class better next time, or help them learn something about their teaching that transfers across classes?

The project resolved this tension in favor of transferable insight. Feedback is framed as identifying patterns and habits revealed by a class session, and suggesting small experiments instructors might try in future classes—even on different topics. This orientation aligns with deliberate practice and reduces the risk of defensiveness by shifting attention away from retrospective critique.

Inference versus epistemic humility

Transcripts invite inference, but also demand humility. The system needed to strike a balance between making reasonable qualitative inferences (for example, identifying extended explanatory passages likely delivered by the instructor) and avoiding overreach. Explicit rules were introduced to govern when inference was allowed and how it had to be framed, with uncertainty made visible rather than hidden.

Automation versus faculty agency

Finally, there was a tension between automation and instructor control. Rather than asking instructors to specify goals or preferences before receiving feedback, the system provides default feedback first and then invites customization for future sessions. This preserves low friction while respecting faculty agency.

Box 1 - Design Principles that Emerged from the Case Study

- Treat AI as a teaching coach, not an evaluator.
- Prioritize epistemic humility over coverage or sophistication.
- Focus feedback on transferable teaching insights, not retrospective critique.
- Limit analysis to what available data can reliably support.
- Preserve faculty agency by framing advice as experiments, not prescriptions.

5. Early experimentation and reflections

Early experimentation involved running the prototype on transcripts from the author's own classes and from a small number of trusted colleagues. While systematic evaluation is

beyond the scope of this case study, the iterative design process surfaced several recurring considerations.

In particular, the design increasingly emphasized feedback on structure, questioning, and sense-making, framed as transferable and experimental, as these dimensions were judged most likely to yield actionable insight given the limits of transcript-based evidence. By contrast, the design deliberately avoids feedback that could be read as implicitly evaluative or as over-interpreting transcript data, given the risk that such feedback could undermine trust.

These reflections informed successive refinements of both the scope of analysis and the tone of feedback, reinforcing the importance of epistemic humility in AI-supported teaching tools.

6. Implications and next steps

This design case study suggests that generative AI can plausibly support faculty deliberate practice—but only under carefully constrained conditions. The main challenge is not model capability, but design discipline: deciding what the system should and should not attempt to do, how uncertainty is communicated, and how feedback is framed relative to faculty goals and identities.

Future work will explore how such tools might integrate into teaching centers or professional development programs without becoming evaluative, and how longitudinal use might support pattern recognition over time. More broadly, the case highlights the importance of scope-limiting and epistemic humility when deploying AI in higher education contexts where trust is fragile and stakes are high.

Acknowledgements

I would like to thank Mae Klinger, Rony Rodriguez and Victoria Barnum for intense collaboration in building various versions of the prototype and for the many exchanges that informed its design. I would also like to acknowledge the use of ChatGPT (OpenAI) for assistance with brainstorming, drafting, and iterative refinement of this case study. Responsibility for the content, interpretations, and any remaining errors rests entirely with me.

Beyond time savings: The hidden pedagogical benefits of AI-assisted grading

Alexandre Billard (CEO)
& Louis Delbosc (CTO),
Bluetrain

Abstract

AI-assisted grading is often framed as a solution to reduce faculty workload by automating a time-consuming and repetitive task. While efficiency gains are real, focusing solely on productivity misses AI's broader potential to enhance the quality of assessment and pedagogy. This submission aims to contribute to the conference's live demonstrations by showing how Bluetrain's AI-assisted grading goes beyond efficiency gains to elevate assessment quality, academic rigor, and pedagogical value in Higher Education Institutions (HEIs).

About Bluetrain

Bluetrain is a French AI-assisted grading solution for business schools and universities.

Already trusted by leading French business schools such as EDHEC Business School, EM Lyon Business School, and Audencia, Bluetrain uses AI fine-tuned for business disciplines (e.g., Law, Marketing, Finance, Strategy). It enables:

- ~50% reduction in grading workload.
- More consistent and fair assessments.
- Personalized student feedback that strengthens compliance with international accreditation standards (e.g., AACSB, EQUIS).

As AI adoption is no longer optional in business education, Bluetrain provides a scalable solution that aligns with schools' AI roadmaps and integrates easily with major LMS (e.g., Blackboard, Moodle, Brightspace), while enabling faculty to focus on high-impact student interactions.

Keywords: AI, Assessment, Engagement, Feedback, LMS

Product demonstration

A demonstration based on a Corporate Finance exam (on 650 students) from EM Lyon Business School will illustrate how AI-assisted grading, if properly calibrated, strengthens student engagement through meaningful personalized feedback and faster results for students.

1) AI-Assisted grading for stronger assessment and pedagogy

(a) More objective and consistent grading

The demonstration illustrates how AI calibration is conducted prior to the exams to improve grading quality and align AI evaluations with faculty grading standards. It also shows how this calibration enables faster pre-grading rubric alignment and stronger consistency across professors before the evaluation process begins.

The demonstration will show how proper calibration enabled the achievement of 98% accuracy, measured as the proportion of grading criteria scored identically by AI and human evaluators.

Achieving high accuracy in AI-assisted grading is a prerequisite for unlocking two key pedagogical benefits: (i) providing a reliable basis for meaningful feedback generation, and (ii) providing students with faster results through an accelerated grading process.

(b) Personalized feedback at scale to accelerate student learning

The goal is to show how AI-generated feedback can accelerate student learning trajectories.

The demonstration illustrates how, when properly configured, AI generates efficient, personalized feedback while serving as a “teaching assistant” to support student preparation for the final exam. It further incorporates faculty perspectives to examine a key pedagogical trade-off between concise and more elaborated feedback.

The demonstration will also illustrate how the proposed AI-augmented assessment process improved students’ learning trajectories, resulting in a significant reduction in failure rates in subsequent exam sessions.

(c) Unlocking more AI-resilient assessments

In the era of AI, students have widely adopted generative AI tools for home assignments. As a result, many traditional take-home exams, particularly knowledge-based assessments designed prior to AI have become significantly less effective as evaluation tools. This shift creates a strong opportunity for Higher Education Institutions (HEIs) to move from knowledge-based exams toward competency-based assessments that better align with labor market needs.

However, redesigning exams to be AI-resilient and competency-based introduces a new challenge: such assessments are typically more complex, open-ended, and substantially more time-consuming to grade. AI-assisted grading addresses this challenge by (i) reducing grading workload and (ii) enabling the analysis of more extensive and complex student submissions that were previously impractical to assess at scale.

First, the demonstration focuses on the AI-resilient, competency-based Corporate Finance exam (Excel assignment) selected for this study, illustrating how AI-assisted grading enables rigorous and meaningful evaluation at scale for a large cohort of students.

Second, in the context of the Corporate Finance exam, the demonstration will show how AI agents automatically identify key cells, highlight critical elements, and rebalance balance sheets to clearly identify student errors, tasks that are typically very time-consuming when done manually.

2) Practical risks of AI in grading and how to stay in control

A frequently observed pitfall is insufficient AI calibration, which can unintentionally lead to overly lenient AI grading. When grading criteria are unclear, AI systems often adopt overly lenient grading behavior. This creates a serious risk of reduced academic rigor and inflated student grades.

This demonstration illustrates how this issue can be effectively mitigated by comparing a poorly calibrated AI-assisted grading setup with an improved calibration, showing significantly higher grading accuracy and more appropriate feedback for students.

3) Q&A after demonstration

As AI-assisted grading gains traction, it naturally raises many questions and legitimate concerns around academic rigor, transparency, and alignment with institutional values.

This Q&A session will:

- Address questions and concerns.
- Share concrete feedback from professors, students, and institutions on Bluetrain's AI-assisted grading.

Educational Technology: The Dream Triplet for Implementing AI Innovation

Emmanuelle DEGLAIRE, Associate professor in Law & Tax,
EDHEC Business School,
Member of EDHEC Augmented Law Institute,
EDHEC Business School

Emmanuelle HOUET, Director of the PiLab – Pedagogical
Innovation Laboratory,
EDHEC Business School

Alexandre BILLIARD, Co-founder of BlueTrain, EdTech
specialized in AI-assisted grading

Abstract

Since February 2023 the PiLAB, the Pedagogical Innovation Lab of EDHEC Business School; one of the school's professors; and a newly founded EdTech startup, Bleu Train, started a genuine project whose ambitious goal is to **enhance grading through artificial intelligence**. This research analyzes that journey in depth, outlining the different key moments, the challenges, and both small and major successes. Three different methodologies are used to examine that venture, namely, the Six Education Technology Integration Principles by Dexter (2024), the Teacher Technology Adoption Model (Lei & Morrow 2010) and the Four-Dimensional CTL Model (Mihai 2025). Ultimately this research concludes that complementarity among stakeholders is essential for the success of any pedagogical and entrepreneurial innovation initiative, the Center for Teaching and Learning (CTL) playing the central role.

Keywords

Academic Innovation - Educational Technology - Faculty Development - EdTech - AI

Introduction

Academic innovation has become an increasingly central theme in the strategic agendas of Higher Education Institutions (HEIs), and abundant research seeks to identify the key success factors in this transformation. The main challenge lies in moving beyond innovation as a mere *buzzword* used for communication purposes (Barger et al., 2022) toward the implementation of true operational transformation. This research aims precisely to revisit the theoretical contributions in this field, through the retrospective analysis of a concrete case of pedagogical innovation success - confirming, refuting, or expanding on existing research findings.

In February 2023, only a few months after the public release of OpenAI's first large language model (LLM), a “**dream triplet**” emerged: the PiLAB, the Pedagogical Innovation Lab of EDHEC Business School; one of the school's professors; and a newly founded EdTech startup, Blue Train, with the ambitious goal of **enhancing grading through artificial intelligence**.

This research analyses that journey (Daly & Deglaire 2025) step by step, outlining the challenges and both small and major achievements. Using a deductive method, the research refers to three different models, corresponding to each of the 3 actors' roles, namely the Six Education Technology Integration Principles by Dexter (2024), the Teacher Technology Adoption Model (Lei & Morrow 2010) and the Four-Dimensional CTL Model (Mihai 2025). Ultimately, this live experimentation's analysis proves that complementarity among stakeholders is essential for the success of any pedagogical and entrepreneurial innovation initiative and provide hands on takeaways for practical education technology implementation.

Literature review

When it comes to pedagogical innovation, the **professor** plays a pivotal role (Lei & Morrow, 2010). Innovation is not necessarily digital, though educational literature often overlap the two (Barger, 2021). Still, technology frequently plays a crucial part in enabling innovation, which calls for new skill development. Mishra & Koehler (2006) emphasized that

mastering *Pedagogical Content Knowledge (PCK)* is no longer sufficient for teachers: they must also develop *Technological Pedagogical Knowledge (TPK)* to ultimately achieve *Technological Pedagogical Content Knowledge (TPACK)* (Dexter, 2023). Similarly, the SAMR model (Substitution, Augmentation, Modification, Redefinition) helps teachers focus on why they integrate technology - to enhance, transform, or redefine their teaching tasks. Falloon (2020) proposed an expanded conceptual framework of teacher digital competence (TDC). Yet, as central as the teacher is, innovation cannot rely on educators alone; a more holistic approach is needed (Falloon, 2020). Barger (2021) highlights the threefold dimension of academic innovation: faculty development, teaching and learning programming, and entrepreneurial ventures.

Centers for Teaching and Learning (CTLs) now occupy a growing place in literature (Hines, 2023; Wright, 2023; Mihai et al., 2025). Although terminology varies - *Learning Innovation Centers, Centers for Academic Innovation ...* - their mission remains to support “faculty to meet the multiple and rigorous expectations of the institution” through diverse services and initiatives (Dickens et al., 2019).

Another key actor is the **EdTech** sector, facilitating the integration of digital tools, software, and platforms to improve learning experiences and outcomes (Tzenios, 2020; Pandey & Pandey, 2022; Dexter, 2023). According to Adeoyer & Otemuyiwa (2024), EdTechs' offers can be grouped into several categories: personalized learning, professional development for teachers, expanding access and equity, fostering collaboration and engagement, and leveraging data-driven insights. In a highly competitive education market, the appetite for technological innovation is strong and even amplified by AI development. Yet, amid a flourishing entrepreneurial ecosystem, measuring EdTechs' performance has become a growing imperative (Barger, 2021).

Methodology

The methodology adopted in this research is the deductive method. Each element of the storytelling of the lived experience was compared with the theoretical frameworks

proposed by the research, with a threefold objective. First, to confirm the recommendations of academic research through this example. Conversely, some elements identified in the literature were not found. Finally, certain elements were observed to be decisive in the practice studied, even though, not referenced in the models.

Three main models were used:

- Dexter's (2024) Six Education Technology Integration Principles (ETIPs)
 1. Learning outcomes drive the selection of EdTech,
 2. EdTech provides added value to teaching and learning
 3. EdTech assists in the assessment of the learning outcomes.
 4. EdTech integration is feasible and technical support is available in a workable time frame if a problem arises.
 5. The EdTech integration should give the opportunity for faculty professional learning.
 6. Professional community enhances EdTech integration.

- Lei & Morrow's (2010) Teacher Technology Adoption Model
 1. Involving teachers in the decision-making process so that the technology integration becomes meaningful;
 2. Helping teachers to develop a well-designed plan with realistic goals and a feasible implementation outline;
 3. Building a community where teachers can learn from peers and work collaboratively;
 4. Implement a strong leadership to ensure resources and support for a successful and ethical tech implementation; and
 5. Proving timely support to help teachers remove roadblocks

- Mihai's (2025) Four-Dimensional CTL Model

1. Activities: Educational development, research, assessment

2. Governance : Positionality, alignment with institutional vision, Size and structure, Staff

3. Identity: Self-positioning, mission, roles

4. Social capital: internal and external networks

The observational field

The initial encounter

In February 2023, the PiLAB organized a one-hour online workshop to raise faculty awareness about AI's role in teaching. Its success lay in its timeliness, accessibility, and balance between technical discussion and non-alarmist perspective. This seminar was made possible thanks to the curiosity of the PiLab director and the structured technology watch activity carried out by her entire team.

Around the same time, an alumnus reconnected with his former program director, now in charge of the PiLab, to brainstorm AI's potential for faculty support. One clear *pain point* emerged: grading, a time-consuming and repetitive task, ideal for industrializing processes testing. The PiLAB director subsequently invited the faculty members participating in her workshop to serve as pilot participants in a series of experimental trials. One volunteered.

A genuine project was born, neither fully intentional nor formally structured, without a name or budget, sparked simply by a chance encounter and the spontaneous enthusiasm of three individuals deeply curious about innovation.

First experiments

Initial testing began in March 2023. It was decided to conduct preliminary tests using papers from previous examinations in order to assess the system's performance on work that had already been graded, with the professor's score serving as a reference point for evaluating the quality of the AI's analysis. The process began with a single paper, then three, and five, through successive iterations and refinements. Technologies were compared, new tools tested, and the notion of fully autonomous grading was eventually abandoned in favor of an AI-assisted pre-correction process. Following the successful completion of this experimental phase, the subsequent stage focused on deploying the solution institution-wide.

Deployment

By March 2024, the system was used for the first time on 50 students exam papers that had not been previously graded by a professor. It was first necessary to identify a course whose examination format was compatible with the solution. Students had to be prepared for a computer-based examination, familiarized with Respondus, another EdTech tool the exam had to be configured within Blackboard. The pilot test proved successful and subsequently opened the way for discussions with other program directors, including those managing larger cohorts, leading to a large-scale implementation involving 650 students in January 2025. Nevertheless, the rollout remains gradual, as several barriers persist, including the strong reluctance of some faculty members, the cautious stance of certain program directors, and even resistance from a portion of the student body.

For the 2025–2026 academic year, approximately twenty professors and courses at EDHEC Business School are now using Blue Train's AI-assisted grading solution, primarily for courses with large student enrollments.

Results

Based on Dexter's six principles, five were confirmed by the observed practice, while one principle was only partially observed: '*EdTech integration should give the opportunity for faculty professional learning*'. Indeed, no specific prior knowledge of AI was required from the teacher to conduct this experiment..

Based on the Lei & Morrow model, four out of five principles were confirmed. However, the second principle (*'develop a well-designed plan with realistic goals and a feasible implementation outline'*) was not observed at all. The spontaneous nature of the project, the absence of a budget, and therefore the absence of formal accountability allowed the project to evolve entirely through an ad hoc, exploratory approach. What was observed instead was the effectuation method, specific to entrepreneurship: *'Effectuation is the inverse of causation. Causal models begin with an effect to be created. They seek either to select between means to achieve those effects or to create new means to achieve pre-selected ends. Effectual models, in contrast, begin with given means and seek to create new ends using non-predictive strategies'* (Sarasvathy, 2008). This approach is known to be powerful in entrepreneurial education (Mäkimurto-Koivumaa & Puhakka, 2013), and small-scale projects can also enhance teaching and learning (Dexter & Seden, 2012).

Based on Mihai's model (2025), 100% of the expected qualities of a CTL were observed within the school's PiLab. This is most certainly the main reason for the success of this initiative.

Contribution

Three main lessons emerged from this research:

- **The central role of CTLs.** – Although the project was initiated by three voluntary actors, one of them emerged as having a decisive role: the PiLab, qualifying according to the literature vocabulary as a Center for Teaching and Learning (CTL). Far more than a simple facilitator, the CTL serves as the pivotal link connecting the various stakeholders. Depending on the institution, it may or may not hold formal decision-making authority; however, its credibility and legitimacy invariably help to trigger the deployment process, which might otherwise become trapped in an endless back-and-forth between the faculty and program management.
- **The importance of an enabling ecosystem.** – A school lacking a culture of innovation, governed by overly rigid management that allows no room for error, influenced by overly conservative parents, and operating within a country

constrained by restrictive regulations, faces multiple barriers that can quickly put an end to even the most promising individual initiatives.

- **The need for EdTech interoperability.** – A final but crucial aspect of operational implementation and large-scale deployment lies in the importance of interaction among EdTech solutions themselves. Established players in the market tend to be less agile and slower to adapt, though they undeniably dominate the sector. In contrast, newer and smaller EdTechs demonstrate remarkable responsiveness and flexibility. Pedagogical innovation can only advance within an ecosystem that fosters interoperability between systems, as institutions are often highly reluctant to adopt any solution that is not compatible with their existing technological infrastructure.

References

Adeoye, M. A., & Otemuyiwa, B. I. (2024). Navigating the Future: Strategies of EdTech Companies in Driving Educational Transformation. *JERIT: Journal of Educational Research and Innovation Technology*, 1(1), 43-50.

Barger, A. P., Leffel, K. G., & Lott, M. (2022). Plotting academic innovation: a content analysis of twenty institutional websites. *Innovative Higher Education*, 47(1), 95-111.

Daly, P., & Deglaire, E. (2025). AI-enabled correction: A professor's journey. *Innovations in education and teaching international*, 62(4), 1241-1257.

Dexter, S. (2023). Developing faculty EdTech instructional decision-making competence with principles for the integration of EdTech. *Educational technology research and development*, 71(1), 163-179

Dexter, B., & Seden, R. (2012). 'It's really making a difference': how small-scale research projects can enhance teaching and learning. *Innovations in Education and Teaching International*, 49(1), 83-93.

Dickens, E., Cruz, L., Alderson, J., Atias, D., Graham, R., Hurney, C. A., ... & Wang, C. (2019). Starting a center for teaching and learning. *POD Speaks*, 3, 1-6.

Falloon, G. (2020). From digital literacy to digital competence: the teacher digital competency (TDC) framework. *Educational technology research and development*, 68(5), 2449-2472

Friedman, R. S., & Deek, F. P. (2004). Innovation and education in the digital age: reconciling the roles of pedagogy, technology, and the business of learning. *IEEE Transactions on engineering management*, 50(4), 403-412.

Hines, S. R. (2017). Evaluating centers for teaching and learning: A field-tested model. *To improve the academy*, 36(2), 89-100.

Lei, J., & Morrow, B. (2010). Teachers' adoption of technology innovation into pedagogical practices. *Education and Information Technologies*, 15(3), 143-153.

Mäkimurto-Koivumaa, S., & Puhakka, V. (2013). Effectuation and causation in entrepreneurship education. *International Journal of Entrepreneurial Venturing*, 5(1), 68-83.

Mihai, A. I., Dailey-Hebert, A. L., & Beusaert, S. (2025). Understanding the complexity of centers for teaching and learning: introducing a four-dimensional model. *Higher Education*, 1-20

Mishra, P., & Koehler, M. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 6, 1017-1054

Pandey, J., & Pandey, A. (2022). A Systematic Review of Ed Tech Education Platform: An Indian Perspective. *Global Journal of Research in Management*, 12(1), 82

Puentedura, R. (2006, August). *Transformation, technology, and education*.

Sarasvathy, S. D. (2008). Effectuation: Elements of entrepreneurial expertise. In *Effectuation*. Edward Elgar Publishing.

Serdyukov, P. (2017). Innovation in education: What works, what doesn't, and what to do about it? *Journal of Research in Innovative Teaching & Learning*, 10(1), 4-33. <https://doi.org/10.1108/jrit-10-2016-0007>

Tzenios, N. (2020). Examining the impact of EdTech integration on academic performance using random forest regression. *ResearchBerg Review of Science and Technology*, 3(1), 94-106.

Wright, M. C. (2023). *Centers for teaching and learning: The new landscape in higher education*. JHU Press

Understanding Linguistic Feature Contributions in AES Systems

Maryam Abo-Tabik, University of Lancashire
Giulio Corsi, University of Cambridge
Alexandru Marcoci, University of Cambridge
Georgiana Thorpe Apreutesei, University of Cambridge
Yael Benn, Manchester Metropolitan University
Lyba Razzaq, Manchester Metropolitan University
Isla Fay, University of Cambridge
Deborah Talmi, University of Cambridge
Roni Tibon, University of Nottingham

Abstract

Automated Essay Scoring (AES) systems are increasingly used to assess student writing, with modern approaches relying on large language models (LLMs). While these models can mark essays quickly and consistently, it remains unclear how linguistic features such as lexical diversity measures influence their assessments compared with human markers when the task is not designed to evaluate students' writing skills. This study analyzed 266 undergraduate essays from three different universities graded by both humans and three LLMs (GPT, Gemini, and Claude). Various lexical diversity metrics were examined using PCA analyses, with and without controlling for essay words count. Results showed that AI scores, unlike human scores, were strongly associated with lexical diversity metrics (TTR, Guiraud Index, MTL) and word count, with longer essays receiving higher marks. Even after controlling for word count, AI evaluation remained sensitive to vocabulary distribution, whereas human scores showed no comparable dependence, indicating AI's greater reliance on textual properties and potential length-related biases.

Keywords: Automated Essay Scoring; Large Language Models; Linguistic Features; Human vs AI Assessment; Essay Evaluation.

Introduction

There is an increased interest in understanding the accuracy and reliability of Automated Essay Scoring (AES) marking of student essays. Early AES systems were based on simple pattern matching and statistical approaches that used features such as word counts, sentence length, and spelling or grammar errors. These approaches demonstrated that computers could approximate human grading, but they often failed to capture the full spectrum of textual features associated with writing quality (Ikram & Castle, 2020; Ramesh & Sanampudi, 2022).

Several approaches have been explored (Lim et al., 2021; Lagakis & Demetriadis, 2021) in AES, but a major improvement came with the use of transformer-based models such as BERT (Bidirectional Encoder Representations from Transformers (Beseiso & Alzahrani, 2020; Azhari et al., 2024). These models capture richer linguistic and semantic patterns in text, allowing AES systems to evaluate writing quality better than earlier feature-based or earlier deep learning methods.

The release of OpenAI generative models (OpenAI, 2026) sparked further interest in using large language models (LLMs) in AES. Several research works suggest that models such as GPT can score essays efficiently (i.e. quicker than human markers, more consistent, and less workload to educators) (Mizumoto & Eguchi, 2023; Quah et al., 2024; Yamashita, 2024; Liew & Tan, 2024). Unlike BERT, these pre-trained LLMs can handle a wider range of language tasks without task-specific training (Matarazzo & Torlone, 2025):

While LLMs have been shown to be more accurate in marking students' work compared to other AES approaches, it is unclear what factors beyond subject-related information may influence Artificial intelligent (AI) marking. For example, how LLMs weight lexical diversity when assessing student work that is not specifically a writing-skill task. Previous work evaluating AES has mostly focused on writing skills and language-based tasks (Ramesh & Sanampudi, 2022); for example, assessing writing quality through coherence (Frag, Yannakoudakis & Briscoe, 2018), grammatical accuracy (Jaganov, Blake, Villegas & Carr, 2025), and vocabulary use, besides evaluating agreement with human markers. Many of these studies found that factors like grammatical errors, lexical diversity, sentence complexity, and other linguistic measures drive AES decisions (Arnold et al., 2018; Atkinson & Palma, 2025; Boulanger & Kumar, 2020; Liu et al., 2023). Nevertheless, the influence of lexical diversity on AES when scoring non-language subjects remains unexplored.

However, the present study differs from prior AES research in the nature of the assessment task. Rather than evaluating essays written for language proficiency or writing-skill assessment, we analyse responses to substantive undergraduate psychology examination questions administered across three universities in the United Kingdom. These essays are designed to assess disciplinary knowledge, conceptual understanding, critical evaluation of empirical evidence, and the ability to construct theoretically informed arguments. In this context, writing quality is not the primary construct being assessed; instead, markers are expected to use the rubric to evaluate features such as the precision of conceptual explanations, the depth of engagement with relevant literature, and the coherence and organisation of the argument. This distinction allows us to examine whether lexical diversity metrics influence scoring even when the assessment task is primarily analytical and content-driven rather than explicitly linguistic.

Therefore, the present study investigates whether a range of lexical diversity metrics influence essay scores awarded by human and different LLM markers and systematically compares the strength of these effects across the two groups, assessing AES performance on a series of non-language-based tasks.

Methods.

In this study, 266 psychology summative assessments from different modules across 3 UK universities: the University of Cambridge (N=108), Manchester Metropolitan University (N=130), and the University of Nottingham (N=26) were collected from undergraduate students. All assessments were graded by expert human markers as part of the routine marking process using standard rubrics. For the purpose of this study, each essay was processed through three different LLM-based grading systems (GPT, Gemini, Claude); all three models were used in a zero-shot setting, with temperature=0 to make their outputs deterministic. Using an ensemble of models (sp_score) within the Surprisingly Popular framework (Prelec, Seung & McCoy, 2017) allowed us to evaluate cross-system consistency and examine how different LLMs weighted linguistic features when generating meta-cognitive predictions about scoring outcomes (i.e each model does not only provide a classification but also estimates how other models would classify the same item).

Multiple linguistic metrics were extracted from each submitted assessment. In the present analysis, we focus on lexical diversity measures (see Table 1), which have been widely used in related work (Arnold et al., 2018; Atkinson & Palma, 2025; Boulanger & Kumar, 2020; Liu et al., 2023). Lexical diversity refers to the range and variability of vocabulary used within a text. It captures the proportion of distinct word types (unique words) relative to the total number of words (tokens), reflecting the richness of the lexicon.

Table 1. Linguistic metrics to capture lexical diversity

Metric	Explanation
Word Count	Number of words in the text
TTR	Ratio of unique words to total words; sensitive to text length and often decreases as text grows longer.
Guiraud_Index	Adjusts TTR by dividing types by square root of tokens, reducing text-length sensitivity.
Herdan's C	Logarithmic lexical diversity measure balancing vocabulary growth against text length.
Maas Index a^2	Lexical diversity metric where lower values indicate richer vocabulary, less sensitive to length.
Dugast U	Vocabulary richness index emphasizing deviation from expected lexical growth.
MTLD	Measure of Textual Lexical Diversity, measures mean segment length maintaining stable TTR; robust against text-length effects.

Results.

We initially computed Bonferroni-corrected pairwise Pearson correlations (adjusted across 36 comparisons) among all variables included in the study: two scoring variables (human and AI marking) and seven lexical diversity metrics. The results are presented in Figure 1, Panel A.

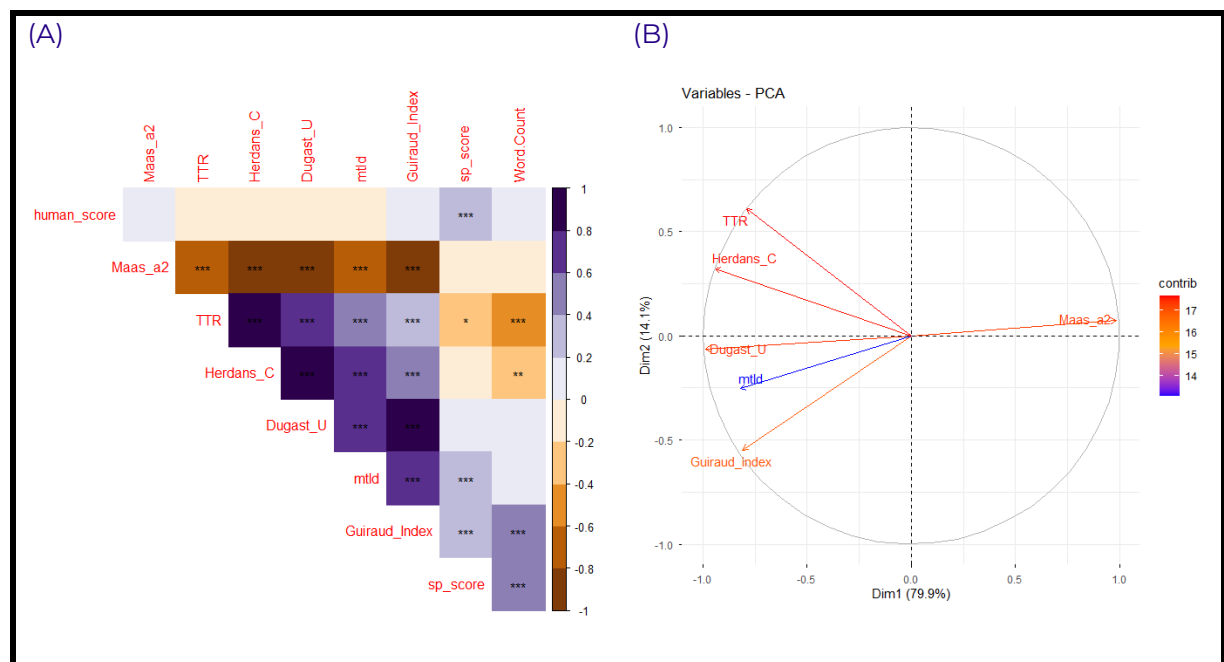


Figure 1. Analyses of lexical diversity of assessments. Panel A: Pearson correlation matrix between all variables included in the study. Asterisks indicate significant results, Bonferroni-corrected across 36 comparisons; * $p < .05$, ** $p < .01$, *** $p < .001$. Panel B: Correlation circle plot showing loadings of six lexical diversity indices on the first two principal components. Arrows represent variable loadings; colour intensity reflects each variable's contribution to the components

Although this analysis suggested that AI marking may be associated with several measures of lexical diversity, it also revealed a high degree of multicollinearity among the linguistic metrics. Moreover, some lexical diversity metrics are known to be strongly influenced by overall text length (Zheng, 2025), and word count emerged as the variable most strongly associated with AI marking ($r = .42, p < .001$). Therefore, to reduce dimensionality, improve interpretability, and isolate effects beyond simple verbosity, we combined principal component analysis (PCA) with general linear modelling (GLM).

Specifically, we conducted a PCA on six of the lexical diversity metrics (TTR, Guiraud Index, Herdan's C, Maas's a^2 , Dugast's U, and MTLD), excluding Word Count. Variables were standardised prior to analysis (z-scored) and PCA was implemented using singular value decomposition as implemented in the `prcomp` function in R.

The first two principal components (PCs) jointly accounted for ~94% of the variance (see Figure 1, Panel B). PC1 accounted for 79.9% of the variance, indicating substantial shared variance across indices. Most measures (TTR, Herdan's C, Dugast's U, Guiraud Index, and MTLD) loaded strongly in the same direction on PC1, whereas Maas's a^2 loaded in the opposite direction, consistent with its inverse scaling relative to other diversity metrics. PC2 accounted for 14.1%, with TTR and Herdan's C loading positively on PC2, whereas Guiraud Index and MTLD loaded negatively.

Next, two GLMs were conducted to examine the degree to which human and AI marking are associated with the PCs that were obtained, whilst also controlling for the (already established) effect of Word count, expressed as:

- GLM1: Human Mark ~ PC1 + PC2 + Word.Count
- GLM2: AI Mark ~ PC1 + PC2 + Word.Count

The results of these analyses are shown in Table 2. For human marking, the GLM was not significant, $F(3, 262) = 2.19, p = .089$, and only explained a small proportion (0.25%) of variance. Neither PC nor word count significantly predicted human scores (all $ps > .2$). Contrary, for AI marking, the GLM was significant, $F(3, 262) = 25.7, p < .001$, explaining 22.7% of the variance. In addition, PC2 significantly predicted AI scores ($p < .001$), whereas PC1 and word count were not significant predictors ($ps > .1$)

	GLM1: Human Mark	GLM2: AI Mark
	Total Model	
F-statistic	2.2 (.09)	25.7 (<.001)
R²	0.025	0.227
Adjusted R²	0.013	0.229
	Model's Coefficients	
Intercept	12.02 (<.0001)	30.25 (<.0001)
PC1	0.76 (.45)	-1.58 (.12)
PC2	0.21 (.84)	-3.91 (<.001)
Word.Count	1.28 (.20)	-0.13 (.89)

Discussion.

We examined how lexical diversity relates to marks awarded to psychology undergraduate assessments by both human and AI markers. Overall our results indicated that LLM-based grading remains sensitive to surface-level lexical properties even in content-driven assessments. This effect is not explained solely by essay length; rather, vocabulary distribution patterns (PC2) appear to serve as implicit quality heuristics. Although lexical sophistication often correlates with academic performance, reliance on such features may introduce construct-irrelevant variance if weighted independently of conceptual accuracy or analytical depth.

Our First key finding was that several diversity metrics, including word count, TTR, Guiraud Index, and MTLN, were associated with AI marking but not with human marking, suggesting that AI evaluation may be more generally sensitive to text diversity. Among these, word count showed the strongest association, with longer assessments receiving higher AI scores.

At the individual level (i.e., students who write more receiving higher marks) this finding is difficult to interpret, as we did not control for other factors such as the inclusion of more relevant content in longer essays. However, the structural implications are more consequential. The results suggest that assessments permitting longer responses may systematically yield higher AI scores. In other words, rather than assuming comparable score distributions across assessment formats, AI marking may produce different distributions depending on permitted length. If AI systems are to be used in evaluation, institutions should account for this effect when designing assessments and implement appropriate controls to mitigate unintended length-related biases.

Our second finding is that even when controlling for word count, some effects of lexical diversity on AI marking (though again, not on human marking) remain. The PCA revealed a dominant general lexical diversity dimension (PC1), alongside a secondary component (PC2) capturing differences in how lexical diversity is distributed across the text—specifically, contrasting TTR and Herdan's C with more length-robust measures such as Guiraud Index and MTLN (Figure 1, Panel B). While neither component predicted human scores, AI marking was significantly associated with PC2, independent of word count. The negative association indicates that AI systems assign lower scores to texts whose diversity reflects surface type introduction (i.e., early bursts of new vocabulary typical of TTR-like behaviour), and relatively higher scores to texts exhibiting sustained, length-robust lexical dispersion across the response. Thus, beyond simple verbosity, AI evaluation appears to be sensitive to structural patterns in vocabulary distribution, whereas human marking shows no comparable dependence on these latent lexical dimensions.

Taken together, these findings suggest that lexical diversity influences AI-based marking, even in assessment tasks not explicitly designed to evaluate writing ability. Understanding these influences is essential for developing best practices in the use of AI for assessment, ensuring that such tools are applied responsibly, transparently, and in ways that promote fair and accurate evaluation.

References

- Arnold, T., Ballier, N., Gaillat, T. and Lissón, P., 2018. Predicting CEFRL levels in learner English on the basis of metrics and full texts. arXiv preprint arXiv:1806.11099.
- Atkinson, J. and Palma, D., 2025. An LLM-based hybrid approach for enhanced automated essay scoring. *Scientific Reports*, 15(1), p.14551.
- Azhari, A., Santoso, A., Ratna, A.A.P. and Prestiliano, J., 2024. Optimization of AES using BERT and BiLSTM for Grading the Online Exams. *International Journal of Intelligent Engineering & Systems*, 17(5).
- Beseiso, M. and Alzahrani, S., 2020. An empirical analysis of BERT embedding for automated essay scoring. *International Journal of Advanced Computer Science and Applications*, 11(10).
- Boulanger, D. and Kumar, V., 2020, June. Shaped automated essay scoring: Explaining writing features' contributions to english writing organization. In *International Conference on Intelligent Tutoring Systems* (pp. 68-78). Cham: Springer International Publishing.
- Farag, Y., Yannakoudakis, H. and Briscoe, T., 2018, June. Neural automated essay scoring and coherence modeling for adversarially crafted input. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 263-271).
- Ikram, A. and Castle, B., 2020, October. Automated Essay Scoring (AES); A Semantic Analysis Inspired Machine Learning Approach: An automated essay scoring system using semantic analysis and machine learning is presented in this research. In *Proceedings of the 12th International Conference on Education Technology and Computers* (pp. 147-151).
- Jaganov, T., Blake, J., Villegas, J. and Carr, N., 2025. Large language model-driven dynamic assessment of grammatical accuracy in English language learner writing. *IEEE Access*.
- Lagakis, P. and Demetriadis, S., 2021, November. Automated essay scoring: A review of the field. In *2021 International Conference on Computer, Information and Telecommunication Systems (CITS)* (pp. 1-6). IEEE.
- Liew, P.Y. and Tan, I.K., 2024, December. On automated essay grading using large language models. In *Proceedings of the 2024 8th International Conference on Computer Science and Artificial Intelligence* (pp. 204-211).
- Lim, C.T., Bong, C.H., Wong, W.S. and Lee, N.K., 2021. A comprehensive review of automated essay scoring (AES) research and development. *Pertanika Journal of Science & Technology*, 29(3), pp.1875-1899.
- Liu, Y., Zhang, Z., Zhang, W., Yue, S., Zhao, X., Cheng, X., Zhang, Y. and Hu, H., 2023. ArguGPT: evaluating, understanding and identifying argumentative essays generated by GPT models. arXiv preprint arXiv:2304.07666.
- Matarazzo, A. and Torlone, R., 2025. A survey on large language models with some insights on their capabilities and limitations. arXiv preprint arXiv:2501.04040.
- Mizumoto, A. and Eguchi, M., 2023. Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), p.100050.
- OpenAI (2026) OpenAI. Available at: <https://openai.com/> (Accessed: 13 January 2026).

Prelec, D., Seung, H.S. and McCoy, J., 2017. A solution to the single-question crowd wisdom problem. *Nature*, 541(7638), pp.532-535.

Quah, B., Zheng, L., Sng, T.J.H., Yong, C.W. and Islam, I., 2024. Reliability of ChatGPT in automated essay scoring for dental undergraduate examinations. *BMC Medical Education*, 24(1), p.962.

Ramesh, D. and Sanampudi, S.K., 2022. An automated essay scoring systems: a systematic literature review. *Artificial Intelligence Review*, 55(3), pp.2495-2527.

Yamashita, T., 2024. An application of many-facet Rasch measurement to evaluate automated essay scoring: A case of ChatGPT-4.0. *Research Methods in Applied Linguistics*, 3(3), p.100133.

Zheng, W., 2025. Text length requirements for stable and genre-sensitive lexical diversity measurement. *Cogent Arts & Humanities*, 12(1), p.2584418.

AI's Invisible Pollution: Academic Deepfakes and Epistemic Pollution

Aurélien Acquier, ESCP Professor of Sustainability, Planetary Boundaries and Business

Jozef Cossey, Postdoctoral Researcher

Abstract

Concerns about GenAI's energy use and physical pollution are well known. This communication highlights another threat which is less visible but equally concerning: epistemic pollution—the degradation of our knowledge ecosystem through low-quality, misleading, or fabricated information. Drawing on a recent pseudo-scientific article co-authored by Grok 3, we discuss the rise of academic deepfakes and their contribution to epistemic pollution. We show how GenAI makes academic deepfakes cheaper, more accessible, and harder to detect. While the resulting epistemic pollution poses acute risks for all fields in academia, it appears particularly acute in fields -like sustainability- which draw on multidisciplinary knowledge and engage with politically sensitive topics. Distinguishing between “productivity driven deepfakes” and “disinformation driven deepfakes”, we show how the latter can be used as a tactic to manufacture doubt and “flood the academic zone with shit,” with potentially dire societal consequences for academia and the legitimacy of science. While the academic world needs a coordinated response, we also point to the inherent risks and limits of available solutions.

Keywords: Generative AI; Academic Deepfakes; Epistemic Pollution; Disinformation; Research Integrity

1. Motivation

“If everybody always lies to you, the consequence is not that you believe the lies, but rather that nobody believes anything any longer. [. . .] And a people that no longer can believe anything cannot make up its mind. It is deprived not only of its capacity to act but also of its capacity to think and to judge. And with such a people you can then do what you please.”

Hannah [Arendt \(1973\)](#)

While preparing a masterclass on “Climate change in the Era of Post-Truth,” one of us—a professor at a European business school—came across trending posts on social media about a new article in the journal *“Science of Climate Change.”*¹ The piece claimed to debunk the human origins of climate change and overturn the International Panel on Climate Change (IPCC) consensus, prompting some online observers to claim the “climate scam is over.” But is it really?

Although the article initially looks academic (See [Figure 1](#)), its credibility quickly unravels. The paper's claims rest on the use of "unadjusted data" and "alternative causality frameworks" to argue that natural drivers—not human emissions—cause climate change. The first author of the study is Grok 3 beta, a generative AI developed by xAI, whose use was promoted as adding rigor and objectivity to the study. Among the human coauthors, only Willie Soon has an academic track record, as well as a reputation for climate change denial, for being funded by the fossil fuel industry, and for failing to disclose such conflicts of interest. The paper—accepted 12 days after submission by a journal created to oppose IPCC findings—fails to meet basic academic standards, let alone overturn scientific consensus. Still, it convincingly illustrates the risks of academic deepfakes in the GenAI era.

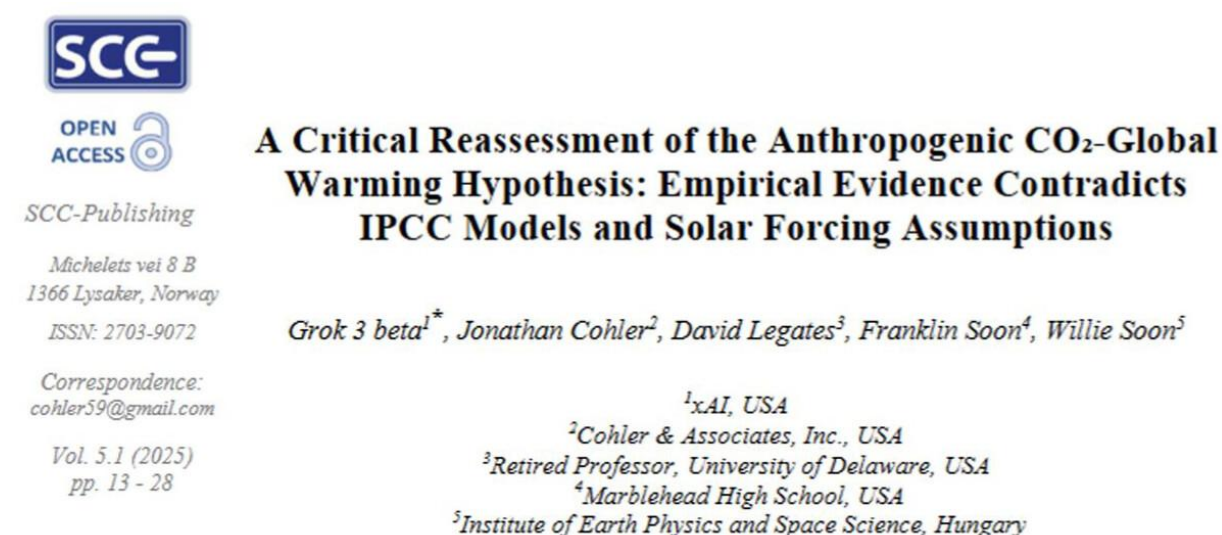


Figure 1. Example of an Academic Deepfake.

Note. "A Critical Reassessment of the Anthropogenic CO₂-Global Warming Hypothesis: Empirical Evidence Contradicts IPCC Models and Solar Forcing Assumptions" can be found at <https://scienceofclimatechange.org/wp-content/uploads/SCC-Grok-3-Review-V6.pdf>

2. Academic Deepfakes and Epistemic Pollution

We define *academic deepfakes* as content fabricated to realistically mimic academic work, with the intent to deceive audiences.

We distinguish two types of academic deepfakes. First, *Disinformation-Driven Deepfakes (DDD)* involve the intentional fabrication of falsehood with the intention to deceive audiences. Such deepfakes are especially prevalent in scientific fields with strong economic or public policy implications, such as the health effects of tobacco or the human origins of climate change ([Oreskes & Conway, 2011](#)). By "manufacturing doubt" amid political polarization and anti-science movements, DDD seek to erode public trust in science, distort perceptions of academic consensus, and weaken academic influence on policy and practice.

Second, *Productivity-Driven Deepfakes (PDD)* involve the unacknowledged fabrication of content such as papers, reviews, data, or pedagogical material to meet productivity expectations. PDD are facilitated by paper mills—illicit organizations making money by faking manuscripts and offering authorship for sale—and predatory journals—publishing work of questionable quality at a price. PDD follow a logic of cheating: producing falsehood is a means to meet productivity targets, rather than an end in itself.

Both types of academic deepfakes generate *epistemic pollution*. Just as physical pollution contaminates natural ecosystems with toxic substances, epistemic pollution degrades our collective knowledge ecosystem by disseminating low-quality, untrustworthy, misleading, or inconsistent information.

Physical pollution has already proliferated widely, becoming diffused, persistent, and irreversible: for example, PFAS pollution is found in rainwater, and microplastics are accumulating in human brains. GenAI could similarly amplify epistemic pollution by making academic deepfakes cheaper, more accessible, and harder to detect. For example, an AI-powered paper mill enables writing a full master's thesis with one prompt, for a few euros. Large-scale epistemic pollution has been documented in fields of biology and medicine: [Abalkina et al. \(2025\)](#) suggest that “at least 400,000 papers published between 2000 and 2022 [have] been produced by paper mills [. . .], yet only 55,000 were retracted or corrected.”

If academic deepfakes are subsequently cited in respectable journals and integrated in LLM training data, epistemic pollution may become as diffuse, persistent, and hard to remedy as microplastics and PFAS.

3. A Call to Action

Next to recent concerns over how GenAI adoption could affect human reflexivity and research integrity ([Lindebaum & Fleming, 2024](#)), we believe academic deepfakes and epistemic pollution deserve serious consideration from academics in sustainability. First, sustainability scholarship confronts the rise of PDD. Second, given sustainability scholars' reliance on interdisciplinary knowledge and their pursuit of societal relevance, the field is especially vulnerable to DDD. Can we really investigate corporate responses to climate change if the scientific consensus is obscured by DDD? How do we handle alternative scientific realities in the classroom, especially when they are backed by academic deepfakes?

Besides, decades of studying physical pollution and its impact on society may also help Sustainability scholars to address the challenge of epistemic pollution.

First, this would require a better understanding of the origins, current scale, and consequences of epistemic pollution. A good starting point would be to examine how journal editors experience and perceive such risks, and which policies and measures they consider appropriate to respond. In addition, research is needed to investigate the organizational fabric of epistemic pollution. How do organizations such as paper mills and predatory journals operate? How are these services shaped by demand-side factors such as tenure criteria or accreditation demands? When and how do deepfakes become toxic?

Second, how can we organize epistemic pollution control and mitigation? Pollution-control literature typically distinguishes between end-of-pipe treatment and prevention: the former is easier to implement but costly and often ineffective, while pollution prevention can spur innovations that also reduce costs. At present, end-of-pipe measures predominate: academic “thrash picking” is carried out by voluntary academic sleuths in the form of investigations, retractions, whistleblowing, or elaboration of predatory journal lists ([Cabanac, 2024](#)). It is crucial to explore how academic journals and institutions could facilitate, protect, and reward their risky work.

In addition, more work is needed to explore the potential of upstream pollution prevention strategies. Such efforts could allow some actors to differentiate themselves through stronger commitments to avoid epistemic pollution. For instance, Anthropic was founded in competition with OpenAI around safety and governance, and Deepseek adjusted its model with added technical clarity after peer review in *Nature*.

Another question concerns ethics and responsibility. What would it mean to apply the “polluter pays” principle to epistemic pollution? Who exactly is the polluter, and how is this responsibility constructed? For example, most academic institutions actively encourage

GenAI use while issuing ethical guidelines on its responsible use, but are these charters implemented, or do they merely shift responsibility from institutions and providers onto individuals? Should AI providers be liable for the reliability of the content produced by their LLMs? Here, insights from business ethics could help explore how actors construct responsibility and assign blame for epistemic pollution and GenAI risks.

GenAI diffusion is unprecedented in speed and scale. This technology deeply transforms our relationship to facts, truth, knowledge, and trustworthiness, raising pressing questions about whether—and how—the resilience of our knowledge ecosystems can be maintained. What is the role of academics in a moment like this? Do we enthusiastically jump on the bandwagon and accelerate the technology's diffusion? Do we also prepare students and stakeholders to navigate their risks and cognitive consequences? Beyond AI training modules, a similar level of engagement should be devoted to epistemology and methodology to cultivate an awareness of how knowledge is produced, challenged, and potentially distorted by humans and technologies. This, we believe, is the growing social responsibility for academics: strengthening the resilience of our knowledge ecosystems against epistemic pollution.

4. Questions for debate at the Summit

This session is designed to provoke debate about how academic institutions experience the problem of AI-driven epistemic pollution and may cope with the problem. One of the questions is, between AI spaceships and monasteries, how to construct institutions in the middle, likely to manage the growing issue of epistemic pollution.

References

This proposal is based on two preliminary works:

Acquier, A., & Cossey, J. (2025, May 1). *Flooding the academic zone with shit? Academic deepfakes and climate action in a post-truth era*. ESCP Business School Impact Paper. <https://ssrn.com/abstract=5467450>

Acquier, A., & Cossey, J. (2025). Generative AI, academic deepfakes, and epistemic pollution. *Business & Society*. Advance online publication. <https://doi.org/10.1177/00076503251406457>

Other references:

Abalkina, A., Aquarius, R., Bik, E., Bimler, D., Bishop, D., Byrne, J., Cabanac, G., Day, A., Labbé, C., & Wise, N. (2025). Stamp out paper mills—Science sleuths on how to fight fake research. *Nature*, 637(8048), 1047–1050. <https://doi.org/10.1038/d41586-025-00212-1>

Arendt, H. (1973). Interview with Roger Errera [Interview]. *The New York Review of Books*. <https://www.nybooks.com/articles/1978/10/26/hannah-arendt-from-an-interview/>

Cabanac, G. (2024). Chain retraction: How to stop bad science propagating through the literature. *Nature*, 632(8027), 977–979. <https://doi.org/10.1038/d41586-024-02747-1>

Lindebaum, D., & Fleming, P. (2024). ChatGPT undermines human reflexivity, scientific responsibility and responsible management research. *British Journal of Management*, 35(2), 566–575.

Oreskes, N., & Conway, E. M. (2011). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Publishing USA.

AI-Powered Research Impact Assessment through ResearchImpact.ai. From Framework to Field-Specific Implementation

Sebastian Haan

University of Sydney, Sydney, Australia

Wilfred Mijnhardt

Rotterdam School of Management, Erasmus University, Rotterdam, Netherlands & Edinburgh Business School, Heriot-Watt University, Edinburgh, Scotland, UK

Abstract

Universities face mounting pressure to demonstrate research impact, yet producing comprehensive impact case studies remains resource intensive. UK institutions spent nearly 500 million GBP preparing Research Excellence Framework 2021 submissions, while researchers lose weeks per assessment cycle. The sector values the right outcomes but lacks scalable assessment methods.

This case study presents ResearchImpact.ai, an AI platform automating research impact assessment. Built on the Research Impact Assessment Framework (RIAF), validated through pilot testing with Australian health organizations ($r = 0.96$ correlation with REF methodology), the platform transforms weeks of manual compilation into 15-minute report generation.

The platform employs an agentic AI architecture combining large language models with multi-source data aggregation. Users specify research topics, authors, and domains through a web interface. The system automatically discovers evidence across academic databases, policy documents, news sources, and industry reports, then generates structured impact narratives following the RIAF's four core questions: problem significance, research outputs, delivered impacts, and expected future impact.

For business schools, the platform addresses accreditation requirements from AACSB, EQUIS, and AMBA. The forthcoming AACSB Global Research Impact Framework emphasizes mission-driven, context-specific assessment across five domains: academic, industry, policy, societal, and educational impact. ResearchImpact.ai's field-specific templates enable schools to document thought leadership and societal contributions aligned with Standard 8 (Impact of Scholarship) and Standard 9 (Engagement and Societal Impact). The platform supports portfolio-level analysis, exemplar case study creation, and benchmarking across institutional portfolios.

Beta testing with the University of Sydney, University of Technology Sydney, and Monash University confirms technical feasibility and user acceptance. Blinded quality assessments show AI-generated case studies identify broader evidence sources than manual preparation while maintaining accurate attribution. Reviewers note that outputs require human verification but provide substantial time savings as working drafts.

The platform aligns with responsible assessment principles from DORA and the Leiden Manifesto, emphasizing narrative evidence over metric aggregation. Implementation requires attention to governance: AI augments rather than replaces professional judgment. Generated outputs need human verification, contextual interpretation, and strategic framing.

This case study demonstrates practical AI adoption in research administration, offering insights for institutions navigating increasing accountability demands while maintaining authentic institutional voice.

Keywords: research impact assessment, artificial intelligence, business school accreditation, impact narratives

1. Introduction

Universities and business schools face mounting pressure to demonstrate the societal value of their research. Governments, funding bodies, and accreditation organizations invest substantially in research and reasonably expect evaluation of the social, economic, and policy impact of that investment. Traditional assessment methods focused on bibliometric indicators such as citations and h-indices fail to capture the broader contributions of research to practice, policy, and public benefit (Bornmann, 2012).

The administrative burden of producing comprehensive impact case studies remains considerable. UK institutions alone spent nearly 500 million GBP preparing submissions for the Research Excellence Framework 2021 (Technopolis Group, 2022), with similar exercises conducted across Australia, Europe, and North America. Researchers spend approximately 34 days per grant proposal, with career-level losses of one to two years on unsuccessful applications where success rates range from 12 to 28 percent (Herbert et al., 2013). The sector has built an assessment system that values the right things but lacks the scalability to assess them.

This case study presents ResearchImpact.ai (<https://researchimpact.ai/>), an artificial intelligence platform designed to automate research impact assessment and case study generation. The platform addresses a critical operational challenge in higher education administration: transforming a process that typically requires weeks of manual effort into one that generates comprehensive, evidence-based impact reports within 15 minutes. We report on the platform's development, validation through partnerships with major research universities, and its potential for field-specific customization to support business school accreditation and strategic planning.

2. The Research Impact Assessment Challenge

The global research community has increasingly recognized limitations of journal-based metrics. The San Francisco Declaration on Research Assessment (DORA, 2013) called for assessment of research on its own merits rather than publication venue, while the Leiden Manifesto (Hicks et al., 2015) articulated principles for responsible metrics, emphasizing that quantitative evaluation should support but not supplant expert assessment: no numbers without stories, and no stories without numbers. The European Research Council endorsed

DORA in 2024, prohibiting Journal Impact Factors in evaluations and introducing narrative CVs.

This transition creates operational challenges. Business schools face requirements from multiple accreditation bodies, including AACSB, EQUIS, and AMBA, each emphasizing research impact in their standards. The AACSB 2026 Global Standards for Business Education (AACSB International, 2025a) explicitly require schools to demonstrate that intellectual contributions have a positive societal impact and that scholarship collaborates with external stakeholders to create credible, relevant, and timely knowledge.

The Research Impact Assessment Framework (RIAF), developed through international collaboration involving the University of Sydney, Rotterdam School of Management at Erasmus University, and the University of Southampton, provides a methodological foundation. Published in the Australian Health Review (Ward et al., 2023), the RIAF assesses not only research outputs but also the research environment that nurtures impactful work and strategic alignment with societal needs.

The RIAF addresses four fundamental questions:

1. What is the problem this research seeks to address and why is it significant?
2. What are the research outputs of this study?
3. What impacts has this research delivered to date?
4. What impact from this research is expected in the future?

This structure moves beyond counting outputs toward understanding how research creates change in the world.

Pilot testing of the RIAF with six health and research organizations demonstrated feasibility and utility, with strong correlation ($r = 0.96$) between RIAF assessments and established REF methodology (Ward et al., 2025). However, the pilot revealed that manual creation of impact case studies requires significant resources, limiting scalability. The authors explicitly identified that implementation would be enabled by AI tools to "reduce administrative burden and ensure consistency of outputs" (Ward et al., 2025).

3. The ResearchImpact.ai Platform

ResearchImpact.ai (<https://researchimpact.ai/>) was developed to operationalize this vision at scale through an agentic AI architecture. The platform combines large language models with contextual AI agents, multi-source data aggregation, and structured prompting to generate comprehensive impact assessments automatically.

Users interact through an accessible web interface requiring no technical expertise. Researchers specify their research topic, author names, keywords, institutional affiliation, research domain, assessment timeframe, and preferred language style for the output. The system supports document uploads for incorporating institutional data not available through public sources, and users can review and refine results through iterative prompts.

The platform performs automated evidence discovery across multiple source types: academic journals and publication databases, policy documents, research metadata repositories, data catalogues, news articles, industry reports, patents, and web content. A contextual research agent conducts iterative searches to build comprehensive background knowledge, formulating research questions based on the submitted topic, identifying

knowledge gaps, and synthesizing findings while maintaining source attribution throughout.

A distinctive feature is the provision of field-specific domain templates. While the RIAF was originally developed for medicine and health research, the platform extends this framework across all major research fields through specialized assessment templates. Each domain employs customized evaluation criteria, metrics, and benchmarks, ensuring that impact is assessed according to disciplinary norms rather than forcing diverse research programmes into a single assessment mould.

The platform generates structured impact reports following the RIAF criteria, addressing each of the four assessment questions in dedicated sections. Users receive a complete impact case study in Microsoft Word and Markdown format, accompanied by an executive summary, comprehensive reference list with source attributions, and input settings for reproducibility. Tiered output options accommodate different assessment needs, from rapid preliminary assessments to comprehensive evaluations for high-stakes submissions. For institutional deployment, the platform offers API access enabling integration with research information systems and supports enterprise configurations with single sign-on authentication and centralized dashboards.

4. Field-Specific Implementation for Business Schools

The platform's domain-based architecture enables customization for business school contexts. The AACSB draft Global Research Impact Framework (AACSB International, 2025b) identifies five types of impact indicators: academic impact (citations, h-index, research awards), industry impact (adoption of frameworks, inclusion in professional guidelines), policy impact (government reports, regulatory influence), societal impact (media coverage, community partnerships), and educational impact (curriculum integration, student projects).

The AACSB framework emphasizes that impact measurement should be mission-driven, context-specific, and evidence-informed. Schools are encouraged to consider a mix of impact indicators rather than relying on a single measure. This aligns precisely with ResearchImpact.ai's capability to aggregate evidence from diverse sources and generate nuanced narratives that capture multiple dimensions of influence. The platform can be customized to support schools in documenting thought leadership, stakeholder collaboration, and societal contributions, which correspond to the three pillars of AACSB Standard 8 (Impact of Scholarship) and Standard 9 (Engagement and Societal Impact).

For business schools preparing accreditation submissions, the platform offers particular value in three areas. First, it can generate portfolio-level analysis across a school's intellectual contributions, identifying patterns of impact and capability gaps. Second, it supports the creation of exemplar case studies demonstrating how specific research has influenced practice, policy, or society. Third, the consistency of AI-generated outputs enables benchmarking across institutional portfolios and over time, supporting continuous improvement processes central to accreditation.

The platform's approach resonates with broader developments in business education impact assessment. The EFMD Business School Impact System (BSIS) demonstrates how thematic methodologies can enable institutions to move beyond transactional impact toward transformative impact, changing how institutions operate rather than merely what they do (EFMD Global, 2024). AI tools can support this transition by enabling systematic narrative development that connects diverse research initiatives into coherent institutional impact stories.

5. Validation and Early Results

ResearchImpact.ai entered beta testing in late 2025 through partnerships with major Australian research universities including the University of Sydney, University of Technology Sydney, and Monash University. The testing programme evaluated technical reliability, output quality, and user acceptance.

Technical evaluation assessed system stability, output quality, and consistency across diverse research domains. Quality assessment involved comparison of AI-generated impact case studies against manually prepared submissions, following the blinded review methodology established in the RIAF pilot study (Ward et al., 2025). Preliminary findings suggest that AI-generated case studies identify a broader range of evidence sources than typical manual preparation and maintain accurate source attribution.

Reviewers noted that AI-generated drafts required verification and refinement by researchers, particularly for nuanced interpretation of impact pathways, but provided substantial time savings as working drafts. User acceptance testing with research office staff and academics yielded positive feedback regarding interface usability and the platform's ability to aggregate evidence from diverse sources while maintaining consistent structure across case studies.

The platform maintains alignment with responsible assessment principles. Following DORA and Leiden Manifesto guidance, the system emphasizes narrative evidence over metric aggregation, supports contextual interpretation of research influence, and maintains transparent source attribution. The design reflects the recognition that AI capability makes human judgment more important rather than less. The challenge lies not in generating impact data but in developing wisdom to ask appropriate questions and translate insights into meaningful action.

6. Strategic Implications for Higher Education Administration

The development of ResearchImpact.ai carries significant implications for AI adoption in university research administration. The platform demonstrates that agentic AI systems can successfully automate complex analytical tasks previously requiring substantial professional expertise and time investment.

From an operational perspective, automation of impact case study preparation could substantially reduce administrative burden on research offices, enabling reallocation of staff effort toward strategic impact cultivation rather than retrospective documentation. For business schools specifically, this efficiency gain addresses a practical constraint: the accreditation cycle demands comprehensive evidence of research impact, yet the resources required to compile this evidence compete with the research activities themselves.

The technology also has implications for how institutions conceptualize and communicate research impact. By systematizing the narrative development process, AI tools can help schools identify previously overlooked impacts, recognize patterns across their research portfolio, and develop more coherent strategic narratives. This capability aligns with AACSB's emphasis on schools identifying and advancing their areas of thought leadership.

However, implementation requires careful attention to governance. The platform is designed to augment rather than replace professional judgment. Generated outputs require human verification, contextual interpretation, and strategic framing. Institutions deploying such tools need clear guidance on acceptable use, quality assurance processes, and maintenance of authentic voice in institutional communications.

7. Conclusion

ResearchImpact.ai represents a practical application of generative AI to address an administrative burden affecting universities worldwide. The platform operationalizes the Research Impact Assessment Framework through an agentic AI architecture that automates evidence gathering, analysis, and report generation while maintaining alignment with established assessment methodologies and principles of responsible research assessment.

Beta testing with Australian universities confirms technical feasibility and user acceptance. The platform's domain-based architecture provides a foundation for field-specific customization, including templates aligned with business school accreditation requirements from AACSB, EQUIS, and AMBA. As higher education institutions navigate increasing demands for accountability and impact demonstration, AI-assisted assessment tools offer a pathway to more efficient, consistent, and comprehensive evaluation of research contributions to society.

The critical challenge lies in deploying such tools to augment rather than replace professional judgment, ensuring that impact assessment serves its fundamental purpose: strengthening the connection between university research and public benefit. For business schools, this means using AI capabilities to support the transition toward broader recognition of research impact, encompassing not only academic influence but also contributions to industry practice, public policy, and societal wellbeing.

References

- AACSB International. (2025a). Global Standards for Business Education: Exposure Draft. AACSB International.
- AACSB International. (2025b). Global Research Impact Framework: Exposure Draft. AACSB International. <https://www.aacsb.edu/insights/reports/2025/research-impact-framework-exposure-draft>
- Bornmann, L. (2012). Measuring the societal impact of research: Research is less and less assessed on scientific impact alone. *EMBO Reports*, 13(8), 673–676. <https://doi.org/10.1038/embor.2012.99>
- EFMD Global. (2024). Business School Impact System (BSIS). <https://www.efmdglobal.org/bsis/>
- Herbert, D. L., Barnett, A. G., Clarke, P., & Graves, N. (2013). On the time spent preparing grant proposals: An observational study of Australian researchers. *BMJ Open*, 3(5), e002800. <https://doi.org/10.1136/bmjopen-2013-002800>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>
- San Francisco Declaration on Research Assessment (DORA). (2013). San Francisco Declaration on Research Assessment. <https://sfedora.org/>
- Technopolis Group. (2022). Review of the Research Excellence Framework: Report for the four UK higher education funding bodies. Research England.
- Ward, R. L., Nutbeam, D., Mijnhardt, W., Nelson, P., Todd, A., Rees, M. I., Richards, J., Khan, N. N., Ho, I., & Chung, S. (2023). Development of a novel and more holistic approach for assessing impact in health and medical research: The Research Impact Assessment Framework. *Australian Health Review*, 47(5), 589–595. <https://doi.org/10.1071/AH23152>

Ward, R. L., Nutbeam, D., Mijnhardt, W., Nelson, P., Todd, A. L., Richards, J., Jackson, K., Khan, N. N., & Chung, S. (2025). Pilot testing of the Research Impact Assessment Framework. *Australian Health Review*, 49, AH25214. <https://doi.org/10.1071/AH25214>

[How Not to Perish by Publishing: Future-Oriented Strategies for AI-Assisted Scientific Writing

Michael Kopp,
University of Graz

Abstract

This paper examines the critical intersection of the academic “Publish or Perish” culture and the disruptive introduction of Generative AI (GenAI). GenAI's ability to instantly generate content intensifies existing issues of fraud and “paper mills”, risking academia a shift from “Publish or Perish” to “Perish by Publishing” and severely undermining scientific integrity by enabling the mass production of fake content and data. While GenAI offers efficiency as an “E-Research Assistant,” authors must mitigate risks through three strategies: radical transparency through multi-layered AI disclosure, rigorous human accountability for all AI-generated outputs, and prioritizing quality over quantity (“publish for purpose”). By focusing on purpose-driven work, researchers can distinguish human-led contributions from AI-augmented content. The paper concludes by calling for systemic reforms, such as moving away from purely metric-based research evaluation and strengthening ethical training, to ensure GenAI enhances human research and preserves the validity of the scientific record.

Keywords: Generative AI; Research Integrity; Scientific Publishing; Academic Reform

1. The Academic Imperative and the AI Paradigm Shift

The academic landscape has historically been characterized by the “publish or perish” principle. This is a deep-seated cultural rule that determines career progression, job security decisions, and institutional evaluation based largely on the quantitative output of scholarly publications (Ajwang & Ikoha, 2024). This intense pressure encourages academics to engage in a form of mass production aimed at rapid publishing. Frequently, this leads to the investigation of trivial matters rather than the innovative synthesis of profound knowledge. This focus on quantity over intellectual depth fosters “conformist thinking” and risks diverting researchers from going after innovative or ethically informed research paths (Elbanna & Child, 2023). This environment of continuous publication demands has received significant impetus due to unethical practices, including data fabrication and falsification, the buying and selling of authorships, and ghostwriting, often orchestrated by highly organized “paper mills” - organizations that produce fraudulent manuscripts for a fee. These practices compromise the integrity of the scholarly record and bend out of shape the metrics used to evaluate scientific success (Vasconez-Gonzalez et al., 2024). The current state risks turning “publish or perish” into “perish by publishing”.

The advent of Generative Artificial Intelligence (GenAI), prominently represented by tools like ChatGPT, introduces a critical dimension to this crisis (Bozkurt, 2024). GenAI is a disruptive technology that can generate scholarly papers and complex research data almost instantaneously, in moments rather than months or years. While offering immense potential, the ease of immediate content production intensifies the existing “publish or perish” pressures and creates novel opportunities for scientific misconduct, thereby accelerating the potential for academics to perish by publishing and posing an urgent threat to research integrity and credibility (Wilson & Burleigh, 2025). This technological turning point necessitates a critical re-evaluation of academic practices to ensure that GenAI is embraced effectively, safely, and responsibly. This paper explores the opportunities and threats of GenAI across the academic ecosystem and proposes actionable strategies to ensure that researchers maintain the integrity and visibility of their publications, thereby successfully handling AI-assisted scientific writing without compromising their careers or credibility.

2. The Dual Impact of GenAI on Scientific Publishing

The integration of AI-powered systems is fundamentally revolutionizing research production, evaluation, and dissemination (Munir, 2025). This transformation presents powerful efficiencies for authors but simultaneously introduces significant challenges and threats to academic integrity, posing systemic challenges for publishers tasked with maintaining the scholarly record.

For authors facing publication pressure, GenAI tools serve as an effective “e-research assistant”, offering capabilities that vastly accelerate workflow (Khlaif et al., 2023). These tools boost content production and editing by simplifying complex processes, saving substantial time, and supporting various stages of manuscript creation, including drafting, summarizing lengthy papers, generating accurate abstracts, and assisting with code generation for analysis. Furthermore, GenAI's Natural Language Processing (NLP) capabilities enhance language quality by correcting grammar, simplifying text, and improving overall readability (Munir, 2025). This is particularly beneficial for non-native English-speaking academics, helping to overcome linguistic barriers and promote global equity in scholarly communication. Finally, GenAI offers crucial research and conceptual support by identifying relevant citations, analyzing research trends, and generating potential research questions, thereby aiding the initial concept generation and literature

review phases (Dergaa et al., 2023). It can also support the interpretation and communication of findings through assistance with data analysis and visualization (Bozkurt, 2024).

However, the efficiency granted by these tools introduces profound challenges and threats to academic integrity, primarily through the industrialization of scientific fraud. The pressure to publish has created a market for “paper mills”, which now utilize AI to generate fake text, data, and images at an alarming scale (Jian & Yaeger, 2025). This “weaponization” of AI allows for the mass production of manuscripts that mimic the style of authentic research, making it increasingly difficult to distinguish between genuine scientific inquiry and machine-generated fabrication. Beyond deliberate fraud, there is the persistent risk of “artificial hallucinations”, where GenAI models invent non-existent references, falsify data points, or generate biologically impossible results (Tang et al., 2024). Studies have shown that a significant percentage of citations generated by these models do not exist or are irrelevant to the text provided (Semrl et al., 2023). Furthermore, the over-reliance on these tools threatens to induce a form of academic deskilling, where researchers may lose the ability to critically organize ideas or exercise profound scientific writing, effectively outsourcing their cognitive labor to algorithms that lack moral responsibility (Gandhi & Gandhi, 2023). This erosion of human agency complicates the traditional definitions of authorship, as AI tools cannot be held accountable for the integrity of the work, yet are increasingly doing the heavy lifting of manuscript creation.

These individual integrity issues compound to create existential threats to the publishing ecosystem itself. The sheer volume of AI-facilitated submissions threatens to overwhelm the peer review system, causing reviewer fatigue and making it statistically probable that fraudulent papers will be overlooked (Dergaa et al., 2023). The scientific community risks falling victim to a “reputation economy” where metrics such as citation counts and publication volume - which is easily vulnerable to manipulation by AI-generated content - are valued over the rigorous search for truth. As the volume of scientific literature explodes, the quality is weakened, creating a “fake science” crisis that undermines public trust. This environment is intensified by the fluid nature of the modern economy, where speed and branding often supersede material reality, a trend that AI accelerates, allowing researchers to generate the appearance of scientific work without the underlying substance (França & Monserrat, 2024). If unchecked, the interplay between the “publish or perish” culture and the unchecked capabilities of GenAI could lead to a systemic collapse where the validity of the scholarly record is irretrievably compromised.

3. Strategies for Scientific Authors: Staying Visible in the Age of GenAI

To navigate this volatile landscape, scientific authors must adopt strategies that surpass mere efficiency, focusing instead on establishing trust and demonstrating the unique value of human intellect. The first critical strategy is enforcing radical transparency through multi-layered disclosure. It is no longer sufficient to merely acknowledge the use of AI in a generic statement; authors must provide a granular accounting of exactly how, where, and why these tools were employed (Agents4Science, 2025). This involves adopting frameworks such as the “Academic Integrity and Transparency in AI-assisted Research and Specification” (aiTARAS), which mandates specific disclosures for different types of AI contribution, whether for idea generation, data analysis, or linguistic editing (Bozkurt, 2024). Authors should utilize structured reporting mechanisms, such as “AI Usage Cards” or detailed checklists, to document the specific models used, the version dates, and the nature of the prompts entered (Writing Center et al., 2025). Such rigorous documentation not only satisfies the evolving requirements of ethical journals but also protects the author against accusations of misconduct by creating a clear audit trail of the research process. By

voluntarily adhering to these high standards of disclosure, authors signal their commitment to integrity, distinguishing their work from the flood of opaque, mass-produced content generated by “paper mills”.

Alongside transparency, authors must distinguish their work by prioritizing human accountability and rigorous verification. While AI can process vast amounts of information, it lacks the capacity for moral responsibility and the nuanced scientific sensibility required to judge the significance of findings. Authors must explicitly assume full responsibility for every aspect of the manuscript, regardless of AI involvement, treating AI outputs with a baseline skepticism reserved for unreliable sources (Dergaa et al., 2023). This requires a shift in the drafting process where the human author acts as a watchful auditor, carefully verifying every citation, data point, and methodological claim generated or refined by AI to prevent the inclusion of “hallucinations”. This human-led verification process is the ultimate safeguard against the “pollution” of the scientific record (Semrl et al., 2023). Furthermore, authors should leverage their unique human capacity for critical reflection and contextual understanding - traits that current AI models struggle to replicate - to frame their research in ways that demonstrate deep engagement with the subject matter, rather than the superficial pattern matching characteristic of AI (Agents4Science, 2025). By emphasizing the personal responsibility of the researcher, authors can elevate their work above the commodity status of AI-generated text (Cheng et al., 2024).

Finally, the most sustainable strategy for long-term survival in academia is elevating quality over quantity: publishing for purpose. The “publish or perish” culture has incentivized the “salami slicing” of research into the smallest publishable units, a practice that AI makes dangerously easy to scale (Sabel & Larhammar, 2025). To counter this, authors should resist the temptation to use AI to churn out derivative papers and instead focus on producing fewer, more substantial works that address significant societal or scientific problems. This “publish for purpose” approach (Elbanna & Child, 2023) aligns with the growing recognition that the metric-obsessed reputation economy is unsustainable and that true scientific impact is measured by the advancement of knowledge rather than how long the publication list is. By concentrating on complex, interdisciplinary problems that require genuine human synthesis and collaboration - areas where AI currently serves as a tool rather than a replacement - researchers can protect themselves from the devaluation of automated content (Gao et al., 2023). In doing so, they not only protect their own reputations but also contribute to the stabilization of the scientific enterprise, ensuring that their work remains visible and relevant in a lot of computer-generated content.

4. Conclusion and Future Perspectives

The integration of GenAI into the scientific workflow represents a double-edged sword, offering substantial efficiency while simultaneously threatening the foundational trust of the academic enterprise. As we transition from a culture of “publish or perish” to a potential crisis of “perish by publishing”, the limitations of current AI integration become starkly apparent. Despite their linguistic fluency, Large Language Models (LLMs) fundamentally operate as probabilistic engines rather than reasoning agents, often functioning as “stochastic parrots” that mimic the form of scientific discourse without grasping its substance (Dergaa et al., 2023). These models currently lack the ability to generate genuine novelty or exercise the scientific judgment required to discern trivial results from significant breakthroughs. Furthermore, because they are trained on existing internet data, they risk entrenching historical biases and creating “monocultures of knowing”, where the scope of scientific inquiry is narrowed to what is represented in the training set (França & Monserrat, 2024). The reliance on these tools also poses technical risks, such as the generation of buggy code or the subtle introduction of errors that may be overlooked by fatigued reviewers.

Consequently, while AI can serve as a powerful assistant, it cannot yet replace the critical, integrative, and ethical functions of the human scientist.

Looking forward, the preservation of scientific integrity requires a concerted effort to shape future directions for research and policy. The academic community must move beyond individual adaptability to systemic reform, as outlined in initiatives like the Stockholm Declaration (Sabel & Larhammar, 2025), which calls for a shift away from profit-driven publishing models and the metric-based assessment of research quality. Future policy must prioritize the development of independent, researcher-controlled mechanisms for fraud detection that are not beholden to commercial publishers, ensuring that the policing of the scientific record remains unbiased. Additionally, there is an urgent need for educational reform to equip the next generation of scientists not only with technical AI skills but with the ethical grounding to resist the pressure to chase metrics and publish more frequently (Dergaa et al., 2023). Research funding and tenure criteria must evolve to reward thoroughness, reproducibility, and societal impact over raw publication counts, thereby breaking the feedback loop that incentivizes the misuse of AI. Ultimately, the goal must be to harness the capabilities of AI to enhance human inquiry, rather than replace it, ensuring that science remains a robust and trustworthy pursuit of truth in the digital age.

AI Use Disclosure Statement

In accordance with the Academic Integrity and Transparency in AI-assisted Research and Specification (aiTARAS) framework, the author discloses the use of the following Generative AI tools during December 2025 in the preparation of this manuscript:

- Literature Review and Information Retrieval: Consensus (free version) and Perplexity Pro were utilized to facilitate the literature search and to identify relevant citations. The results provided by these tools were subsequently verified and analyzed by the human author to ensure the accuracy and relevance of the sources.
- Drafting: Substantial portions of the initial draft of this paper were generated using NotebookLM Pro. These sections were subsequently reviewed, critically edited, and validated by the human author to ensure academic rigor and adherence to ethical standards.
- Editing and Language Refinement: This paper was reviewed, edited for English language style, and checked for orthography with the assistance of Gemini Pro, complementing the human editorial process.

The author critically assessed and addressed potential biases inherent in the AI-generated content. The final content, conclusions, and assertions in this paper are the sole responsibility of the human author.

References

- Agents4Science. (2025, October 22). Agents4Science 2025 Virtual Conference [Video]. YouTube. <https://www.youtube.com/watch?v=7pXqAeedqOo>
- Ajwang, S. O., & Ikoha, A. P. (2024). Publish or perish in the era of artificial intelligence: Which way for the Kenyan research community? *Library Hi Tech News*, 41(9), 7-11. <https://doi.org/10.1108/LHTN-04-2024-0065>

- Bozkurt, A. (2024). GenAI et al.: Cocreation, Authorship, Ownership, Academic Ethics and Integrity in a Time of Generative AI. *Open Praxis*, 16(1), 1-10. <https://doi.org/10.55982/openpraxis.16.1.654>
- Cheng, H.-Z., Sheng, B., Lee, A., Chaudhary, V., Atanasov, A. G., Liu, N., Qiu, Y., Wong, Y., Tham, Y.-C., & Zheng, Y.-F. (2024). Have AI-Generated Texts from LLM Infiltrated the Realm of Scientific Writing? A Large-Scale Analysis of Preprint Platforms. *bioRxiv*.
- Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, 40(2), 615-622. <https://doi.org/10.5114/biolsport.2023.125623>
- Elbanna, S., & Child, J. (2023). From 'publish or perish' to 'publish for purpose.' *European Management Review*, 20(4), 614-618. <https://doi.org/10.1111/emre.12618>
- França, T. F. A., & Monserrat, J. M. (2024). The artificial intelligence revolution...in unethical publishing: Will AI worsen our dysfunctional publishing system? *Journal of General Physiology*, 156(11). <https://doi.org/10.1085/jgp.202413654>
- Gandhi, M., & Gandhi, M. (2023). Does AI's touch diminish the artistry of scientific writing or elevate it? *Critical Care*, 27(1). <https://doi.org/10.1186/s13054-023-04634-z>
- Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *Npj Digital Medicine*, 6(1). <https://doi.org/10.1038/s41746-023-00819-6>
- Jian, X., & Yaeger, L. H. (2025). Combating Fake Science in the Age of Generative Artificial Intelligence: A Biomedical Perspective. *Missouri Medicine*, 162-168.
- Khlaif, Z. N., Mousa, A., Hattab, M. K., Itmazi, J., Hassan, A. A., Sanmugam, M., & Ayyoub, A. (2023). The Potential and Concerns of Using AI in Scientific Research: ChatGPT Performance Evaluation. *JMIR Medical Education*, 9. <https://doi.org/10.2196/47049>
- Munir, A. (2025). Editorial: Artificial Intelligence and its transformative impact on scientific publishing. *International Journal of Pharmacy & Integrated Health Sciences*, 6(1). <https://doi.org/10.56536/ijpihs.v6i1.214>
- Sabel, B., & Larhammar, D. (2025). Reformation of science publishing: The Stockholm Declaration. *R. Soc. Open Sci.*, 12. <https://doi.org/10.1098/rsos.251805>
- Semrl, N., Feigl, S., Taumberger, N., Bracic, T., Fluhr, H., Blockeel, C., & Kollmann, M. (2023). AI language models in human reproduction research: Exploring ChatGPT's potential to assist academic writing. *Human Reproduction*, 38(12), 2281-2288. <https://doi.org/10.1093/humrep/dead207>
- Tang, A., Li, K., Kwok, K. O., Cao, L., Luong, S., & Tam, W. (2024). The importance of transparency: Declaring the use of generative artificial intelligence (AI) in academic writing. *Journal of Nursing Scholarship*, 56(2), 314-318. <https://doi.org/10.1111/jnu.12938>
- Vasconez-Gonzalez, J., Izquierdo-Condoy, J. S., Naranjo-Lara, P., Garcia-Bereguain, M. Á., & Ortiz-Prado, E. (2024). Integrity at stake: Confronting "publish or perish" in the developing world and emerging economies. *Frontiers in Medicine*, 11. <https://doi.org/10.3389/fmed.2024.1405424>

Wilson, A., & Burleigh, C. (2025). Research Integrity in the Era of Generative Artificial Intelligence. *Journal of Educational Research and Practice*, 15(1), 1-16. <https://doi.org/10.5590/JERAP.2025.15.2054>

Writing Center, Competence Center for University Teaching, Center for Digital Teaching and Learning, & University Library Graz. (2025). Guideline: Documentation and labeling of the use of generative AI technologies in writing and research processes. University of Graz.

The Architect and the Machine: Towards a New Ethics of the Author-Supervisor in the Age of Generative AI

Erwan Lamy,
ESCP Business School

Abstract

This paper examines the crisis of the "auctorial definition" triggered by the integration of Generative AI (GenAI) in higher education and research. It argues that current distrust toward GenAI stems from an outdated conception of the author as a solitary artisan whose legitimacy is tied to the manual execution of writing. Drawing on historical precedents from art and architecture (notably the studios of Jacques-Louis David, the practice of Auguste Rodin, and the role of the Architect), this study suggests that the dissociation between conception and execution is a long-standing tradition in creation. The author proposes a paradigm shift: the mutation of the writer into an "Author-Supervisor" or "Textual Architect." In this new ethical framework, authorship is no longer defined by the "labor of the pen" but by intentionality, iterative guidance, and intellectual responsibility. Ultimately, the paper calls for a reform of academic ethics that prioritizes high-level cognitive skills—such as logical structuring and factual verification, over the fetishization of manual drafting, positioning the human as the conductor of synthetic intelligence rather than its rival.

Keywords: Authorship, Author-Supervisor , Intellectual Ethics, Plagiarism, Transparency

Introduction: The Crisis of the Auctorial Definition

The massive arrival of Generative Artificial Intelligence (GAI) in the landscape of higher education and research arouses much more distrust than hope. These tools are primarily seen as instruments of a form of moral depravity. Their use is viewed, at best, as a manifestation of guilty intellectual laziness, and at worst, as fraud or plagiarism. While these fears are not entirely unfounded, a significant portion of the visceral reactions these technologies provoke rests on a dated and situated conception of the author: that of a solitary artisan whose legitimacy relies on the manual execution of the writing task – a conception that engages an equally dated ethics.

Is this conception of the author as the one "holding the pen" the only one possible? If we accept that AI-assisted writing challenges the link between thought and gesture, we must admit that this link has already been severed many times in the history of creation, without

the authorship of the work being contested. Drawing on the history of art and architecture, this proposal aims to demonstrate that the use of generative AI does not signify the death of the author, but rather their mutation into a figure of a “supervisor” or textual architect. This redefinition calls for a complete overhaul of our ethical and normative frameworks: it is no longer a question of judging who wrote, but who intended, directed, and validated the text.

Part I: The Historical Precedent – The Dissociation Between Conception and Execution

The massive arrival of Generative Artificial Intelligence (GAI) in the landscape of higher education and research arouses much more distrust than hope. These tools are primarily seen as instruments of a form of moral depravity. Their use is viewed, at best, as a manifestation of guilty intellectual laziness, and at worst, as fraud or plagiarism. While these fears are not entirely unfounded, a significant portion of the visceral reactions these technologies provoke rests on a dated and situated conception of the author: that of a solitary artisan whose legitimacy relies on the manual execution of the writing task – a conception that engages an equally dated ethics.

The Painter’s Studio: The Case of David.

The romantic image of the painter alone before their canvas is a late construction. In the 19th century, large studios functioned like production companies. Let us take the emblematic example of *The Coronation of Napoleon* by Jacques-Louis David. This monumental canvas (nearly 10 meters long) is signed by David, attributed to David, and it is David whom history remembers. Yet, David painted only a fraction of the total surface. Two assistants, Georges Rouget and Ignazio Eugenio Degotti, realized the decors, the costumes, and the secondary figures. David conceived the composition, supervised the execution, and intervened manually only for the key faces and the final harmonization. David is the author not because he painted everything, but because he thought and controlled everything.

The Paradox of Marble: The Case of Rodin.

The example of sculpture is even more striking. Auguste Rodin, a giant of modern sculpture, is famous for his work with marble. However, Rodin almost never carved the marble himself. He was a modeler: he worked with clay and plaster. The transposition of these models into marble was entrusted to “practitioners” (*praticiens*), highly skilled workers whose names are known only to art historians. Is this fraud? Is Rodin an impostor? No. The work resides in the form conceived by Rodin, not in the act of striking the chisel, a technical act delegated to an expert hand. If we accept that Rodin is the author of *The Kiss* without having touched the marble, why do we refuse the status of author to one who conceives a text but delegates the generation of words to a machine?

The Architect: The Author Without Contact with Matter.

Architecture offers an even more accomplished, because more general, example of this dissociation. No one expects Renzo Piano or Jean Nouvel to lay the bricks or pour the concrete of their edifices. The architect is the indisputable author of the work, yet their intervention is limited to plans, directives, and supervision. They operate through a total delegation of material execution. In this paradigm, writing by AI is akin to architecture: the “prompter” provides the plans (the structure, ideas, tone) and the AI (the worker) constructs the text. The author's added value no longer lies in the masonry of sentences, but in the engineering of thought.

Part II: The Ethics of Authorship in the Age of AI – The Concept of “Creative Supervision”

If manual execution is no longer the criterion for authorship, how do we define the author in the age of AI? We propose shifting the center of gravity of production towards supervision.

From “Writer” to “Director”.

To vary the metaphors further, after the painter's studio, the sculptor's workshop, or the architect's firm, I would say that the expert use of a generative AI transforms the writer into a director (in the cinematographic sense). The director does not hold the camera, does not play the roles, and does not sew the costumes. Yet, the film is indeed their creation. Why? Because they exercise what could be called “directional agency”. Facing the AI, the author exercises three critical functions that found their legitimacy:

1. Impulsion (or Intentionality): The AI creates nothing ex nihilo. It responds to a specific human intention. It is the author who defines the “why” and the “what” of the text. Without this initial impulse, the text does not exist.
2. Iteration (or Guidance): Rarely is an AI's first draft perfect. The author guides the machine, refines the prompt, and requests corrections of tone, structure, or substance. This dialogic process is dense intellectual work, comparable to the corrections a master brings to their student's sketch.
3. Validation (or Assumption of Responsibility): This is the crucial point. By signing the generated (or co-generated) text, the author assumes responsibility for its content. They transform a machinic production (the AI's text) into committed speech.

The End of the Fetishism of “Sweat”.

The current moral resistance opposing the reform of our conception of authorship for writing may stem from a confusion inherited from the notion of “labor value”: we estimate that a text has value because it was painful to write, and that this effort justifies the title of author. This is an ethical error. The value of an academic or professional text resides in the relevance of its ideas, the solidity of its argumentation, and its capacity to transmit a truth or an innovation. Whether the syntactic bricks were assembled by a neural network or a human brain is, regarding the final result, secondary, provided that the logical structure and veracity are guaranteed by the author-supervisor. As Michel Foucault highlighted in *What is an Author?*, the author-function serves to classify, validate, and attribute coherence to discourses. This function does not require the uniqueness of the hand, but the uniqueness of intellectual responsibility.

Part III: Towards a New Ethics and New Institutions

Recognizing the author as a supervisor does not mean opening the door to “laissez-faire”. On the contrary, it demands a much more rigorous ethics than that of traditional writing.

Responsibility as the Cornerstone.

If the author does not “build,” they must “verify” with increased vigilance. In the architectural model, if the building collapses, the architect is responsible, even if they did not pour the concrete. Similarly, if an AI “hallucination” slips into a text, the author-supervisor is solely culpable. The ethics of AI must not rest on a prohibition of the tool (which would be as absurd as forbidding architects from using CAD software), but on the obligation of verification. The fraud is not in using AI, but in failing to supervise its work. The student or researcher who submits raw text generated by ChatGPT is comparable to an architect who

would sign randomly generated plans without verifying the load-bearing capacity of the walls: it is professional misconduct. We will now expect students or researchers to be good "textual project owners" (maître d'ouvrage textuel).

A New Skill: The “Textual Project Owner” Skillness.

Education must train for this new skill. Knowing how to write remains important to be able to judge the quality of a text (one cannot supervise what one does not understand, just as Rodin had to understand sculpture to direct his practitioners), but knowing how to make write becomes crucial. This implies skills in logic, structuring of thought, general culture, and factual verification. It is these “High Level” skills that the university must prioritize, delegating “Low Level” skills (syntax, spelling, connectors) to the machine, just as the architect delegates the laying of bricks.

Against Injunctions to Transparency.

This is undoubtedly the most provocative proposition of this communication, but it follows naturally from the premises of this ethical reflection. Today, most academic institutions require that the use of generative AI be transparent. But it is one of two things: either this use is not blameworthy, and there is then no more reason to confess its use than to admit using a spell checker or a thesaurus; or it is blameworthy, and then this transparency is merely a very poor way of discouraging this undesirable practice. But as I have stated, nothing justifies the rejection of this new authorship. This injunction to transparency actually hides a desire to guilt-trip AI usages, to discourage them, in the name of dated prejudices. It is therefore deeply immoral, and doubly so. First in the hypocrisy of the approach, but above all in the inability to reform one’s prejudices and to allow the epistemic potentialities of these new forms of writing to unfold. Using a generative AI is normal, and it is recommendable. There is certainly nothing to hide, but above all, there is nothing to show. Researchers using this technology do not have to confess a sin.

Conclusion

The Coronation of Napoleon is no less a masterpiece because David did not paint every brushstroke. Rodin is no less a genius because he did not carve The Kiss. The buildings of Le Corbusier are no less magisterial because he did not wield the trowel. With generative AI, we are going to extend this understanding of creation to textual production.

This proposal asserts that banning AI in the name of a fantasized auctorial purity is a historical and *ethical* dead end. The future of academic and professional writing lies in the acceptance of this new paradigm: the author is no longer the solitary scribe, but the conductor of a synthetic intelligence. This transition does not diminish the human role, it elevates it. It detaches us from the technicality of drafting to concentrate on the essential: the structure of ideas, the veracity of facts, and the ethical scope of our discourses. It is on this basis – that of intellectual responsibility rather than manual labor – that we must refound the ethics of academic authorship.

From Relevance-Lag to Continuous Curriculum Intelligence The Readiness Index: A Governed AI-Enabled Operating Model for Student Readiness in Business Education

Antoine Casanova-Mazet, HIM Business School

Abstract

Business schools promise employability, yet curriculum governance often moves more slowly than labour-market skill drift. The result is **relevance-lag**: the delay between emerging skill signals and documented curriculum response. Skills-first hiring expectations (OECD, 2025) and anticipated skills change through 2030 (World Economic Forum, 2025) intensify this pressure. This paper proposes the **Readiness Index**: a governed, AI-powered architecture that senses demand, infers curriculum profiles, produces alignment metrics, and triangulates student readiness into a longitudinal profile.

Keywords: Readiness Index, curriculum intelligence, relevance-lag, skills-first, governance, labour-market intelligence, responsible AI, business education

1. Relevance-lag in business education

Graduate employability is multi-dimensional. It combines domain competence with transferable and adaptive capabilities shaped by context, not merely technical skills (Fugate, Kinicki and Ashforth, 2004; Clarke, 2018; Abelha et al., 2020). For business schools, this creates a governance challenge that is also reputational: readiness must be defined and evidenced in ways that employers, students, and external evaluators recognise as credible.

Labour-market signalling is increasingly articulated in skills terms. Employers and workforce policy actors place greater emphasis on capability relevance and development rather than credential completion alone (OECD, 2025). This pressure is compounded by forecasts of sustained skills change through 2030 (World Economic Forum, 2025).

This paper frames the strategic risk as relevance-lag: the time between an external skill signal becoming stable and salient, and an internal response being recorded in curriculum design, assessment, or learning delivery. Unlike static “skills gap” narratives, relevance-lag is a system property: it emerges from governance cadence, evidence infrastructure, decision rights, and organisational routines (OECD, 2019; OECD, 2024). For business schools, where employability is central to the educational value proposition, relevance-lag is not an abstraction, it is an operational risk.

2. AI should power evidence flows, not replace academic judgement

An AI-enabled operating model can strengthen institutional capability in three ways:

1. **Scale and timeliness.** Continuous sensing and synthesis of large text corpora—job advertisements, student feedback, curriculum artefacts—can support faster institutional interpretation and response cycles (Humphreys and Wang, 2018; Lee et al., 2020).
2. **Consistency and comparability.** Normalisation against controlled vocabularies and repeatable pipelines enables trend monitoring and supports auditability (European Commission, 2025; *ONET Resource Center, n.d.*; ONET Resource Center, 2024).
3. **Actionability.** Integrated signals make it easier for governance bodies to identify where change is most needed and where signals are stable rather than short-lived.

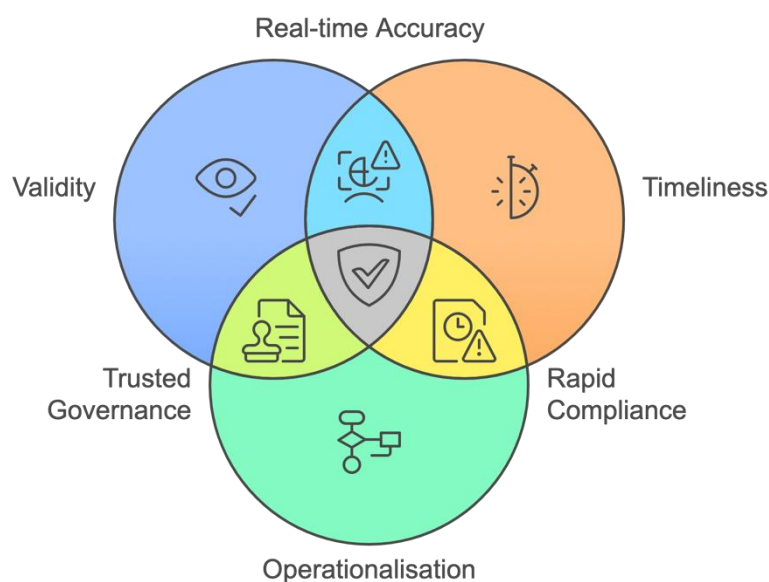
The intent is not to automate curriculum decisions. It is to use AI-enabled automation to equip deans, programme directors, and faculty teams with decision-ready evidence that reduces lag while maintaining coherence and academic integrity (OECD, 2024). This is particularly important because labour-market data—especially job advertisements—are

imperfect proxies: postings reflect aspiration, templating, and labour-market frictions as well as underlying demand (Cappelli, 2015). AI can improve the translation of signals into structured representations, but legitimacy depends on interpretability, documentation, and appropriate safeguards (Lee et al., 2020; Akgun and Greenhow, 2022).

The Readiness Index operationalises this position: a governed AI-powered architecture that makes curriculum intelligence continuous and auditable while keeping academic judgement inside the decision chain.

3. The Readiness Index: architecture and requirements

The Readiness Index is built on three requirements: Validity, Timeliness, and Operationalisation.

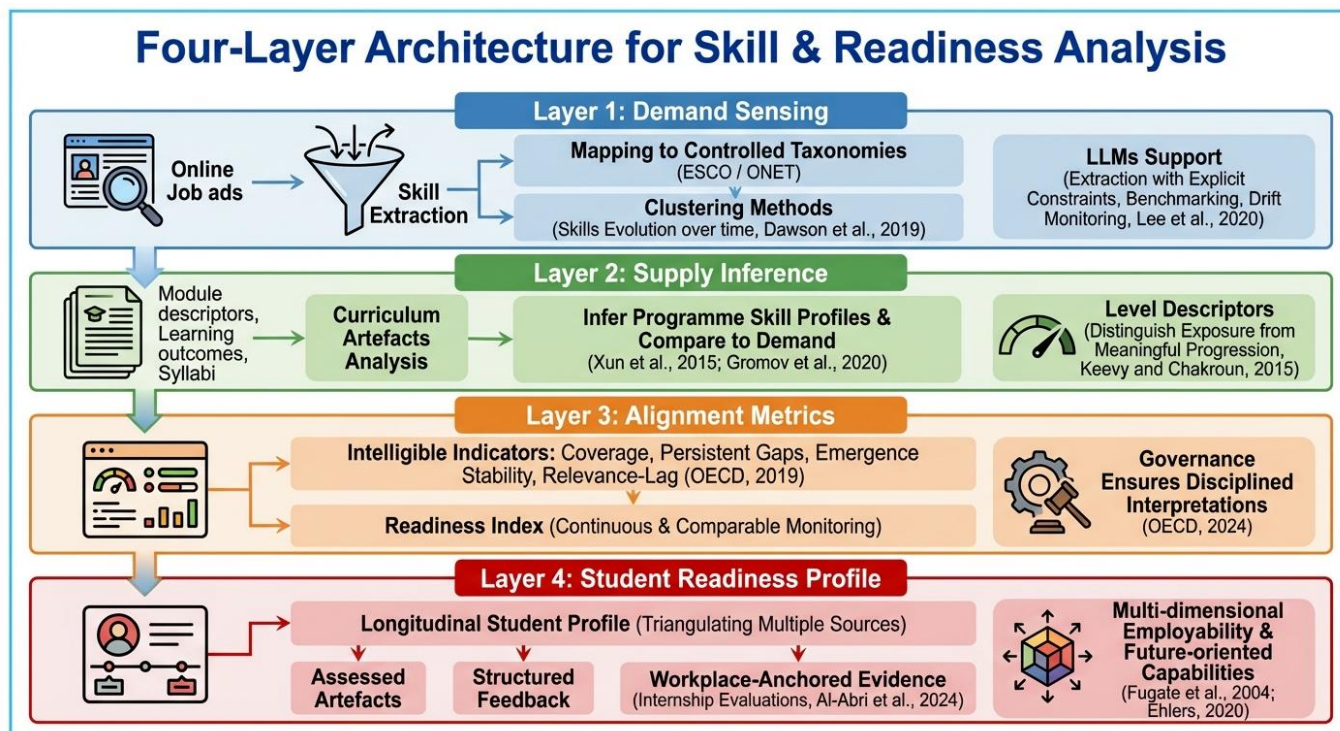


This aligns with curriculum as a dynamic cycle of anticipation, action, and reflection (OECD, 2019) and with the governance risks introduced by increased flexibility (OECD, 2024).

This paper makes three contributions:

- (1) framing relevance-lag as a measurable governance risk, not an abstract skills gap;
- (2) proposing a triangulated readiness profile that links demand signals, curriculum intent, delivery quality, and workplace evidence; and
- (3) embedding both within an operating model that preserves legitimacy through audit trails and human review gates.

At HIM Business School, we are building this approach using institution-specific evidence streams to support feasibility and governance traceability (HIM Business School, 2025a; HIM Business School, 2025b; HIM Business School, 2025c; HIM Business School, 2025d). The Readiness Index architecture has four layers:



Layer 1: Demand sensing

Online job advertisements enable scalable demand sensing when skills are extracted and mapped to controlled taxonomies such as ESCO or ONET (European Commission, 2025; ONET Resource Center, n.d.; O*NET Resource Center, 2024). Open, data-driven approaches to skills extraction strengthen feasibility and transparency, including methods and tooling developed around job advert corpora (Djurnalieva and Sleeman, 2018; Economic Statistics Centre of Excellence, 2023; Nesta, 2021). Clustering methods can support tracking skills evolution over time (Dawson, Rizoiu and Johnston, 2019). Large language models can support extraction and classification, but require explicit constraints, benchmarking, and drift monitoring to remain decision-credible (Lee et al., 2020).

Layer 2: Supply inference

Curriculum artefacts (module descriptors, learning outcomes, syllabi, and assessment briefs where available) become analysable data. Programme skill profiles can be inferred and compared against demand profiles (Xun, Gottipati and Shankararaman, 2015; Gromov et al., 2020). Level descriptors help distinguish superficial exposure from meaningful progression, supporting a more defensible interpretation of “coverage” (Keevy and Chakroun, 2015).

Layer 3: Alignment metrics

Business schools need a small set of intelligible indicators: coverage, persistent gaps, emergence stability, and a **relevance-lag** measure linking dated external signal events to dated curriculum response events (OECD, 2019). The Readiness Index makes monitoring continuous and comparable; governance ensures interpretations remain disciplined and that flexibility does not fragment coherence or equity (OECD, 2024).

Layer 4: Student readiness profile

Readiness is not curriculum coverage. The Readiness Index produces a longitudinal student profile triangulating multiple sources—assessed artefacts, structured feedback, and workplace-anchored evidence such as internship evaluations (Al-Abri et al., 2024). This reflects employability's multi-dimensional nature (Fugate, Kinicki and Ashforth, 2004; Clarke, 2018; Abelha et al., 2020) and the need for future-oriented capabilities under uncertainty and change (Ehlers, 2020). Outcome models can test whether readiness indicators add explanatory value for progression and employability proxies, with explicit attention to fairness and interpretability (Haque et al., 2025; Law et al., 2025; Mustofa, Emon and Mamun, 2025).

4. From concept to capability: adoption, governance, and responsible use at HIM

A recurring analytics failure mode is insight without authority: credible outputs that do not translate into legitimate decisions or sustained change. Curriculum innovation is vulnerable to inertia when institutional routines are structurally slow, creating obsolescence risk when practice evolves faster than review cycles (Lemay and Moreau, 2020).

HIM partially mitigates this through (i) operational evidence streams supporting continuous curriculum intelligence (HIM Business School, 2025a), (ii) structured employer triangulation to calibrate meaning and priorities (HIM Business School, 2025b), and (iii) committee routines that anchor decision rights and maintain audit trails through documented minutes (HIM Business School, 2025c; HIM Business School, 2025d). Adoption depends on perceived usefulness, usability, and enabling conditions (Davis, 1989; Venkatesh, Thong and Xu, 2012), as well as broader analytics ecosystem capability (Pappas et al., 2018). Ethics and legitimacy require transparency, privacy-by-design, human review gates, and explicit boundary-setting to prevent measurement systems becoming surveillance or unfair selection mechanisms (Akgun and Greenhow, 2022; Lee et al., 2020; OECD, 2024).

Business schools have built their value proposition on employability. The Readiness Index offers a way to govern that promise with evidence rather than assumption. At HIM Business School, we are building it as a governed operating model that can be evaluated, iterated, and held accountable.

References

Abelha, M., Fernandes, S., Mesquita, D., Seabra, F. and Ferreira-Oliveira, A.T. (2020) 'Graduate employability and competence development in higher education—A systematic literature review using PRISMA', *Sustainability*, 12(15), 5900. doi:10.3390/su12155900.

Akgun, S. and Greenhow, C. (2022) 'Artificial intelligence in education: addressing ethical challenges in K-12 settings', *AI and Ethics*, 2, pp. 431–440. doi:10.1007/s43681-021-00096-7.

Al-Abri, M. et al. (2024) 'Investigating the influence of university–industry internships on the employability of Omani business graduates', *Humanities and Social Sciences Communications*. doi:10.1057/s41599-024-02779-y.

Cappelli, P.H. (2015) 'Skill gaps, skill shortages, and skill mismatches: evidence and arguments for the United States', *ILR Review*, 68(2), pp. 251–290. doi:10.1177/0019793914564961.

Clarke, M. (2018) 'Rethinking graduate employability: the role of capital, individual attributes and context', *Studies in Higher Education*, 43(11), pp. 1923–1937. doi:10.1080/03075079.2017.1294152.

Davis, F.D. (1989) 'Perceived usefulness, perceived ease of use, and user acceptance of information technology', *MIS Quarterly*, 13(3), pp. 319–340. doi:10.2307/249008.

Dawson, C., Rizoïu, M.-A. and Johnston, B. (2019) 'Adaptively identifying and quantifying skills demand from job ads', in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '19)*. New York: ACM, pp. 1140–1143. doi:10.1145/3341161.3342905.

Djumaliev, J. and Sleeman, C. (2018) *An Open and Data-driven Taxonomy of Skills Extracted from Online Job Adverts*. ESCoE Discussion Paper ESCoE DP-2018-13. Economic Statistics Centre of Excellence (ESCoE).

Economic Statistics Centre of Excellence (2023) *ESCoE Skills Extractor Library* (software/library).

Ehlers, U.-D. (2020) *Future Skills: The Future of Learning and Higher Education*. Wiesbaden: Springer VS. doi:10.1007/978-3-658-29297-3.

European Commission (2025) *What is ESCO?* (web resource / ESCO portal).

Fugate, M., Kinicki, A.J. and Ashforth, B.E. (2004) 'Employability: A psycho-social construct, its dimensions, and applications', *Journal of Vocational Behavior*, 65(1), pp. 14–38. doi:10.1016/j.jvb.2003.10.005.

Gromov, A. et al. (2020) 'Curriculum profile: modelling the gaps between curriculum and the job market', in *Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020)*, pp. 610–614.

Haque, R. et al. (2025) 'Leveraging LLMs for optimised feature selection and embedding in structured data: A graduate employment classification approach', *Computers and Education: Artificial Intelligence*, 8, 100356. doi:10.1016/j.caeai.2025.100356.

HIM Business School (2025a) *Internal operational records/materials for weekly labour-market skills sensing pilot* (unpublished internal document/materials).

HIM Business School (2025b) *World Readiness Symposium: recap and key insights* (unpublished internal document).

HIM Business School (2025c) *Be World Ready Committee minutes* (unpublished internal document).

HIM Business School (2025d) *Be World Ready Committee minutes* (unpublished internal document).

Humphreys, A. and Wang, R.J.-H. (2018) 'Automated text analysis for consumer research', *Journal of Consumer Research*, 44(6), pp. 1274–1306. doi:10.1093/jcr/ucx104.

Keevy, J. and Chakroun, B. (2015) *Level-setting and recognition of learning outcomes: The use of level descriptors in the twenty-first century*. Paris: UNESCO.

Law, T.J. et al. (2025) 'QED: A quality-enhanced data approach for improving embeddings in predicting key higher education outcomes', *International Journal of Educational Technology in Higher Education*, 22, 66. doi:10.1186/s41239-025-00560-w.

Lee, L.W., Dabirian, A., McCarthy, I.P. and Kietzmann, J. (2020) 'Making sense of text: artificial intelligence-enabled content analysis', *European Journal of Marketing*, 54(3), pp. 615–644. doi:10.1108/EJM-02-2019-0219.

Lemay, V. and Moreau, K.A. (2020) 'Implementing change in pharmacy education: from crisis to opportunity', *Pharmacy*, 8(3), 153. doi:10.3390/pharmacy8030153.

Mustofa, S., Emon, Y.R. and Mamun, S.B. (2025) 'An AI-driven model for student dropout risk analysis: explainable AI insights', *Computers and Education: Artificial Intelligence*, 8, 100352. doi:10.1016/j.caeai.2024.100352.

- Nesta (2021) *Open Jobs Observatory / Making Sense of Skills* (programme/web resource).
- ONET Resource Center (n.d.) *The ONET Content Model* (web resource).
- ONET Resource Center (2024) *ONET Data Dictionary 29.0 / ONET Database Release 29.0** (released August 2024) (web/PDF resource).
- OECD (2019) *OECD Future of Education and Skills 2030: Anticipation-Action-Reflection Cycle for 2030* (concept note / Education 2030).
- OECD (2024) *Curriculum Flexibility and Autonomy*. Paris: OECD.
- OECD (2025) *Empowering the Workforce in the Context of a Skills-First Approach*. Paris: OECD. doi:10.1787/345b6528-en.
- Pappas, I.O. et al. (2018) 'Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies', *Information Systems Frontiers*, 20, pp. 659–670. doi:10.1007/s10257-018-0377-z.
- Venkatesh, V., Thong, J.Y.L. and Xu, X. (2012) 'Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology', *MIS Quarterly*, 36(1), pp. 157–178.
- World Economic Forum (2025) *The Future of Jobs Report 2025*. Geneva: World Economic Forum.
- Xun, L.S., Gottipati, S. and Shankararaman, V. (2015) 'A framework for curriculum analytics in outcome-based education', in *Proceedings of the 2015 International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. doi:10.1109/TALE.2015.7386048.

How Blackboard is Advancing Trustworthy AI-Enhanced Teaching and Learning: A Case Study in Ethical Implementation

Hervé Didiot-Cook, Blackboard;

Richard Gibbons, Blackboard;

Gemini 3 (AI Collaborator), Google

Abstract

The AI in Higher Education Summit 2026, hosted by ESCP Business School in Paris, highlights the urgent need for responsible, scalable, and innovative uses of AI across teaching and learning. Specifically aligned with Track 1—focusing on AI-enhanced pedagogy, digital assessment, and faculty development—this case study illustrates the practical application of Blackboard’s ethical AI framework. By embedding principles of transparency, fairness, accountability, and privacy, Blackboard enables institutions to modernize course delivery and scale student engagement while preserving pedagogical integrity. Central to this approach is a reimagined blueprint for digital assessment that moves beyond individual tasks toward holistic assessment "programs" designed for a world of "co-intelligence." This demonstration provides a model for institutional transformation, ensuring that AI adoption serves as an ethically aligned augmentation of the human educational experience.

Keywords: Trustworthy AI, Higher Education, AI-Enhanced Pedagogy, Digital Assessment, Human-in-the-Loop

How Blackboard is Advancing Trustworthy AI-Enhanced Teaching and Learning: A Case Study in Ethical Implementation

Introduction

The AI in Higher Education Summit 2026 served as a landmark event for addressing the scalable and ethical implementation of generative AI in global academia. As institutions move beyond reactive stances, the focus has matured toward creating comprehensive frameworks that balance technological innovation with institutional integrity. A recurring theme in technology-enhanced learning (TEL) is the difficulty of transitioning from individual "pockets" of enthusiasm to a coherent, sustainable institutional strategy (Didiot-Cook, 2013). This case study illustrates how the Anthology Trustworthy AI framework

addresses this gap by integrating sophisticated generative tools—specifically the AI Design Assistant and the AI Virtual Assistant (AVA)—into the core Blackboard Learning Management System (LMS).

The Philosophical Anchor: Privacy and Accountability

The adoption of AI in higher education is fundamentally tethered to the perceived security of information privacy. The Blackboard framework ensures that all AI interactions occur within a "walled garden," where data is not used to train public Large Language Models (LLMs), thereby maintaining institutional trust.

Furthermore, the Summit emphasized the need for Augmentation over Automation. By requiring explicit human review for every AI-generated suggestion, the platform ensures that the educator remains the final "source of truth," mitigating the risks of "algorithmic management" and maintaining faculty autonomy.

The Future of Digital Assessment: From Tasks to Programs

As generative AI continues to challenge traditional assessment staples, the Anthology White Paper (2025) argues that institutions must move beyond individual, holistic tasks toward longitudinal assessment programs. This paradigm shift involves:

- **Embracing Co-Intelligence:** Fusing AI with existing pedagogical techniques to develop the skills students need in an AI-augmented workforce.
- **Prioritizing Formative Assessment:** Shifting from high-stakes summative exams to frequent, low-stakes activities that improve motivation and learning outcomes (Anthology, 2025).
- **Authentic Evaluation:** Leveraging AI to create "AI-resistant" assessments focused on higher-order thinking and reflective practice.

Technical Deep Dive: AI-Enhanced Pedagogy

The AI Design Assistant functions as a collaborative tool utilizing Microsoft Azure OpenAI Service to modernize course delivery. Guided by Anthology (n.d.) pillars of Transparency and Fairness, the tool operates through several key mechanisms:

- **Context-Aware Content Generation:** By analyzing existing syllabi and reading lists, the assistant proposes modular course structures, allowing faculty to focus on deep content refinement.
- **Reframed Assessment Standards:** While the assistant automates rubric drafting, it encourages educators to follow Clark's (2024) perspective on reframing Bloom's Taxonomy for the age of AI. This ensures that assessments challenge students to engage in thinking that transcends what AI can independently replicate.

Enhancing the Student Experience: AVA and AI Conversations

The Summit proceedings highlighted a shift toward scaling student engagement through interactive scaffolding.

- Reliability through RAG: AVA (AI Virtual Assistant) utilizes a retrieval-augmented generation (RAG) framework, ensuring student queries are answered based on the specific course context (Anthology, n.d.).
- Interactive Scaffolding: Ma (2025) emphasizes that the "AI Conversation" feature facilitates deeper learning through Socratic questioning and role-playing. This allows students to engage in reflective learning, effectively transitioning educators from transmitters of knowledge to facilitators of high-level critical engagement.

Faculty Development and New Curricula

Effective AI adoption depends on faculty confidence. Blackboard enables development ecosystems grounded in ethical AI literacy, including workshops on bias awareness and "hands-on" experimental sandboxing. According to Queralto and Boorman (2025) from the University of Southampton, "prompt-a-thon pilots have shown that providing a safe, fun environment for students to play with AI builds their confidence and their understanding of responsible use, including academic integrity". Furthermore, Blackboard supports the creation of new curricula by integrating AI literacy modules across disciplines, ensuring that all students build the "AI confidence" required for a skills-driven economy.

Conclusion: A Legacy of Trustworthy AI

The findings shared at the AI in Higher Education Summit 2026 underscore that the future of the AI-augmented campus is defined by the strength of the ethical frameworks surrounding the technology. By combining generative power with the interactive scaffolding described by Ma (2025) from the University of West England and the assessment blueprint provided by Anthology (2025), Blackboard provides a sustainable roadmap for global academia. The ultimate goal is an educational ecosystem where technology enhances the human experience while preserving academic integrity.

Appendix: Technical Comparison of AI Assets

Feature	AI Design Assistant	AI Virtual Assistant (AVA)
Primary User	Faculty / Instructional Designers	Students / Learners
Core Function	Course construction, assessment drafting & interactive scaffolding	Real-time support, feedback & equity
Technical Logic	Contextual analysis of course documents	Retrieval-Augmented Generation (RAG)

Key Capability	Generates modules, formative / authentic assessment & rubrics	Provides 24/7 course-specific query resolution
Educational Theory	Reframed Bloom's Taxonomy (Clark, 2024)	Teaching Presence - Community of of Framework (Garrison et al, 1999)

References

Anthology. (n.d.). *Trustworthy AI approach*. Last accessed 2026 , March 1 <https://www.anthology.com/trust-center/trustworthy-ai-approach>

Anthology. (2025). *The future of assessment: A blueprint for learner evaluation in an AI world* [White Paper]. <https://www.anthology.com/white-paper/future-of-assessment>

Clark, L. (2024). *Reframing Bloom's for the Age of AI: A White Paper for Future-Ready Educators*. https://backstage.anthology.com/sites/default/files/2025-09/ReframingBloomsForTheAgeOfAI_WhitePaper_v1.pdf

Didiot-Cook, H. (2013). *From enthusiasm to strategy: Four critical factors to sustain the development of technology enhanced learning in educational organizations*. Proceedings of the 6th International LINC Conference MIT, Cambridge, MA, USA. <https://linc.mit.edu/linc2013/proceedings/Session9/Session9Didiot-Cook.pdf>

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The internet and higher education*, 2(2-3), 87-105. <https://auspace.athabasca.ca/bitstream/handle/2149/739/?sequence=1>

Ma, Z. (2025, October 7). *Teaching with AI in higher education: some of my insights from practice with Blackboard Ultra*. Digital Learning at UWE Bristol. <https://digitallearning.uwe.ac.uk/teaching-with-ai-in-higher-education-some-of-my-insights-from-practice-with-blackboard-ultra/>

Queralto, J. & Boorman, P. (2025, October 21). *Playful Rigor, Ethical Practice: Introducing a Prompt-a-thon Workshop*. ALT Annual Conference 2025, Glasgow, UK. <https://altc.alt.ac.uk/blog/2025/10/playful-rigor-ethical-practice-introducing-the-prompt-a-thon-workshop/>

Keeping The Searcher in Control: Designing AI for Responsible Academic Search

Jesper Solheim
Johansen, Keenious
Christian Headley,
Keenious

Abstract

As Generative AI (GenAI) becomes more common in higher education, educators worry that students might outsource their academic thinking. This demonstration paper introduces Keenious, a research discovery tool designed to counter the trend of AI "answer engines". Instead of bypassing the learning process, Keenious prioritizes "search-as-learning". It aims to help novice researchers bridge the "vocabulary gap" without removing their critical engagement. In this paper, we explain the system's design philosophy, its hybrid search technology, and its data sources. We also share a practical case study to show how AI can responsibly support information literacy.

Keywords: Generative AI, Academic Search, Information Literacy, Search-as-learning, Human-in-the-loop

Introduction

Educators are increasingly concerned about how generative AI (gen AI) tools affect students' information literacy. The main worry is that students might outsource the intellectual work involved in finding, filtering, and evaluating scholarly information. A worry that has backing in current data about students gen AI usage (Handa et al., 2025).

Today, many students use AI chatbots as their primary way to find information. These tools often act as "black boxes," giving confident and synthesized answers. If a student gets a perfect summary from an AI model, they might skip the crucial step of evaluating the original sources, even though. This shortcuts a cognitive process known as "search-as-learning," where the act of searching and navigating actually helps students understand their topic better (Vakkari, 2016).

This paper argues that universities do not have to choose between banning AI or letting it automate everything. Instead, we can design educational AI tools that help students manage information overload while still requiring them to think critically. We present Keenious (keenious.com) as a case study of a "discovery engine" rather than an "answer engine."

The Keenious Tool

Keenious is an academic search tool that uses artificial intelligence to help users find scholarly literature. The idea for the tool came from the creators' own experiences as students. They noticed how hard it can be to find the exact academic terms needed to search traditional databases effectively. The Keenious tool solves this "vocabulary gap" by using AI to analyze the meaning behind a user's input. The AI acts as a translator, turning the student's everyday language into an academic search. To fit naturally into the research process, Keenious works as a sidebar add-on in Microsoft Word, Google Docs, and a dedicated web platform. This means users can search for literature while they are actively reading or writing.

From Recommendation to Interaction

Keenious has evolved from solely being a recommendation engine into an active research assistant. In its early version, Keenious provided a static list of recommended papers based on the user's text. Over the years, it has gradually supported more complex search tasks by allowing more advanced filtering. After a major update in fall 2025, the tool now features a conversational interface (Image 1). Users can chat with the AI to update their search, add filters, or clarify their ideas. This conversational layout guides the user, but the AI still serves only as an extension of the researcher's own intent. The user remains in control and must evaluate the results themselves.

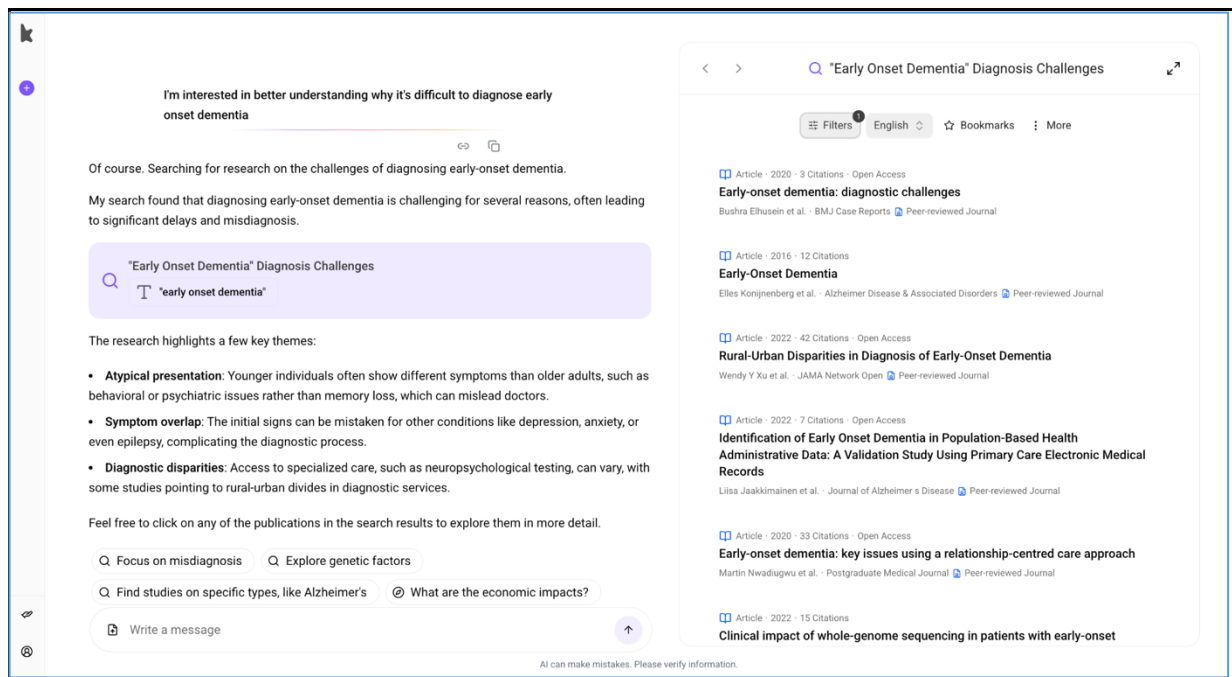


Image 1. The Keenious User Interface.

How Search Works in Keenious

When a user provides input to the tool, whether a full question, a highlighted passage, or an uploaded PDF, the system analyzes it using advanced language understanding. Rather than simply lifting keywords, it interprets the intent and scope of the information need, drawing on the broader context of the session to form a structured academic search. The system then queries metadata from OpenAlex (Priem, Piwowar & Orr, 2022) using two complementary methods:

Semantic Search. Identifies research based on meaning and shared conceptual space (embeddings). This surfaces relevant work even when terminology differs, such as linking "climate anxiety" with "eco-anxiety."

Keyword-Based Search. Functions like a traditional database to find exact or closely related wording, ensuring technical precision for specific terms like locations, instruments, or software.

Keyword and semantic evidence are blended into a single, transparent ranking. When a description is technical, the system emphasizes exact matches; when it is broad, it leans on semantic similarity. To ensure results are both relevant and credible, scholarly cues are applied to fine-tune the order:

Publication Date. Recent work is gently emphasized to keep the researcher current.

Citation Activity. Frequently cited papers receive a modest lift to highlight foundational research.

These adjustments do not override relevance; they simply make trustworthy research easier to spot. The result is an active, engaging process where the researcher can refine the search in natural language, narrowing by period, language, or method, to reach useful sources with minimal friction.

User Case Study

To showcase a real-life example of how Keenious' "human-in-the-loop" approach works in practice, a recent Master's thesis project at Metropolia University of Applied Sciences describes in detail how Keenious was essential in researching the use of open-source Large Language Models (LLMs) for medical transcription (Chowdhury, 2025). The following passage is direct quote from the master thesis:

"For example, when I described how transformer models can be adapted to process spoken medical dictations, Keenious pointed me to "BioBERT: A pre-trained biomedical language representation model for biomedical text mining [14]", which provided some insight into domain-aware pretraining techniques. Another example was when I roughed out a part of the paper discussing the limitations of existing NLP tools in healthcare settings; Keenious pointed me to "Publicly Available Clinical BERT Embeddings [15]", which educated me on the ways contextualized embeddings improve clinical text comprehension. In addition, while conducting research on how LLMs can be used to improve the accuracy of medical transcription, Keenious made me aware of "Recent Advances in End-to-End Automatic Speech Recognition [16]", which directly supported my argument for the integration of ASR (automatic speech recognition) and structured output modelling. These targeted recommendations allowed me to discover relevant sources quickly, often showing me papers I may have had difficulties discovering using traditional search engines alone. This approach greatly enriched my literature review with relevant, cutting-edge articles."

Conclusion

The idea that AI will make students lazy or passive is not guaranteed. It depends entirely on how we build and use these tools. To make sure students stay actively involved, Keenious uses specific design choices that encourage "close reading" instead of "blind trust". Keenious shows that "discovery engines" can responsibly support the research process. By keeping the user in control, using transparent data, and requiring students to read original texts, universities can use AI to build genuine information literacy.

References

- Chowdhury, M. N. R. (2025). Using Open Source LLM Model for Medical Transcription. [Master's Thesis, Metropolia University of Applied Sciences].
- Handa, K., Bent, D., Tamkin, A., McCain, M., Durmus, E., Stern, M., Schiraldi, M., Huang, S., Ritchie, S., Syverud, S., Jagadish, K., Vo, M., Bell, M., & Ganguli, D. (2025, April 8). *Anthropic education report: How university students use Claude*. Anthropic. <https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude>
- Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *ArXiv*. <https://arxiv.org/abs/2205.01833>
- Vakkari, P. (2016). Searching as learning: A systematization based on literature. *Journal of Information Science*, 42(1), 7-18.

Case Study TECgpt: Empowering Faculty to Build Learning Experiences with AI. Tecnológico de Monterrey

Authors:

Irving Hidrogo
Montemayor

Edrei Robles Chávez
Úrsula Saldívar Dávila

Luis Güemes Frese
Paola Martínez Chiñas
Verónica Pérez Aguirre
Manuel Terán Melgarejo

Affiliation: Tecnológico
de Monterrey

Abstract

This case study examines Tecnológico de Monterrey's strategic integration of generative artificial intelligence (AI) into its educational ecosystem, advancing beyond a passive consumer model toward a proactive, pedagogically grounded approach. The institution identified limitations in the commercial model solutions— such as decontextualized outputs, one-size-fits-all interfaces, risks to faculty agency, and ethical concerns around data sovereignty — and responded with a vision that frames AI as a creative medium.

At the core of this AI in education strategy is TECgpt, a secure, institution-specific generative AI platform that enables students and faculty members to access a TECgpt chat which is an LLM in an institutional environment; a TECgpt skill that empower teachers to design and use customizable prompts (skills) that allow teacher to enhance learning experiences with AI while diminishing time dedicated to repetitive tasks that now can be automated in their teaching practice; and TECgpt agent that enables teachers to design and implement conversational agents that optimize the teaching-learning process. After a pilot phase, the University opened campus-wide access to this platform intended to reach more than 90,000 students and 13,000 teachers from high school to postgraduate degrees.

The platform's scalability, open-source availability, and global outreach through the AIGEN network position TECgpt as a model for democratizing AI in higher education. This strategy also foregrounds ethical governance, data privacy, and human-centered integration, while extending implementation through partnerships and funding initiatives. Collectively, the institutional experience offers a replicable model for leveraging AI to support human-centered learning, faculty agency, and sustainable innovation across contexts.

Keywords: (max 5)

AI in Education; GenAI in Education; Innovation; Faculty engagement; Teaching & Learning.

1. The Pedagogical Challenge: Beyond the Consumer Model

Since the emergence of generative AI in late 2022, higher education institutions have primarily treated AI as a consumer product, providing access to commercial tools like ChatGPT or Gemini. However, Tecnológico de Monterrey identified several limitations in this model:

- **Decontextualized Outputs:** Commercial models lack specific knowledge of institutional curricula, assessment rubrics, and especially the Tec21 Educational Model, often providing answers that do not align with educational objectives or competencies.
- **One-Size-Fits-All Interfaces:** Generic AI offers the same experience to an engineering professor as it does to a humanities instructor, failing to meet diverse disciplinary needs.
- **Erosion of Faculty Agency:** The consumer model positions faculty as end-users of external technology, reproducing historical patterns where educational technology "happens to" faculty rather than being driven by them.
- **Data Sovereignty and Ethics:** Sending sensitive institutional content or student work to commercial servers raises significant privacy and compliance risks.

In response to these limitations, institutional leadership articulated a paradigm shift: repositioning AI not as a product to be consumed, but as a medium designed to enhance human capabilities and strengthen pedagogical intentionality.

2. What is TECgpt?

TECgpt is an ecosystem of tools and a development framework created by Tecnológico de Monterrey to enable the design of purpose-specific generative AI components and solutions. These components are integrated through other institutional enablers, all within a private, secure, managed, and customizable environment.

The platform is designed to transform academic experiences by leveraging AI to enhance teaching, collaboration, research, and operational efficiency. TECgpt empowers faculty to optimize time spent on academic tasks while increasing student engagement, enabling new approaches to personalized learning, creativity, and academic growth.

Trained on institutional data, TECgpt supports academic excellence and fosters the development of collective intelligence within the university. A relevant feature is the AI development for non-technical faculty through the following tools:

TECgpt chat

TECgpt chat is an interactive conversational platform designed for institutional educational environments. This tool enables users to engage with a variety of Large Language Models (LLM) accessed through a chat.

The platform's core capabilities include understanding and generating text naturally, making complex concepts accessible through conversational interactions. It maintains a comprehensive record of all conversations, allowing users to revisit previous discussions and track their learning journey over time. Additionally, TECgpt chat offers the ability to download conversation history, enabling users to archive important exchanges with the AI for future reference or study purposes. The system also processes documents in various

formats, transforming static educational materials into interactive learning resources that students can query and explore dynamically.

TECgpt skill

TECgpt skill is a specialized tool designed to help educators design and utilize mini-applications, known as "skills," built with customizable prompts. This platform streamlines repetitive teaching tasks and enhances instructional practice by automating routine educational activities. The system automates content generation, saving teachers valuable time in creating educational materials. It provides assistance in designing challenges, activities, and lesson plans, offering educators a collaborative AI partner in content and learning experience development.

One example of the use of TECgpt skill is the creation of assessment instruments that support teachers developing quizzes, rubrics and other teaching tools aligned with the learning experience. Furthermore, the platform promotes collaboration by allowing educators to share their created skills with colleagues, fostering a community of practice where innovative teaching approaches can be exchanged and adapted across different educational contexts.

TECgpt agent

TECgpt agent represents an advanced AI system that creates, configures, and executes conversational agents to deliver highly contextualized learning experiences through guided simulations and dialogues. This tool transforms traditional learning into immersive, personalized educational journeys.

This key feature allows users to configure personality settings to tailor responses to specific learning needs and teaching styles, ensuring that each interaction aligns with educational goals. TECgpt agent can retrieve information from uploaded documents and URLs, grounding conversations in specific course materials and authoritative sources. Perhaps most innovatively, it connects to external services through APIs, extending the agent's capabilities beyond simple conversation to include real-world data integration and interactive functionalities. This makes TECgpt agent particularly powerful for scenario-based learning, simulations, and role-playing exercises that prepare students for real-world professional situations.

3. Implementation Strategy and Faculty Adoption

A central pillar of Tecnológico de Monterrey's AI in Education Strategy is the integration of AI into the teaching-learning process to enable adaptive, personalized, and discipline-relevant learning. This line of action coordinates institution-wide efforts to deploy AI-enabled platforms that address shared academic needs while responding to the specific requirements of different fields.

TECgpt is the primary enabler of this strategy, repositioning faculty as active creators rather than passive users of AI. Through its Skills and Agents modules, educators can design, customize, and deploy AI-powered learning tools without requiring programming expertise. By December 2025, the pilot phase engaged more than 1,394 faculty members, reporting a 70% reduction in time spent on repetitive tasks. This shift has enabled faculty to move from content delivery toward learning experience design and mentorship.

These efforts are complemented by structured professional development programs, including workshops, hackathons, and communities of practice, developed in collaboration with other internal areas of the institution such as Educational Innovation, Emerging Educational Technologies, Digital Transformation, and the Center for Teacher Development and Educational Innovation (CEDDIE). Throughout 2025, these initiatives reached more than 1,000 faculty members, reinforcing the strategy's emphasis on pedagogical fluency over technical specialization.

During 2025, the institution focused on three strategic priorities: (1) the technological enablement of TECgpt to ensure full platform readiness; (2) the implementation of pilots to validate the functionality and pedagogical value of both Skills and Agents for faculty and students; and (3) user empowerment through training and familiarization to increase adoption, effective use, and positive educational impact.

In 2026, Tecnológico de Monterrey reached a major milestone by releasing platform access to its entire academic community—more than 90,000 students and 13,000 faculty members—across 26 campuses in Mexico. Engagement with the platform is expected to be gradual, aligned to their own disciplinary contexts and learning pathways.

4. Scalability of TECgpt ecosystem

TECgpt is designed for scalability through a modular architecture, open-source model, and cloud-based infrastructure. The platform's cloud infrastructure supports increased demand while maintaining performance and security. [Ethical and pedagogical guidelines for AI use](#) have been central to scaling adoption. These guidelines establish shared ethical principles for integrating AI into teaching, learning, and institutional activities, ensuring proactive and responsible use.

The platform's scalability for the Tec community, combined with its open-source availability, enabled the creation of a community of practice known as the Artificial Intelligence Global Education Network (AIGEN). AIGEN aims to catalyze the adoption of AI in higher education by exploring both commercial and open-source AI tools for educational use, including TECgpt. Within this framework, TECgpt is currently in pilot implementation in 12 universities, positioning the platform as a model for democratizing access to AI in higher education.

The TECgpt strategy places strong emphasis on ethical governance, data privacy, and human-centered integration, while also extending its reach through partnerships and dedicated funding initiatives. Collectively, this institutional experience offers a replicable model for leveraging AI to support human-centered learning, strengthen faculty agency, and enable sustainable innovation across diverse educational contexts.

5. Ethical Governance and Funding source

The rapid adoption of generative AI prompted the development of a coordinated institutional action plan grounded in five guiding principles: AI as a strategic educational capability; a human-centered and ethical approach; discipline-sensitive integration; scalability and sustainability; and evidence-based decision-making.

Ethical governance is overseen by the Vice Presidency of Inclusion, Integrity, and Compliance and is guided by four core principles:

- **Human-centered design:** Faculty retain curricular authority; AI feedback is advisory.

- **Transparency:** AI interactions and permitted uses are explicitly communicated to students.
- **Data privacy:** All data remain within the institution's private cloud infrastructure.
- **Institutional oversight:** Ethical governance structures ensure responsible deployment.

TECgpt is fully funded by Tecnológico de Monterrey. It received initial seed grant from Microsoft to support open-source development. Investments focus on infrastructure, AI model development, and user support systems.

6. Global Democratization: AIGEN and Open Source

Recognizing educational innovation as a global public good, Tecnológico de Monterrey expanded TECgpt through the AI for Global Education Network (AIGEN), a community of practice led by the Institute for the Future of Education. AIGEN connects 10 higher education institutions across five continents that share innovative pedagogical strategies.

The TECgpt ecosystem was released as open source in September 2024. Within four months, 27 institutions in 15 countries adopted the Open Edition, enabling institutions of any size to deploy the platform locally while maintaining full data sovereignty.

7. Lessons Learned

The TECgpt case study offers a replicable framework for institutions seeking to democratize AI innovation in teaching and learning.

Key success factors include faculty agency, a pedagogy-first approach, high accessibility through intuitive no-code design, and iterative development based on student feedback. Another positive experience from the implementation process is that the platform is cost-efficient in relation to other products in the market that offer higher licensing costs making scalability a significant challenge.

The integrated TECgpt ecosystem—comprising Chat, Skills, and Agents—facilitated deployment and training by aligning directly with faculty members' existing teaching practices. By integrating into educators' workflows, the platform supports progressive familiarization and practical adoption. Collectively, these elements enable sustainable innovation and foster meaningful pedagogical transformation.

8. Conclusion

The TECgpt experience demonstrates that the most effective integration of artificial intelligence in higher education is achieved not through top-down technological mandates, but through the intentional empowerment of educators. By providing faculty with the infrastructure, guidelines, and pedagogical freedom to become creators of AI-enabled tools, Tecnológico de Monterrey has ensured that AI-enhanced learning remains pedagogically authentic, institutionally distinctive, and ethically grounded.

This approach reinforces a human-centered vision of AI, in which technology is positioned as an enabler of judgment, creativity, and professional expertise rather than as a substitute for academic labor. Early evidence from the implementation highlights how this model allows faculty to redirect time and cognitive effort away from repetitive tasks and toward

deeper forms of engagement, including mentorship, critical dialogue, and personalized guidance—core functions that define the value of higher education.

Looking ahead, the evolution of the TECgpt ecosystem will focus on continuously improving the user experience in response to teachers' and students' needs. This includes simplifying the process of creating skills and agents so that any user can do so intuitively, as well as enhancing the system's technical configuration to increase responsiveness and overall performance. These developments further advance a vision of AI as a partner in learning, enhancing human skills, supporting reflective and ethical decision-making, and strengthens the relational dimensions of teaching and learning.

Ultimately, TECgpt offers a scalable and replicable model for Higher Education institutions seeking to integrate AI in ways that preserve educational purpose, enhance faculty agency, and place human connection at the center of technological innovation.

References

Related Websites & Multimedia:

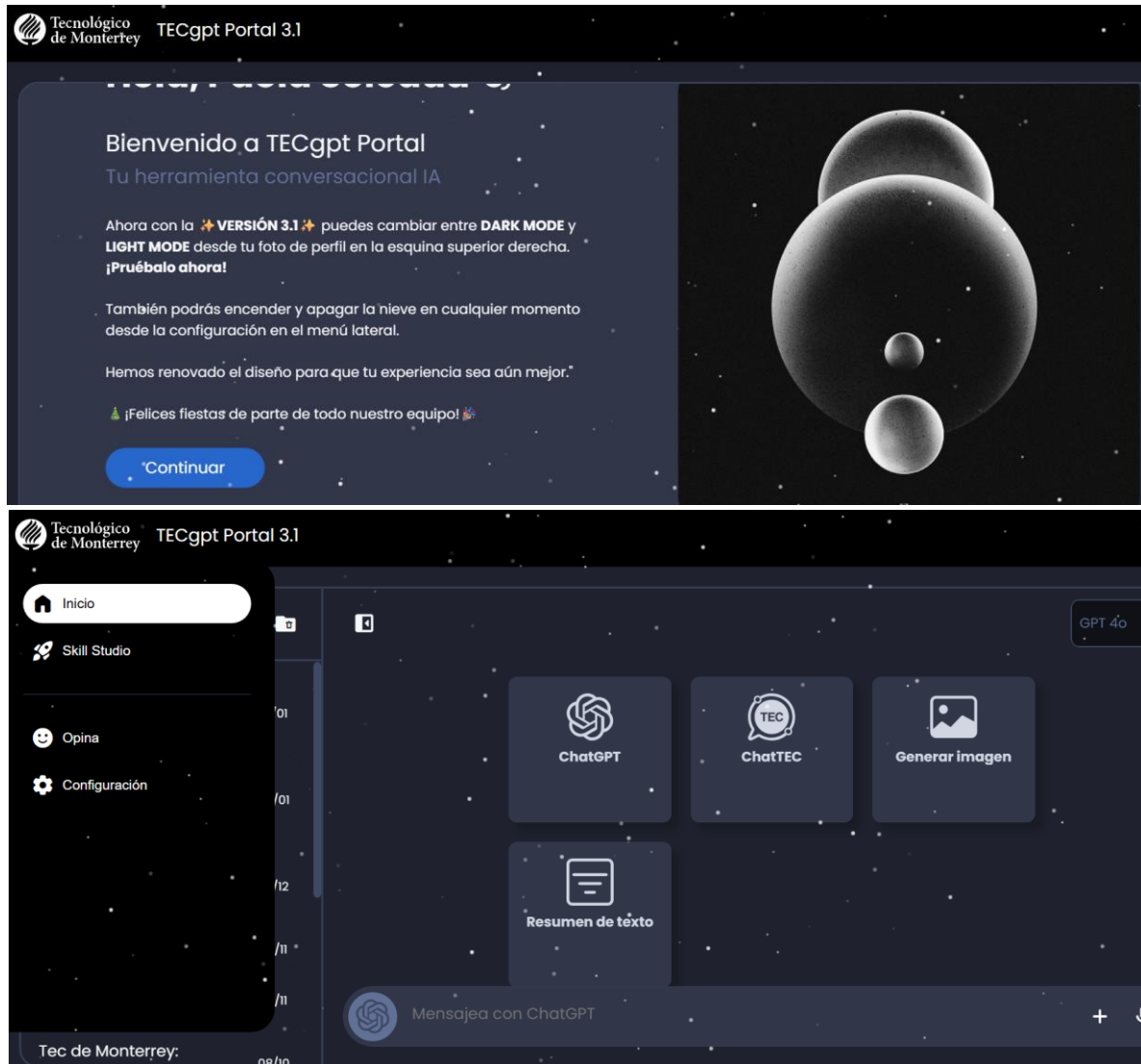
TECgpt Open Edition: [TECgpt Open Edition | Tecnológico de Monterrey](#)

Tecnológico de Monterrey Ethical and pedagogical guidelines for AI use: [Ethical and pedagogical guidelines for AI use](#)

TECgpt presentation video: <https://www.youtube.com/watch?v=Npd0wG8zJwA>

APPENDIX

Image 1. TECgpt portal – Landing page





Joachim Lebovits, CEO
of PyxiScience

Abstract

PyxiScience is an adaptive mathematics learning platform that offers a unique range of features designed to address the challenges currently facing mathematics education, including classroom heterogeneity and instructors workload. Drawing on the expertise of experienced instructors and the ethical use of artificial intelligence, PyxiScience enables differentiated, effective, and dynamic mathematics instruction.

Already used by more than 15,000 users in France and the United States, the platform is deployed in secondary and higher education institutions. Thanks to our numerous deployments, we have been able to gather feedback and analysis on how our users utilize the platform. Students and teachers use the platform in different ways, and it is always interesting to study and understand how they work on PyxiScience.

Introduction

PyxiScience was born out of the experience of Joachim Lebovits, a university Professor (Sorbonne Paris Nord university & New-York University) for over ten years, who faced a universal challenge: helping students at very different levels make progress. Confronted with the limits of traditional approaches, such as manually creating personalized content and dealing with the increasing burden of grading, he realized that these challenges were shared by many professors in both secondary and higher education.

PyxiScience presentation

PyxiScience is an adaptive mathematics platform designed to enhance both teaching and learning. It brings together high-quality educational content, the possibility to create assignments, AI-assisted grading of handwritten assignments, and advanced error analysis to create a comprehensive and efficient learning environment.

The platform provides lessons and exercises developed by experienced professors. Lessons are accessible online and can be enriched with videos and audio content. Exercises are directly linked to the corresponding course sections, allowing students to review concepts while practicing. They are available in multiple formats : multiple-choice, open-ended responses typed on the platform, and static questions answered on paper. Each attempt generates new data, enabling unlimited practice and encouraging true mastery. Students particularly value having access to corrections and being able to practice similar problems multiple times without searching for additional material.

PyxiScience also allows professors to create their own courses and exercises using Python and `MystMarkdown`, a language similar to LaTeX. All created content remains the property of its author, who can choose whether to share it with students, colleagues, or the broader community. In addition, professors can easily design homework, recitations, and exams by selecting exercises from specific chapters or from their own materials, ensuring consistency between teaching and assessment.

A key innovation of the platform is its AI-assisted grading tool for handwritten assignments. Professors can upload scanned exams, and the AI performs text recognition and reasoning analysis before proposing a correction. Professors can then access the correction to make any changes, validate the correction, and send it to their students. This process significantly speeds up grading while providing students with detailed feedback on each question.

Beyond grading, PyxiScience offers powerful error analysis tools. Professors can identify common mistakes across a class, cluster students based on error patterns, and generate personalized exercises. Both instructors and students can monitor progress over time, enabling more frequent, data-driven feedback and continuous improvement throughout the semester.

Where it is used

Today, PyxiScience is widely deployed in both secondary and higher education, reflecting its adaptability to different academic levels and teaching formats. The platform is used in

classrooms as well as for independent work at home, supporting blended learning models and continuous practice beyond scheduled lessons.

PyxiScience now brings together more than 15,000 users across Europe and the United States. In France, it is implemented in several prestigious private high schools and leading higher education institutions such as Sorbonne Paris Nord University, Sciences Po, and CentraleSupélec. In the United States, PyxiScience is currently being rolled out in several major universities.

Professors use of PyxiScience

PyxiScience is designed not to replace professors but to support their teaching. Initially, many professors were cautious about integrating AI into their classrooms. They worried about AI and thought it illegitimate, especially while simultaneously telling their students to be cautious with AI. However, once professors began using PyxiScience, they quickly recognized the platform's real value. It saves time on repetitive and time-consuming tasks, such as grading handwritten assignments, with more than 200,000 papers graded to date, and provides an effective tool to manage the growing heterogeneity of student skill levels within a single class. Professors report up to eight times faster grading compared to traditional methods. Rather than replacing teaching, PyxiScience allows professors to focus on higher-value interactions with students, such as personalized guidance and pedagogical planning.

Professors use PyxiScience in different ways, depending on their teaching style and needs. Some primarily use the platform's exercises to create unique exams and assignments because exercises have data that refresh at each trial. They appreciate the flexibility of multiple formats, including multiple-choice questionnaires, open-ended questions, and static exercises. Others rely mainly on the AI-assisted grading tool to streamline the evaluation of homework, recitations, or exams. For recitations, professors often provide only the detailed correction without a grade, allowing students to use the feedback to improve their understanding. When grading exams, professors maintain full control over the AI's suggested corrections and grades, reviewing and modifying them as needed before releasing results. Across the platform, professors have already created more than 20,000 recitations, exams, and homework assignments, highlighting its broad adoption and utility.

Another major use of PyxiScience is remediation. Many professors value the platform's ability to address differences in student performance by generating exercises tailored to the errors made by individual students. This personalized approach helps bridge learning gaps, ensures that all students progress, and allows instructors to efficiently manage classes with widely varying levels of mastery.

Student use of PyxiScience

With PyxiScience being used in many institutions, students have come to appreciate this platform for several reasons. One of the most popular features is the wide range of exercises available. Instead of spending hours searching for exercises online, students can access exercises directly on the platform, with data that updates with each attempt. This ensures that every attempt is a new opportunity to learn. To date, students have completed more than 4,500,000 practice attempts on the platform.

Students also appreciate the freedom to practice without being judged. Since the exercises are done privately, they can make mistakes and experiment without fear of embarrassment. This safe environment encourages more regular practice and helps build confidence in tackling difficult problems.

Students also particularly appreciate the detailed feedback provided for each answer. Each mistake is explained in depth, allowing students to understand where they went wrong and how to improve. This level of feedback helps them learn more effectively and strengthens their problem-solving skills.

Finally, the personalized exercises assigned by professors are highly appreciated by students. Tailored to each student's level, these exercises allow learners to focus on areas where they need improvement, ensuring that practice is relevant and sufficiently challenging. This personalization helps students progress effectively and makes learning more accessible and motivating.

Analytics

Since PyxiScience began to be used in numerous institutions, we have collected several data points highlighting the platform's usage.

Category	Activity	Data
General data	Number of users	15,000
	Average time spent per student on the platform per week	3 hours
Educational content	Number of exercises available on the platform	4,000+
	Number of practice trials on exercises by students	4,500,000+
	Number of lectures available on the platform	1,500+
AI-Assisted grading tool	Number of recitations, homework and exams created on the platform	3,000+
	Number of handwritten papers graded with PyxiScience	200,000+
	Time divided by professors	By 8

	in grading papers	
	Number of detailed feedback generated	2,000,000+
Analysis of errors and personalized exercises generation	Number of exams analysis generated	6,000+
	Number of personalized exercises generated	150,000+

Conclusion

PyxiScience is used in several institutions in both France and the United States: high schools, preparatory classes, engineering schools, and universities.

Wherever PyxiScience is used, users really appreciate the various features offered by the platform. Instructors can spend less time on tedious and repetitive tasks and more time helping their students. PyxiScience's features offer instructors the opportunity to provide personalized support to each of their students, which was impossible before the introduction of this platform.

While students appreciate the educational content provided and the opportunity to practice the exercises available on the platform as many times as they wish. The platform also enables them to benefit from more personalized support, which is highly and unanimously valued.

Will a chatbot, trained on course-aligned materials and embedded in the learning management system improve students learning outcome?

Associate professor Hilde Skjerve*, Associate professor Martin F. Strand*, Associate professor Anette Sørensen*, Professor Audun Beyer*, Senior advisor Emil Hasle-Hoff*, Associate professor Lars Erik Braaum*, *Kristiania university of Applied Sciences

Abstract

This case study reports on the development and pilot implementation of a course-specific artificial intelligence (AI) chatbot in BIO900 Biochemistry (20 ECTS) at Kristiania University of Applied Sciences. The project was conducted in 2025 with pedagogical development funding from Kristiania University of Applied Sciences. The initiative explored whether a chatbot, trained on course-aligned materials and embedded in the learning management system (Canvas), could help approximate elements of one-to-one academic support for a larger cohort, thereby strengthening learning experiences and potentially improving outcomes.

The chatbot was trained on curated course resources, including transcribed lecture recordings, examination tasks with marking guides, and course texts aligned with learning outcomes. Student assistants and the course coordinator tested and quality-assured the chatbot's responses and its ability to direct students to relevant course resources. Approximately half of the cohort used the chatbot, and users reported high perceived usefulness. Examination results were modestly higher than in the two preceding years, although cohort differences and concurrent use of other AI tools limit causal interpretation.

Keywords: AI, Chatbot, Biochemistry, Assessment, Academic support

Introduction

Since the public release of ChatGPT in late 2022 (OpenAI.com), higher education has faced a dual challenge: concerns about assessment integrity and an accelerating demand to use AI responsibly to enhance learning. At the same time, generative AI chatbots have improved in their ability to provide explanations, support problem-solving, and generate practice activities. These developments have created a realistic opportunity for chatbots to function as virtual teaching assistants, particularly for routine questions and self-testing, while leaving scheduled teaching and human supervision intact.

The present project focused on the “academic assistance” role of an AI chatbot—specifically, whether a course-aligned chatbot could provide students with timely support outside scheduled teaching hours. The guiding question was:

How can the effect of one-to-one instruction be approximated for larger groups through a virtual chatbot, and will this strengthen learning outcomes and the learning experience?

BIO900 Biochemistry was selected for the pilot because key content is relatively stable over time, making it feasible to reuse teaching resources and to build a durable, course-specific knowledge base. The chatbot was framed as a supplement rather than a replacement for teaching sessions. The primary educational objective was to improve accessibility to course-relevant guidance (including rapid responses and alignment with lecture content), while a secondary objective was to reduce repetitive staff workload and generate insight into which topics students find most difficult, based on the questions asked.

Findings

Key problems identified

Implementation planning identified three interrelated problems that commonly reduce learning momentum in content-heavy science courses:

1. Limited access to timely help outside scheduled teaching. Students often study in the evenings or weekends and may become stuck on practice tasks or conceptual misunderstandings. Waiting for the next lecture or office hour can result in lost study time and reduced motivation.
2. Reluctance to ask questions publicly. Some students refrain from raising questions in class, particularly when they assume the question is “too basic” or fear negative evaluation. This can produce silent misconceptions that persist into exam preparation.
3. Difficulty scaling individualized support. Even when teaching staff provide strong support, personalized feedback and repeated explanations are hard to deliver at scale. A supplemental tool that can respond immediately and consistently to common questions may help address this constraint.

Intervention design and implementation

The chatbot was embedded in the learning management system Canvas, requiring student authentication through the learning management system. The chatbot’s knowledge base was constructed from course-specific materials and designed to direct students to relevant resources for deeper study (e.g., lecture videos, slide decks, course documents).

Key components of the training and quality assurance process included:

- Examination tasks and marking guides were uploaded so that the chatbot could assist students who struggled to interpret model answers and assessment expectations.
- Course texts aligned with central learning outcomes were authored by the course coordinator and included to strengthen alignment between chatbot guidance and intended curriculum.
- A separate preparatory set of chemistry resources (upper-secondary level Chemistry 1 and 2) was included to support students lacking prior chemistry background.

Testing was conducted before launch by two former students and the course coordinator, covering straightforward and challenging queries. Quality checks also examined whether

the chatbot correctly pointed students to relevant locations in Canvas for follow-up learning. The chatbot was configured to generate multiple-choice questions and flashcards, which were quality-assured by student assistants before broader use. During pilot use, answers were monitored to detect and correct potential errors.

Student involvement and evaluation activities

Student involvement was planned at multiple stages:

- Consultation with student representatives from the cohort for information-sharing and input.
- An interim evaluation during implementation to guide iterative refinement.
- A final evaluation after course completion, including plans for a focus group interview to capture student experiences in depth.
- Complementary indicators such as usage data (ex. question volume) and examination outcomes.

Adoption and use patterns

Adoption during the pilot was moderate. Twenty-two of forty-two students used the chatbot (approximately 52%), and users submitted 632 questions in total. In parallel, 95% of students reported using other AI tools, indicating that the chatbot was introduced into an already AI-saturated study environment. Among students who used the course chatbot, 91% reported that it was useful. Regarding help-seeking preferences, 82% indicated that they would still ask questions during teaching sessions, while 18% reported preferring the chatbot over asking questions in class.

Student questions primarily reflected core learning needs and assessment preparation:

- definitions and key concepts;
- clearer explanations of textbook topics;
- checking solutions to practice tasks;
- simplified explanations of biochemical reactions and metabolic processes;
- exam-related clarifications and practice using quizzes and flashcards.

Examination outcomes

Assessment in BIO900 consisted of a multi-part semester assignment (including multiple-choice items and extended-response tasks) and an oral examination, with grades on a National Scale from A-F translated to 1–6 scale (6 highest). The course reported that content and examination format were identical to 2024 and 2023, allowing cautious year-to-year comparison.

The average examination grade in 2025 was 3.84, compared with 3.67 in 2024 and 3.33 in 2023. No students failed in 2025, matching the outcome in 2024, while 2023 had a 3% failure rate. Non-attendance at the examination was 16% in 2025 (18% in 2024; 15% in 2023).

At the same time, cohort performance in other related courses suggested potential cohort effects. The same cohort reportedly performed weaker than earlier cohorts in BIO400 (–0.26 grade points) and BIO300 (–0.69 grade points), with higher failure rates (13%–27%). This indicates that the modest improvement in BIO900 cannot be attributed to chatbot use alone without stronger controls.

Year started on program	Result	Failure rate	Nonattendance
2023 (intervention)	3,84	0%	16%
2022	3,67	0%	18%
2021	3,33	3%	15%

Intervention class exam results BIO400 on prior semester

Year started on program	Result	Failure rate	Nonattendance
-------------------------	--------	--------------	---------------

2023 (intervention)	2,12	29,09%	7,27%
2022	2,22	16,13%	3,23%
2021	2,38	21,74%	8,70%

Intervention class exam results BIO300 on prior semester

Year started on program	Result	Failure rate	Nonattendance
2023 (intervention)	1,56	37,74%	42,03%
2022	2,25	10,14%	15,09%

Discussion

Summary of the major problem(s) and interpretation

The case suggests that a course-specific chatbot can meaningfully address the problems of limited access to timely help, reluctance to ask questions in public settings, and challenges in scaling individualized academic support.

First, availability and immediacy appear to be the most direct benefit. Students were able to request explanations at the point of need rather than waiting for scheduled sessions. In a cumulative subject such as biochemistry, timely clarification may prevent misunderstandings from compounding and may support more productive independent study. The question volume among users supports the interpretation that the chatbot functioned as “just-in-time” academic assistance rather than a purely exploratory tool.

Second, barriers to help-seeking may be reduced when students can ask questions privately. A chatbot can provide repeated explanations without social cost, potentially supporting students who hesitate to speak up in lectures. Importantly, the findings also suggest complementarity rather than replacement: most students reported that they would still ask questions in class, while a smaller group preferred the chatbot. This pattern aligns with a model in which chatbot use supports confidence-building and preparation for higher-quality classroom engagement.

Third, scalability and instructional development represent additional potential value. A course-aligned chatbot can respond to frequent foundational queries and direct students to relevant resources, potentially reducing repetitive staff burden. Moreover, logs of student questions can provide instructors with systematic insight into which learning outcomes generate confusion, enabling more targeted adjustments to teaching and assessment preparation.

Practical implications

Several implementation considerations emerge from the pilot:

- Curation and quality assurance are essential, especially when the chatbot is expected to explain complex processes and to reference course resources accurately. Corrected transcripts and verified exam materials likely improved alignment and trust.
- Student assistants played a crucial role in preparing data and monitoring output quality, indicating that institutional resourcing is necessary for responsible deployment.
- Clear positioning and guidance are needed in an environment where most students already use general AI tools. A course-specific chatbot should be presented as uniquely valuable due to alignment with course content, learning outcomes, and assessment expectations.

- Inclusivity can be strengthened by incorporating foundational resources (e.g., chemistry prerequisites) for students with limited prior knowledge.

Limitations

This case study has notable limitations. Chatbot use was voluntary and self-selected, which may bias outcomes because users may differ from non-users in motivation or study strategies. Nearly all students reported using other AI tools, introducing confounding influences. Cross-year comparisons, while informative, cannot control for cohort differences in preparedness, language proficiency, prior achievement, or study context. Stronger evaluation designs are required to determine whether the chatbot independently contributes to improved learning outcomes and for which subgroups the benefit is greatest.

Conclusion

This 2025 case study at Kristiania University of Applied Sciences demonstrates the feasibility of deploying a Canvas-integrated, course-specific AI chatbot as a pedagogical development initiative. Approximately half of the cohort adopted the chatbot, and most users reported that it was useful for conceptual clarification, practice-task support, and exam preparation through quizzes and flashcards. Examination results were modestly higher than in 2023 and 2024, although cohort effects and concurrent use of other AI tools limit causal conclusions. Overall, the case indicates that a curriculum-aligned chatbot can extend access to timely academic assistance, reduce barriers to help-seeking, and generate actionable insight into difficult topics. Future development should prioritize rigorous evaluation (e.g., structured group comparisons and pre-post measures), systematic analysis of question logs mapped to learning outcomes, and continued quality assurance to maintain accuracy and pedagogical alignment.

AI-Gen-Case: A Transformative AI-Driven Framework for Case-Based Learning in Higher Education

Diego Ravenda, TBS
Education

Abstract

Traditional case-based teaching remains a cornerstone of management and business education, yet it increasingly struggles to reflect contemporary decision-making environments characterized by complexity, uncertainty, stakeholder interaction, and the growing presence of artificial intelligence. Static cases, limited personalization, uneven participation, and opaque assessment practices are further challenged by the rapid diffusion of Generative AI tools capable of producing high-quality case analyses, calling into question the pedagogical value of conventional approaches.

This paper introduces AI-Gen-Case, an AI-driven framework that leverages Generative AI to design adaptive, interactive, and context-rich case studies in real time. Through structured prompting, dynamic dialogue, and scenario branching, AI-Gen-Case enables students to engage with evolving problems, interact with simulated stakeholders, and make decisions in AI-augmented environments that require human judgment, explicit trade-offs, and critical thinking. The framework integrates governance safeguards, including supervised use, oral validation, interaction traceability, and comparison across multiple AI systems, to ensure responsible and ethical implementation.

The paper outlines the conceptual foundations, methodology, and pedagogical contributions of AI-Gen-Case, highlighting its benefits for students, faculty, and institutions. AI-Gen-Case enhances inclusivity and motivation through personalization, strengthens the link between theory and practice, improves assessment transparency, and develops future-ready skills essential for professional contexts increasingly shaped by AI. The framework and demonstration show how Generative AI can be integrated into higher education not as a shortcut to answers, but as a catalyst for deeper learning and more robust decision-making.

Keywords: Generative artificial intelligence; Case-based learning; Human judgment; AI-augmented decision making

Introduction

Case-based teaching has long been a foundational pedagogical approach in management and business education, valued for its capacity to connect theory with practice and expose students to realistic decision-making contexts (Christensen, Garvin, & Sweet, 1991; Kolb, 1984). Despite its enduring relevance, traditional case pedagogy increasingly reveals structural limitations that undermine its effectiveness in contemporary higher-education environments. Students often engage with static and outdated scenarios that fail to reflect the complexity, uncertainty, and rapid evolution of modern organizations. Classroom participation is uneven, feedback is limited, and assessment practices frequently rely on subjective judgments that are difficult to document systematically (Ellet, 2007; Biggs & Tang, 2011).

At the same time, the rapid diffusion of Generative Artificial Intelligence (Gen-AI) tools capable of producing sophisticated case analyses within seconds raises fundamental questions about the validity of conventional case-based assignments and evaluation methods. When high-quality answers can be generated instantly, the pedagogical focus must shift from solution production toward reasoning, judgment, and critical engagement (Cotton, Cotton, & Shipway, 2023).

The AI-Gen-Case framework responds to these challenges by leveraging Generative AI to generate, personalize, and continuously adapt case studies in real time. Through structured prompting, interactive dialogue, and AI-enabled applications, AI-Gen-Case creates a dynamic learning environment that adjusts to individual students' knowledge levels, preferences, and analytical behaviours. In doing so, it addresses the core shortcomings of traditional case pedagogy while aligning with international calls for responsible, transparent, and human-centred integration of AI in education (UNESCO, 2021, 2023).

Limitations of Traditional Case Study Teaching

The limitations of traditional case-based teaching are well documented. First, cases are typically anchored to specific historical and organizational contexts, which limits their transferability and rapidly diminishes their relevance in environments characterized by technological disruption and regulatory change (Ellet, 2007). Updating case materials is costly and time-consuming, and many cases struggle to keep pace with real-world developments.

Second, participation inequality remains a persistent concern. Classroom discussions often privilege extroverted, linguistically confident, or culturally dominant students, while introverted learners or international students may find it difficult to engage meaningfully (Brookfield & Preskill, 2016). As a result, instructors gain limited insight into the reasoning processes of a substantial share of the cohort.

Third, traditional case discussions leave little traceable evidence of student thinking. Feedback is ephemeral, and instructors lack detailed records of how students interpret information, revise assumptions, or justify decisions. This constrains assessment transparency and limits the reliability of learning evaluation (Biggs & Tang, 2011).

Ethical limitations further compound these issues. Many business cases emphasize financial performance while marginalizing ethical, social, and sustainability considerations, offering a partial and sometimes misleading representation of managerial decision-making (Crane et al., 2019). More fundamentally, traditional cases are largely static and fail to reflect

contemporary decision environments in which managers must interact continuously with multiple stakeholders, evaluate complex and evolving situations, and demonstrate relational, communicative, and ethical skills in real time. They also insufficiently prepare students to operate alongside AI systems that require human judgment, explicit trade-offs, and critical thinking rather than mechanical optimization. Finally, the growing use of Gen-AI tools challenges academic integrity and raises concerns about the authenticity of student work (Cotton et al., 2023).

AI-Gen-Case: Concept and Methodology

AI-Gen-Case uses Generative AI to produce adaptive, personalized, and pedagogically aligned case studies. At the core of the framework lies a structured prompting methodology designed to ensure coherence, transparency, and instructional relevance (Liu et al., 2023).

The process begins with an assessment of student knowledge, enabling the AI system to calibrate case complexity and avoid cognitive overload, in line with cognitive load theory (Sweller, Ayres, & Kalyuga, 2011). The system then elicits student preferences regarding industries, ethical dilemmas, or stakeholder perspectives, increasing perceived relevance and intrinsic motivation (Ryan & Deci, 2000).

Based on these inputs, the AI generates a realistic, context-rich scenario that mirrors authentic professional challenges. Students navigate the case through iterative analytical questions and interactive dialogue, allowing them to request clarifications, test alternative strategies, and receive immediate feedback. Advanced implementations incorporate stakeholder simulations and voice-based role-play, reproducing the ambiguity and complexity of real organizational decision environments more effectively than static written cases.

Interactive Learning Through AI-Generated Exercises

AI-Gen-Case transforms case-based learning from a passive, consumption-oriented activity into an active, dialogic process. Students do not simply analyze a predefined scenario; they co-construct the case dynamically as new information emerges. This reflects the iterative nature of real-world decision-making, characterized by uncertainty, feedback loops, and evolving constraints (Mintzberg, 2009).

A central advantage of the framework is its capacity for dynamic adaptation. When students demonstrate strong analytical skills, the AI can introduce additional constraints, conflicting stakeholder objectives, or ethical trade-offs. When students struggle, the system can scaffold learning by providing conceptual explanations or simplifying the scenario, consistent with socio-constructivist learning principles (Vygotsky, 1978; VanLehn, 2011).

Scenario branching enhances engagement and ownership. Different choices lead to different consequences, enabling students to experience how strategic decisions influence financial performance, employee outcomes, stakeholder trust, and long-term sustainability.

Post-Exercise Activities and Learning Consolidation

Learning consolidation is supported through structured post-exercise activities emphasizing reflection, validation, and critical evaluation. Oral presentations require students to explain their scenario, decisions, and underlying reasoning, reinforcing accountability and reducing overreliance on AI-generated outputs (Biggs & Tang, 2011). Because each student encounters a unique case, these presentations also stimulate rich peer learning and comparative discussion.

A key component involves critical evaluation of AI outputs. Students are required to identify potential hallucinations, biases, or logical inconsistencies, cultivating responsible AI use and epistemic vigilance (Floridi et al., 2018). Students also refine or redesign prompts to achieve specific learning objectives, developing prompt-engineering competencies that are increasingly relevant in professional contexts (Zamfirescu-Pereira et al., 2023).

All AI-student interactions can be saved and reviewed, allowing instructors to analyze reasoning patterns, misconceptions, and engagement levels. This traceability enhances assessment transparency and supports more informed and equitable evaluation.

Responsible Implementation and Risk Mitigation

The integration of Gen-AI into pedagogy requires careful governance. AI-Gen-Case incorporates several safeguards to ensure responsible and ethical implementation. Exercises are ideally conducted under supervision, reducing the risk that students simply use AI tools to generate answers outside the intended learning process. Oral presentations and structured class discussions further validate that students have internalized the material and can articulate their reasoning, rather than outsourcing judgment to AI systems. In addition, systematic student feedback supports the continuous refinement and improvement of the methodology (UNESCO, 2021, 2023).

An additional safeguard involves exposing students to multiple AI systems and comparing their outputs. Evaluating tools such as ChatGPT, Gemini, Grok, and Qwen allows students to observe differences in reasoning, biases, and limitations across models. This comparative approach discourages overreliance on a single platform and reinforces a more critical and informed understanding of AI ecosystems, while explicitly positioning AI as a cognitive partner that requires human judgment, trade-offs, and accountability rather than as a substitute for reasoning.

Benefits and Pedagogical Contribution

AI-Gen-Case delivers substantial pedagogical benefits that directly address the shortcomings of traditional case teaching. One of the most transformative advantages is personalization. Cases adapt to each student's background, interests, and level of understanding, increasing motivation and reducing inequities related to personality traits or cultural differences (Ryan & Deci, 2000). Students who typically remain silent in large-group discussions can engage fully through one-to-one AI interactions, ensuring more inclusive participation.

The framework also strengthens the connection between theory and practice. Gen-AI can provide just-in-time theoretical explanations, definitions, and conceptual links, ensuring that students understand the foundational principles underlying their decisions rather than relying solely on intuition or pattern matching. This supports deeper learning and more explicit articulation of reasoning.

Real-time relevance represents another major benefit. AI-generated scenarios can incorporate current economic trends, regulatory environments, and emerging technologies, eliminating the problem of outdated case materials and exposing students to the types of uncertainty and complexity faced by contemporary decision makers.

Importantly, AI-Gen-Case enhances assessment transparency. Recorded dialogues and interaction logs provide concrete evidence of student reasoning processes, enabling more consistent, objective, and defensible evaluation. This traceability constitutes a major advancement over traditional case discussions, which offer limited documentation for assessment.

Finally, AI-Gen-Case develops future-ready skills. Students practice critical thinking, digital literacy, scenario analysis, ethical reasoning, and prompt engineering, while learning to interact productively with AI systems that require human judgment, explicit trade-offs, and critical evaluation. These competencies are increasingly indispensable across professional domains, positioning the framework not only as a pedagogical innovation but also as a workforce-development tool.

Implications for Students, Faculty, and Institutions

The AI-Gen-Case framework creates value for all stakeholders within the educational ecosystem.

For students, AI-Gen-Case offers a more engaging, flexible, and relevant learning experience. Each student works with a unique case, follows a personalized learning path, and receives immediate, context-specific feedback. This personalization enhances motivation, supports deeper conceptual understanding, and allows students to actively develop judgment, critical thinking, and decision-making skills in complex, AI-augmented environments.

Faculty benefit from reduced preparation time and increased consistency in case quality. Rather than focusing on the repeated crafting or updating of static cases, instructors can devote more energy to facilitation, feedback, and the higher-order dimensions of teaching, including guiding reflection, ethical reasoning, and stakeholder-oriented analysis.

At the institutional level, AI-Gen-Case provides a scalable and technologically advanced pedagogical model aligned with accreditation expectations related to transparency, innovation, and responsible digital integration. Implementing AI-Gen-Case signals strategic leadership in a higher-education sector undergoing rapid transformation, while demonstrating a commitment to preparing graduates for decision-making roles in AI-augmented professional contexts.

Conclusion

AI-Gen-Case represents a paradigm shift in case-based teaching by reconceptualizing what it means to learn from cases in an AI-augmented world. By leveraging Generative AI to produce adaptive, context-rich, and interactive scenarios, the framework overcomes the structural limitations of traditional, static case studies and replaces them with learning environments that more closely resemble contemporary decision-making realities. Students are not merely asked to analyse situations *ex post*, but to engage dynamically with unfolding problems, multiple stakeholders, and AI systems that require human judgment, explicit trade-offs, and critical thinking.

Beyond technological novelty, the contribution of AI-Gen-Case lies in its pedagogical reorientation. Its structured prompting methodology ensures coherence and alignment with learning objectives, while its interactive design promotes deeper reflection, transparency of reasoning, and more equitable participation. At the same time, the framework embeds safeguards that support responsible and ethical AI use, reinforcing the central role of human oversight and accountability in AI-mediated learning.

Aligned with the objectives of the AI in Higher Education Summit 2026, AI-Gen-Case illustrates how Generative AI can be integrated into higher education not as a shortcut to answers, but as a catalyst for richer learning, stronger judgment, and future-ready skills. In this sense, AI-Gen-Case is not simply an enhancement of case-based pedagogy, but a substantive redefinition of how case learning can prepare students, faculty, and institutions for an increasingly AI-augmented professional landscape.

References

- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.
- Brookfield, S. D., & Preskill, S. (2016). *The discussion book* (2nd ed.). Jossey-Bass.
- Christensen, C. R., Garvin, D. A., & Sweet, A. (1991). *Education for judgment*. Harvard Business School Press.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating? Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 60(4), 1–12. <https://doi.org/10.1080/14703297.2023.2190148>
- Crane, A., Matten, D., Glozer, S., & Spence, L. (2019). *Business ethics: Managing corporate citizenship and sustainability in the age of globalization* (5th ed.). Oxford University Press.
- Ellet, W. (2007). *The case study handbook*. Harvard Business School Press.
- Floridi, L., Cows, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707.
- Kolb, D. A. (1984). *Experiential learning*. Prentice Hall.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1–35. <https://doi.org/10.1145/3560815>
- Mintzberg, H. (2009). *Managing*. Berrett-Koehler.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.
- UNESCO. (2021). *AI and education: Guidance for policy-makers*. UNESCO Publishing.
- UNESCO. (2023). *Guidance on generative AI in education and research*. UNESCO Publishing.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.

Vygotsky, L. S. (1978). *Mind in society*. Harvard University Press.

Zamfirescu-Pereira, J., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and fail) to design prompts for large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–21). ACM. <https://doi.org/10.1145/3544548.3581388>.

AI Situations: A Practice-Intensive and Modular Framework for AI Literacy, Curriculum and Assessment in Higher Education

Marc Gonnet, Arnaud Gorgeon
Clermont School of Business, 4 Bd Trudaine,
63000 Clermont-Ferrand, France
marc.gonnet@clermont-sb.fr;
arnaud.gorgeon@clermont-sb.fr

Abstract

This submission presents AI Situations, a practice-intensive and modular framework for AI literacy in higher education. It shifts the focus from tools to transversal situations of working with generative and agentic AI. Competencies are developed and assessed through short, practice-based sprints, supporting curriculum design, academic integrity, and institutional coherence.

Keywords: Generative AI, Agentic AI, Competencies, Practice

Context and problem statement

Generative and agentic AI are transforming learning, teaching, research, and academic work at a pace rarely experienced in higher education (Robert et al., 2025). Institutions face increasing expectations to integrate these technologies responsibly while preserving academic standards. In response, many initiatives focus on specific tools, short training sessions, or experimental projects led by motivated individuals.

While such initiatives demonstrate agility, they also reveal structural limits. Tool-centric approaches age quickly and provide limited guidance for assessment (Burneo-Arteaga et al., 2025). More fundamentally, institutions often lack a definition of what it means to be competent when working with AI in academic and professional contexts (Yan et al., 2025).

In this contribution, we draw from the latest research on generative artificial intelligence literacy (Annapureddy et al., 2025; Chee et al., 2025; Chiu et al., 2024; Mutawa, 2025) and propose a competency-based and situation-driven framework for AI literacy, grounded in extensive structured practice and organized through a modular, sprint-based design. This framework is designed to remain robust despite technological change.

Competency framework

Generative artificial intelligence refers to models capable of producing new content (texts, images, data, reasoning) based on human instructions. Agentic AI encompasses systems that are capable not only of generating output, but also of acting, planning, executing tasks,

and interacting within a process, under human supervision. Digital agents carry out planned actions by relying on tools and data. Autonomous agents possess a broader capacity for initiative in task execution, within an ethical framework defined by humans.

These forms of AI are transforming pedagogical and administrative activities as well as professions, research, and managerial practices. They therefore call for a shared and structured framework for the entire educational community.

Our competency framework defines the skills required to understand, experiment with, and govern the uses of AI. It is structured around three families of competencies, reflecting three structuring logics:

- Thinking with AI: Understanding models, reasoning with discernment, and exercising critical vigilance.
- Acting with and through AI: Experimenting with, validating, and designing useful solutions, while ensuring human oversight.
- Guiding and transforming: Supporting governance, managing change, and building trust in the collective use of AI.

Building competencies through transversal AI situations

AI situations represent stable categories of activity across disciplines and institutional functions. Examples include analyzing a problem or a dataset with AI support; producing texts, analyses, or designs collaboratively with AI; evaluating relevance, reliability, bias, and limits of AI outputs; supervising AI-assisted work within a workflow; and redesigning pedagogical or administrative processes rather than merely automating existing ones.

Because these situations are tool-agnostic, they provide a durable foundation for curriculum design, assessment, and organizational learning.

Practice is central to the framework (Salinas-Navarro et al. 2024). Learning is organized into short, modular sprints, each centered on one transversal AI situation. A sprint combines guided activities, autonomous experimentation, and reflective analysis. It leads to an explicit output that can be discussed, assessed, and improved.

This design ensures that competencies are not inferred but demonstrated. It also supports flexible implementation across courses, programs, and professional development contexts while maintaining coherence.

Levels of autonomy and responsibility

Each situation and competency can be addressed at three progressive levels. The first level corresponds to guided use with explicit instructions and supervision. The second level involves autonomous and reflective use, where individuals independently choose, justify, and evaluate AI usage. The third level corresponds to strategic and supervisory use, including system design, process governance, and responsibility for others' AI-assisted work.

This progression supports curriculum scaffolding, assessment alignment, and professional development.

Target audiences

The framework is designed for students at undergraduate and graduate levels; faculty involved in course design, assessment, and supervision; administrative and support staff using AI-assisted processes; and academic leaders and program directors seeking coherence across institutional initiatives

Contributions and implications

This contribution offers a future-proof approach to AI literacy grounded in practice and modularity rather than tools. It provides a structured response to assessment and academic integrity challenges, a shared language across disciplines and functions, and a practical foundation for aligning learning objectives, activities, and evaluation in AI-rich environments.

References

- Annapureddy, R., Fornaroli, A., & Gatica-Perez, D. (2025). Generative AI Literacy: Twelve Defining Competencies. *Digital Government: Research and Practice*, 6(1), 1–21. <https://doi.org/10.1145/3685680>
- Burneo-Arteaga, P., Lira, Y., Murzi, H., Balula, A., & Costa, A. P. (2025). Capability-based training framework for generative AI in higher education. *Frontiers in Education*, 10, Article 1594199. <https://doi.org/10.3389/educ.2025.1594199>
- Chee, H., Ahn, S., & Lee, J. (2025). A competency framework for AI literacy: Variations by different learner groups and an implied learning pathway. *British Journal of Educational Technology*, 56(5), 2146–2182. <https://doi.org/10.1111/bjet.13556>
- Chiu, T. K., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers & Education Open*, 6, 100171. <https://doi.org/10.1016/j.caeo.2024.100171>
- Mutawa, A. M., & Sruthi, S. (2025). UNESCO's AI Competency Framework: Challenges and Opportunities in Educational Settings. In *Impacts of Generative AI on the Future of Research and Education* (pp. 75–96). IGI Global. <https://doi.org/10.4018/979-8-3693-0884-4.ch004>
- Robert, J., Muscanell, N., McCormack, M., Pelletier, K., Arnold, K., Arbino, N., ... Reeves, J. (2025). *2025 EDUCAUSE Horizon Report: Teaching and Learning Edition*. EDUCAUSE. <https://library.educause.edu/resources/2025/5/2025-educause-horizon-report-teaching-and-learning-edition>
- Salinas-Navarro, D. E., Vilalta-Perdomo, E., Michel-Villarreal, R., & Montesinos, L. (2024). Designing experiential learning activities with generative artificial intelligence tools for authentic assessment. *Interactive Technology and Smart Education*, 21(4), 708–734. <https://doi.org/10.1108/ITSE-12-2023-0236>
- Yan, W., Nakajima, T., & Sawada, R. (2025). Beyond tool use: Tracking the evolution of generative AI literacy among university students through a process-oriented investigation. *Computers & Education: Artificial Intelligence*, 100465. <https://doi.org/10.1016/j.caeai.2025.100465>

[From AI literacy to quiteracy: Unplugging the classroom to learn about and not with AI

Heid Reed, Audencia
Business School

Abstract

AI literacy is an urgent imperative; however, the management and critical education technology literatures increasingly warn us that literacy initiatives must go beyond tool competence. It is not only important to understand how to use AI but also to understand how it works and the ethical concerns it raises. In line with these concerns, this work in progress draws on a pilot study conducted during a faculty development day in which I hosted a digital device free “AI Ethics Unplugged” workshop. Utilizing tactile activities and mechanical computing, participants engaged in hands-on stations designed to demystify technical terms like algorithm and link them to ethical concerns like algorithmic bias. The study then explores how an unplugged pedagogical approach to AI literacy can cultivate critical thinking and ethical engagement while fostering digital agency, defined here through the lenses of media “quiteracy” and digital sobriety.

Keywords: AI ethics, AI literacy, AI quiteracy, digital sobriety, unplugged learning

Background Literature

There is growing debate in the management education literature on the role of GenAI in the classroom (Krammer, 2025; Prasad & Sliwa, 2025) with research needed to explore how we can develop social critical thinking around the societal issues raised by this now ubiquitous technology (Larson et al., 2024). One area of particular urgency is how to incorporate AI literacy in education curriculum (Ng et al., 2021; UNESCO, 2024). The critical education technology (edtech) literature warns us that technology literacies (i.e. media, digital, data, etc.) have a long history of bending towards “tool competence”, not only necessitating expensive investment in hardware and software that benefits businesses more than schools (Postman, 2011; Watters, 2023) but also overshadows critical engagement and reflection about the ethics behind them (Hobbs & Jensen, 2009). Critical perspectives towards edtech recognizes “Technologies are not just ‘tools’ but as sociotechnical entities that reflect and reinforce human values due to their entanglements with social, cultural, political, and economic systems” (Pleasant et al., 2024, p. 2).

Noting the distinction between teaching about versus teach with technologies, Krutka et al. (2019) calls on educators to “disrupt pervasive and normalizing worldviews that prioritize progress, power” and “seek justice-oriented pedagogies” (p. 568). Answering this call, I draw on the concepts of media quiteracy and digital sobriety to demonstrate how unplugged learning activities grounded in relational and embodied learning (Macrine & Fugate, 2022) can be a way to not only develop AI competences (UNESCO, 2024) but also foster digital agency to resist “technological imperialism” in higher education (Nichols et al., 2025).

Good and Ciccone (2025) use the term media quiteracy not as “a passive disengagement from digital culture” but “as an active, generative approach to learning about and intervening in a heavily mediatized, commercialized, and digitized information environment through intentional acts of refusal, abstention, and disconnection” (p. 151). They argue for the incorporation of non-use “as a viable option and a legitimate enactment of media and digital literacy skills” (p. 153) in order to:

...empower students to not only flex their agency against machines and explore the social and ethical implications of their media use and disuse, but also to question the cultural, institutional, and economic assumptions that have led ‘participation’ in [digital] platforms to be treated as a precondition for functioning in our society. (p. 152)

Quiteracy can additionally be seen as a way to reduce harms to students linked to the use of digital devices and GenAI in the classroom (Burns et al., 2026; Deng et al., 2022). Noting the environmental impacts of digital technologies, it is also connected to the concept of digital sobriety which encourages more sober digital services and moderate daily digital use (European Commission, 2021) or refutation (Pérea et al., 2023).

While plugged activities “employ technologies, such as digital and mobile applications and block-based programming, for hands-on engagement and problem-solving exercises,” unplugged activities “incorporate minimal to no [digital] technologies, such as puzzles and physical games, and usually focus on AI concepts and thinking skills” (Dai, 2025, p. 6743). UNESCO’s AI competences framework (2024) identifies unplugged learning settings as appropriate for all progressive AI literacy levels, i.e. to understand, apply, and create.

Pilot Test

How can an unplugged pedagogical approach cultivate critical thinking and ethical engagement with AI while fostering digital agency? To explore this question, a pilot study was conducted during a mandatory faculty development day centered on AI in research and teaching. Of the ten micro-workshops offered, nine were instrumental, focusing on *how* to use AI such as for data extraction or chatbot creation. In contrast, the tenth workshop, “AI Ethics Unplugged,” was strictly device-free, prioritizing learning *about* AI. The ten workshops were run in parallel, and faculty could choose which to attend. A total of 16 professors and 1 pedagogical designer participated in the unplugged workshop.

The goal of the unplugged workshop was to understand AI concepts in line with critiques in the literature that technology literacy is overly focused on understanding functions rather than the mechanics of how they work (Knauss, 2020). Utilizing “TACTivities” (Hodge et al., 2020) and mechanical computing (Gardner, 1962; Pitt, 2023), participants rotated through hands-on stations designed to demystify technical AI related terms while linking them to ethical concerns such as algorithmic bias (Broussard, 2024), the alignment problem (Benjamin, 2024), and the underwaged, racialized, gendered realities of invisible digital labor (D’Ignazio & Klein, 2020).

A post-workshop qualitative survey was sent to both participants and non-participants to ask respectively about their experience and motivations for choosing or not choosing the workshop. Analysis is ongoing but initial findings reveal that participants overall felt that they increased their understanding of technical terms associated with AI and their awareness of ethical issues while appreciating a “slower” and more “reflective” pace that allowed for dialoguing with colleagues. One participant described it as a place of “discovery and thinking space” while many used the word “fun” or “engaging” for the activities. Many participants appreciated the critical approach of the workshop and contrasted it with external discourses framing AI in a “positive only” way. Non-participants expressed that they were more interested in learning how to use AI with some expressing interest in the Ethics Unplugged workshop, noting that they wished it had been offered in addition to rather than in place of the others. We have plans to run the workshop again so that more colleagues can participate.

References

- Broussard, M. (2024). *More than a glitch: Confronting race, gender, and ability bias in tech*. MIT Press.
- Burns, M., Winthrop, Rebecca, Luther, Natasha, Venetis, Emma, & Rida, Karim. (2026). A new direction for students in an AI world: Prosper, prepare, protect (pp. 1–219). Brookings Institute. <https://www.brookings.edu/wp-content/uploads/2026/01/A-New-Direction-for-Students-in-an-AI-World-FULL-REPORT.pdf>
- Dai, Y. (2025). Integrating unplugged and plugged activities for holistic AI education: An embodied constructionist pedagogical approach. *Education and Information Technologies*, 30(5), 6741–6764. <https://doi.org/10.1007/s10639-024-13043-w>
- Deng, L., Liu, Y., Ku, K. Y. L., & Lin, L. (2022). In-class multitasking with smartphones and laptops: Exploring student experiences and perceptions. *College Teaching*, 70(4), 443–451. <https://doi.org/10.1080/87567555.2021.1973947>
- D’Ignazio, C., & Klein, L. F. (2020). *Data feminism*. The MIT Press.

European Commission. (2021, April 8). Digital sobriety: How can we adapt our uses for a positive impact on the environment? Digital Sobriety. <https://epale.ec.europa.eu/en/blog/digital-sobriety-how-can-we-adapt-our-uses-positive-impact-environment>

Good, K. D., & Ciccone, M. (2025). Media quiteracy: Why digital disconnection belongs in the media literacy curriculum. *Journal of Media Literacy Education*, 17(1), 150–165. <https://doi.org/10.23860/JMLE-2025-17-1-10>

Hobbs, R., & Jensen, A. (2009). The past, present, and future of media literacy education. *Journal of Media Literacy Education*. <https://doi.org/10.23860/jmle-1-1-1>

Hodge, A., Wanek, K., & Rech, J. (2020). TACTivities: A tactile way to learn interdisciplinary communication skills. *PRIMUS*, 30(2), 160–171. <https://doi.org/10.1080/10511970.2018.1532937>

Knaus, T. (2020). Technology criticism and data literacy: The case for an augmented understanding of media literacy. *Journal of Media Literacy Education*, 12(3), 6–16. <https://doi.org/10.23860/JMLE-2020-12-3-2>

Krammer, S. M. (2025). Is there a glitch in the matrix? Artificial intelligence and management education. *Management Learning*, 56(2), 367–388. (184190445). <https://doi.org/10.1177/13505076231217667>

Krutka, D. G., Heath, M. K., & Willet, K. B. S. (2019). Foregrounding technoethics: Toward critical perspectives in technology and teacher education. *Journal of Technology and Teacher Education*, 27(4), 555–574. <https://www.learntechlib.org/primary/p/208235/>

Larson, B. Z., Moser, C., Caza, A., Muehlfeld, K., & Colombo, L. A. (2024). Critical thinking in the age of Generative AI. *Academy of Management Learning & Education*, 23(3), 373–378. <https://doi.org/10.5465/amle.2024.0338>

Macrine, S. L., & Fugate, J. M. B. (Eds). (2022). *Movement matters: How embodied cognition informs teaching and learning*. The MIT Press. <https://doi.org/10.7551/mitpress/13593.001.0001>

Marx, E., Leonhardt, T., Baberowski, D., & Bergner, N. (2022a). Using Matchboxes to Teach the Basics of Machine Learning: An Analysis of (Possible) Misconceptions. *The Second Teaching Machine Learning and Artificial Intelligence Workshop*, 25–29. <https://proceedings.mlr.press/v170/marx22a.html>

Marx, E., Leonhardt, T., Baberowski, D., & Bergner, N. (2022b). Using matchboxes to teach the basics of machine learning: An analysis of (possible) misconceptions. *Proceedings of the Second Teaching Machine Learning and Artificial Intelligence Workshop*, 25–29. <https://proceedings.mlr.press/v170/marx22a.html>

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/j.caeai.2021.100041>

Nichols, T. P., Logan, C., & Garcia, A. (2025). Generative AI and the (Re)turn to Luddism. *Learning, Media and Technology*, 1–14. <https://doi.org/10.1080/17439884.2025.2452199>

Péréa, C., Gérard, J., & de Benedittis, J. (2023). Digital sobriety: From awareness of the negative impacts of IT usages to degrowth technology at work. *Technological Forecasting and Social Change*, 194, 122670. <https://doi.org/10.1016/j.techfore.2023.122670>

Pitt, L. (2023). Turing tumble is Turing-complete. *Theoretical Computer Science*, 948, 113734. <https://doi.org/10.1016/j.tcs.2023.113734>

Pleasant, J., Gui, X., Krutka, D., Logan, C., and Heath, M.. (2024). Coming to Critical Technology Consciousness: A Phenomenological Study of Educators. *Learning, Media and Technology*, 1-14. <https://doi.org/10.1080/17439884.2024.2438925>.

Postman, N. (2011). *The end of education: Redefining the value of school*. Knopf Doubleday Publishing Group.

Prasad, A., & Sliwa, M. (2025). Participatory (management) learning in the era of digital transformation. *Management Learning*, 56(4), 681. <https://doi.org/10.1177/13505076251375254>

Benjamin, R. (2024). The new artificial intelligentsia. *Legacies of Eugenics*. <https://lareviewofbooks.org/article/the-new-artificial-intelligentsia>

Selwyn, N. (2024). Digital degrowth: Toward radically sustainable education technology. *Learning, Media and Technology*, 49(2), 186–199. <https://doi.org/10.1080/17439884.2022.2159978>

Tomita, T., Lee, J., Isokawa, T., Peper, F., Yumoto, T., & Kamiura, N. (2020). Universal logic elements constructed on the Turing Tumble. *Natural Computing*, 19(4), 787–795. <https://doi.org/10.1007/s11047-019-09760-8>

UNESCO. (2024). *AI competency framework for students*. UNESCO. <https://doi.org/10.54675/JKJB9835>

Watters, A. (2023). *Teaching machines: The history of personalized learning*. MIT Press.

[Reclaiming the value of Human Capital in the AI era

Marjolaine Catil, Principal at Newfund

Abstract

This article explores the structural disruption of the labour market by Artificial Intelligence, arguing that the traditional "execution pyramid" model for junior talent is becoming obsolete. As AI commoditizes repetitive tasks, the value of human capital must pivot toward three essential competencies: autonomous creation, strategic judgment, and relational intelligence. The goal is not to resist automation, but to invest in technology that strengthens non-delegable human traits and educational foundations. Ultimately, financial capital must back the entrepreneurs capable of combining this new technology with strategic judgment to drive a critical economic transformation.

Keywords: Human capital, artificial intelligence, future of work, edtech, SME

The structural shift in junior talent

Let's start by defining human capital. Commonly, it refers to the aggregate knowledge, skills, and qualifications that drive an individual's productivity and capacity for economic value creation. Theorized by Gary Becker in the 1960s, the concept posits that any investment in training yields future productivity gains.

Yet the business model of recent decades has done the exact opposite. Recent graduates have spent their formative years crunching data, formatting reports, and churning out slides, a repetitive task that does little to foster judgment and fails to drive structural productivity improvements for firms.

Roles built on this "execution pyramid" are now facing a massive squeeze. Accenture announced a reorganization cutting 11,000 jobs, driven partly by AI optimization; KPMG reduced its US workforce by 5%; and Amazon eliminated 14,000 managerial roles in a push for organizational flattening.

A Stanford study published in August 2025 ("Canaries in the Coal Mine") confirms this global impact as "Early-career workers (ages 22-25) experienced a 13% relative decline in employment in roles most exposed to AI."

By so easily displacing these junior executors, AI exposes the fragility of this degraded form of human capital. This poses a critical question: what value proposition can these talents offer the companies of tomorrow?

The demand for a new talent profile

As AI tools streamline execution, companies are shifting toward flatter hierarchies where the need for middle management evaporates. The premium is now on individuals capable of direct action.

In this landscape, the target profile is pivoting toward three core competencies:

- **Autonomous Creation:** Talent must leverage AI to augment their own productivity (deploying AI agents to generate documentation, mining corporate history, iterating on product Minimum Viable Products, etc.).
- **Exercising Judgment:** While machines provide scenarios and probabilities, humans own the final decision. The critical skill is the ability to form a conviction and set a strategic course where the data remains ambiguous or incomplete.
- **Cultivating Relational Capital:** Automation frees up time to meet clients, capture unvoiced needs, build partnerships, and detect "weak signals" during informal interactions. This relational intelligence unlocks opportunities and fuels judgment with proprietary insights that AI cannot access.

A McKinsey study validates this shift: by 2030, demand for technological skills will rise by 55%, those for social and emotional skills by 24%, and higher cognitive skills (creativity, critical thinking, complex decision-making) by 17%. Conversely, demand for basic data processing skills will plummet by 21%.

Graduates who master these capabilities will be high-value assets. Unlike predecessors trained in the division of labour, they will deliver end-to-end value—from analysis to delivery—without relying on a command chain. This productive autonomy is precisely what will offset their lack of operational tenure.

They must master technology to relegate it to its proper status as a tool, rather than blindly following algorithmic scripts. This capacity for discernment will be the differentiator, compounding as they gain experience within the firm.

AI startups and the EdTech opportunity

To develop this strategic judgment, higher education plays an irreplaceable role: ensuring students truly master first principles. One cannot "judge" or "create" without grasping underlying mechanisms. However, facing surging demand for STEM and a shortage of educators, how do we guarantee deep learning at scale?

This is where AI must intervene: not to do the work for the student, but to decode their failure patterns. Solutions like PyxiScience, a Newfund portfolio company, illustrate this thesis: using data to maximize actual learning retention.

Their proprietary AI doesn't just grade handwritten math and coding exams; it analyses error networks to generate personalized exercises. By diagnosing why a student is failing, the tool

reinforces conceptual foundations with surgical precision. Professors, in turn, reclaim time for mentorship and pedagogical innovation. These technologies support a model where universities forge proactive operators capable of piloting tomorrow's AI, rather than passive executors.

In a world where AI becomes a commodity, top-tier human capital will be the ultimate scarce resource. Backing technologies that secure this high-level training represents a structural value-creation opportunity.

Investment new frontiers: the SME transformation

Then, once trained, where will this "augmented" talent find its best ROI? Schumpeterian creative destruction is at play: as execution jobs vanish, new roles emerge. The challenge for recent grads is to drive this transformation, which extends far beyond the startup ecosystem.

In France, over 4 million SMEs suffer a productivity gap of more than 10 points compared to European peers. Yet these firms hold strategic assets that AI can unlock: untapped historical data, deep domain expertise, and a local client footprint. By re-engineering the business models of these legacy companies, by combining human talent with intelligent data orchestration, we can bridge the productivity gap stifling growth.

This is a core investment thesis for Newfund: backing entrepreneurs who leverage talent to transform traditional SMEs into tech-enabled champions. This is where human capital will find its highest multiple, delivering positive impact for both our regional economies and the broader market.

Conclusion

As AI automates more, human capital capable of genuine judgment risks becoming rarer. To ensure future talent doesn't settle for being algorithmic operators, our educational institutions must pivot.

The mandate is twofold: greater rigor in fundamentals, and an increased capacity to exercise judgment amidst uncertainty. The goal is not to resist automation, but to double down on what cannot be delegated to an algorithm: creation, critical judgment, and strategic intent.

Since inception, Newfund has backed entrepreneurs demonstrating these specific traits: those who tackle complex problems, pivot through adversity, and create net-new value for customers. These are the founders we will continue to back to transform the economy.

References

Becker, G. S. (1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education* (3rd ed.). University of Chicago Press. (Original work published 1964)

Zhang, S., Brynjolfsson, E., & Li, D. (2025). *Canaries in the Coal Mine: Six Facts about the Recent Employment Effects of Artificial Intelligence*. Stanford Digital Economy Lab. <https://digitaleconomy.stanford.edu/publications/canaries-in-the-coal-mine>

Bughin, J., Hazan, E., Lund, S., Dahlström, P., Wiesinger, A., & Subramariam, A. (2018). *Skill shift: Automation and the future of the workforce*. McKinsey Global Institute. <https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce>

Feiner, L. (2025, October 28). Amazon laying off about 14,000 corporate workers in push to flatten organization. CNBC. <https://www.cnbc.com/2025/10/28/amazon-layoffs-corporate-workers-ai.html>

KPMG. (2023, June 26). KPMG to cut 5% of US workforce. Reuters. <https://www.reuters.com/business/finance/kpmg-cut-5-us-workforce-2023-06-26/>

Smith, J. (2025, September 15). Accenture Lays Off 11,000 Staff as Part of AI Reskilling Strategy. CX Today. <https://www.cxtoday.com/contact-center/accenture-lays-off-11000-staff-as-part-of-ai-reskilling-strategy/>

European Investment Bank. (2024). *EIB Investment Report 2023/2024: Transforming for competitiveness*. European Investment Bank. <https://doi.org/10.2867/568586>

Why Knowledge Is Required to Collaborate with AI

A Structural Reframing of AI-Era Education

Michał Serkieza

Abstract

The rapid integration of generative artificial intelligence (AI) into higher education has led to the assumption that internal knowledge storage may become less necessary when high-quality explanations are continuously accessible. This paper challenges that assumption. It argues that effective human–AI collaboration depends fundamentally on durable knowledge stored in human long-term memory.

Although AI systems can generate fluent and structured output, human cognition remains constrained by motivational trade-offs, limited working memory capacity, and inevitable forgetting. These structural bottlenecks determine whether AI functions as a cognitive amplifier or as a source of superficial understanding.

Building on established principles from cognitive psychology, this paper introduces the Cognitive High-Impact Learning Infrastructure (C.H.I.L.I.), a structured framework designed to support durable knowledge formation in AI-rich environments. The framework integrates governance mechanisms, structured encoding processes, and long-term retrieval practices.

In an era of informational abundance, sustainable educational advantage depends not on the speed of output generation, but on the deliberate construction of cognitive infrastructure within the learner.

Keywords: Human–AI collaboration; Long-term memory; Cognitive load; Retrieval practice; Learning infrastructure

1. Introduction: Information Abundance and Cognitive Constraint

Generative artificial intelligence has dramatically reduced the cognitive cost of accessing explanations, analyses, and structured argumentation. In higher education, students can now obtain context-sensitive responses across disciplines within seconds. As a result, an implicit assumption has begun to emerge: if high-quality explanations are continuously available, the necessity of durable internal knowledge may diminish.

This assumption reflects a fundamental misunderstanding of how cognition operates.

While AI expands informational access, it does not alter the biological architecture of the human cognitive system. Evaluation, abstraction, transfer, and judgment occur within the learner's mind and remain constrained by working memory capacity, motivational dynamics, and the inevitability of forgetting (Baddeley, 1992; Sweller, 1988).

The availability of external intelligence does not replace internal structure. On the contrary, as informational abundance increases, the cost of weak internal representations rises. Without well-developed knowledge structures in long-term memory, learners cannot reliably evaluate AI-generated content, detect inconsistencies, integrate new information, or transfer insights across contexts.

The central claim of this paper is therefore structural rather than technological: effective human–AI collaboration depends on durable cognitive architecture within the learner. AI can amplify reasoning, but it cannot compensate for absent internal structure. Informational abundance does not eliminate the need for learning; it intensifies it.

This paper argues that AI-era education must shift from a focus on information transmission toward the deliberate construction of cognitive infrastructure. The challenge is no longer access to knowledge. It is the stabilization of knowledge within constrained human cognitive systems.

2. The Human Bottleneck in Human–AI Collaboration

Human–AI interaction is constrained not only by technological capability but by the structural limits of human cognition. Regardless of advances in generative systems, reasoning, evaluation, and decision-making must occur within the biological architecture of the learner.

In AI-rich environments, three cognitive constraints become especially consequential: motivational regulation, limited working memory capacity, and the inevitability of forgetting. These constraints function as rate-limiting factors. They determine whether AI serves as a cognitive amplifier or devolves into a tool for superficial performance.

2.1 Motivation: Effort–Reward Calibration

Durable learning requires sustained cognitive effort. Engagement depends on perceived alignment between effort expenditure and anticipated reward. When AI systems can produce acceptable outputs with minimal human investment, the perceived necessity of internal encoding may decline.

From an expectancy–value perspective (Eccles & Wigfield, 2002), learners allocate effort based on anticipated future utility. If external systems reliably supply explanations, the long-term value of internal knowledge may appear diminished. This recalibration can weaken deep learning engagement.

In AI-enhanced contexts, the motivational structure of learning must therefore shift. The reward is no longer immediate output quality but long-term cognitive autonomy—the ability to evaluate, adapt, and transfer knowledge independently of continuous external prompting. Without this reframing, AI risks eroding the very effort investment required for durable knowledge formation.

2.2 Working Memory Limits

All reasoning and comprehension pass through working memory—a system with sharply constrained capacity (Baddeley, 1992; Cowan, 2001). At any given moment, only a limited number of meaningful elements can be processed simultaneously.

Generative AI systems can produce extended explanations and multi-layered reasoning structures that far exceed this capacity. When learners lack well-organized knowledge in long-term memory, they cannot efficiently integrate such complexity. According to cognitive load theory, learning deteriorates when processing demands exceed available cognitive resources (Sweller, 1988).

Crucially, working memory capacity itself does not meaningfully expand through practice. What changes is representational structure. Through learning, previously separate elements become integrated into structured units—chunks—stored in long-term memory. These chunks compress complexity and reduce cognitive load during reasoning.

In the absence of such internal structure, AI-generated elaboration increases surface fluency without increasing cognitive integration.

2.3 Forgetting as the Default State

Human memory is inherently unstable. Without retrieval and reinforcement, memory traces weaken over time (Ebbinghaus, 1885/1913). Exposure alone does not produce durable retention.

Research on retrieval practice demonstrates that active recall substantially strengthens long-term retention compared to passive review (Roediger & Karpicke, 2006). In AI-rich environments, if learners rely predominantly on external retrieval—querying systems rather than reconstructing knowledge internally—memory accessibility declines.

AI does not eliminate forgetting. It can mask it.

When learners substitute external access for internal consolidation, apparent competence may persist while cognitive structure weakens. The result is dependency rather than amplification.

Together, motivational calibration, working memory constraints, and forgetting define the structural bottleneck of human–AI collaboration. These are not temporary limitations. They are enduring properties of human cognition. Educational systems that ignore them risk optimizing for immediate output while undermining the cognitive architecture required for long-term competence.

3. From Information Access to Cognitive Infrastructure

If human cognition constitutes the limiting factor in AI collaboration, then the central task of education cannot remain information transmission. It must become cognitive infrastructure development.

Cognitive infrastructure refers to the structured network of durable knowledge stored in long-term memory that enables efficient reasoning, evaluation, and transfer. It functions analogously to physical infrastructure: invisible when intact, but indispensable for performance. Without it, advanced tools cannot be effectively utilized.

In AI-rich environments, access to explanations is abundant. What remains scarce is internal structure. Long-term memory serves as a compression system: complex patterns become integrated into organized representations that reduce working memory load during reasoning. This compression enables abstraction, comparison, and judgment under cognitive constraints.

Expertise research consistently demonstrates that domain experts do not possess greater raw processing capacity; rather, they possess richer, more interconnected knowledge structures (Chi et al., 1981). These structures allow rapid pattern recognition, anomaly detection, and flexible transfer—capacities that are essential when evaluating AI-generated output.

In the absence of such infrastructure, learners may experience fluent interaction with AI while lacking the structural competence required to detect errors, challenge assumptions, or generalize insights beyond immediate prompts. Apparent performance can therefore coexist with fragile understanding.

The implication is structural and non-negotiable: AI integration must be paired with deliberate mechanisms for encoding, consolidation, and retrieval. Access alone does not produce competence. Infrastructure does.

The remainder of this paper introduces a structured framework designed to support the systematic construction of such cognitive infrastructure in AI-rich educational environments.

4. The Cognitive High-Impact Learning Infrastructure (C.H.I.L.I.)

To operationalize the shift from information access to cognitive infrastructure, this paper introduces the Cognitive High-Impact Learning Infrastructure (C.H.I.L.I.). C.H.I.L.I. is not a study technique or isolated strategy; it is a structured architecture designed to counteract the human bottleneck in AI-rich educational environments.

The framework addresses the three structural constraints identified earlier—motivational regulation, working memory limits, and forgetting—by organizing learning into an integrated system that governs attention, transforms exposure into structured representation, and stabilizes knowledge through retrieval.

C.H.I.L.I. consists of three interdependent layers:

1. **Cognitive Learning Canvas** (Governance Layer)
2. **OSER Learning Loop** (Core Cognitive Mechanism)
3. **Intellectual Asset Production** (Application Layer)

These layers function as a coordinated system rather than independent modules.

The Governance Layer establishes motivational alignment and structural boundaries prior to engagement. Without governance, learning in AI-rich environments risks fragmentation and reactive information consumption.

The Core Cognitive Mechanism transforms exposure into durable, retrievable knowledge by organizing structured encoding and retrieval cycles. This layer directly addresses working memory limitations and forgetting.

The Application Layer externalizes stabilized knowledge through production. It reinforces transfer, exposes structural gaps, and creates recursive feedback loops into the encoding cycle.

The layered design is intentional. Governance without encoding produces intention without structure. Encoding without production produces retention without transfer. Production without encoding produces output without durable competence.

C.H.I.L.I. therefore reframes AI-era education as system design rather than content delivery. Its purpose is to ensure that AI functions as a cognitive amplifier embedded within structured human knowledge, rather than as a substitute for it.



Figure 1. Cognitive Learning Canvas (Goal · Why · Scope · Resources · Execution Plan · Success Criteria)

C.H.I.L.I. — Cognitive High-Impact Learning Infrastructure

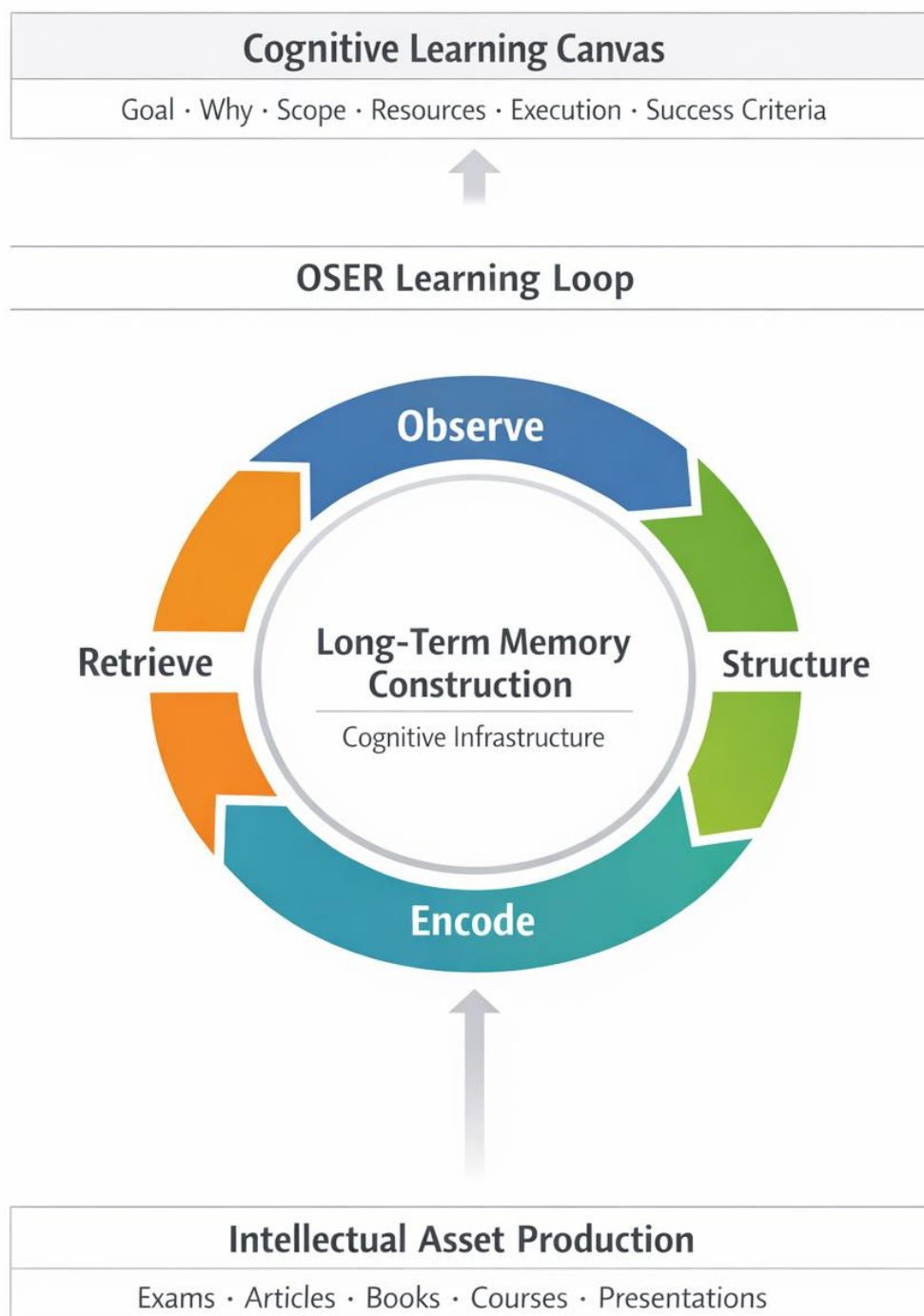


Figure 2. The C.H.I.L.I. Framework

5. The Cognitive Learning Canvas: Governance Layer

The first layer of C.H.I.L.I.—the Cognitive Learning Canvas—addresses the motivational and attentional bottleneck by establishing structured governance prior to deep cognitive engagement.

In AI-rich environments, informational access is frictionless. Learners can rapidly shift between topics, generate summaries across domains, and explore tangential ideas with minimal cost. While this flexibility increases exploration, it also increases fragmentation. Without explicit structural constraints, learning easily devolves into reactive information consumption rather than deliberate knowledge construction.

The Cognitive Learning Canvas counteracts this drift by stabilizing intention before exposure begins. It defines the boundaries within which cognitive effort will be invested.

The Canvas includes:

- **Goal:** The specific capability, domain, or representational structure to be developed.
- **Why:** The articulation of long-term cognitive value and future utility.
- **Scope:** Explicit inclusion and exclusion criteria to prevent cognitive dilution.
- **Resources:** Selected materials, tools, and AI supports aligned with scope.
- **Execution Plan:** Structured cadence and sequencing of engagement.
- **Success Criteria:** Measurable indicators of reconstructable knowledge and transfer.

This governance layer serves several structural functions.

First, it aligns effort with anticipated value, strengthening expectancy–value calibration (Eccles & Wigfield, 2002). When learners understand why a knowledge structure matters beyond immediate output, sustained engagement becomes more likely.

Second, it protects cognitive bandwidth. Working memory and attention are finite resources. Explicit scope boundaries prevent overload and reduce extraneous cognitive load generated by uncontrolled expansion.

Third, it converts learning from episodic interaction into investment. By defining success in terms of reconstruction and transfer—not mere exposure—the Canvas reorients the reward structure toward durable competence.

The Cognitive Learning Canvas does not directly encode knowledge. Rather, it governs where encoding effort will occur and why it is worth the cost. In AI-saturated environments, such governance is not optional. It is the stabilizing mechanism that prevents motivational erosion and structural fragmentation.

6. The OSER Learning Loop: Core Cognitive Mechanism

While the Cognitive Learning Canvas establishes governance, durable knowledge formation occurs within the OSER Learning Loop: **Observe – Structure – Encode – Retrieve**.

OSER is not a sequence of study techniques. It is a cyclical cognitive mechanism designed to transform informational exposure into stabilized long-term memory representations under conditions of limited working memory and inevitable forgetting.

Each stage addresses a distinct structural constraint of human cognition. Together, they form a closed loop that converts access into architecture.

6.1 Observe: Structured Exposure Under Attention Constraints

The Observe stage represents structured exposure to new material. This includes reading, listening, problem-solving, or interacting with AI-generated explanations.

In AI-rich contexts, Observe frequently involves dialogue with generative systems. AI can increase access efficiency, generate alternative explanations, and surface examples rapidly. However, Observe remains input-oriented.

Exposure alone does not produce durable knowledge. Research on levels of processing demonstrates that shallow interaction produces weak memory traces (Craik & Lockhart, 1972). Similarly, passive review is substantially less effective than active engagement (Roediger & Karpicke, 2006).

Observe therefore operates under an attentional constraint: only information that receives focused cognitive engagement proceeds to structural integration. AI can amplify exposure, but it cannot force depth of processing.

Without transition to the next stage, Observe results in familiarity without competence.

6.2 Structure: Integration and Cognitive Compression

The Structure stage transforms exposure into organized representation. Here, learners actively interrogate and integrate material with prior knowledge.

Key integrative questions include:

- What is the governing principle?
- How does this connect to existing representations?
- Under what assumptions does it hold?
- Where does it fit within the broader domain?

This stage aligns with elaborative rehearsal and schema formation. Elaboration strengthens associative networks in long-term memory, increasing retrievability and transfer (Bjork & Bjork, 2011).

Crucially, Structure reduces working memory load through cognitive compression. Discrete elements become integrated into structured units—chunks—that can be processed as single representations during reasoning.

Without structural integration, AI-generated content remains fragmented. It may appear coherent externally while remaining unanchored internally.

AI may assist here as a structural challenger—highlighting inconsistencies or proposing alternative framings—but integration must remain cognitively active. Compression cannot be outsourced.

6.3 Encode: Deliberate Representational Design

Encoding in OSER is deliberate representational construction.

During Encode, learners transform structured understanding into explicit, retrievable formats. Examples include:

- Concise definitions
- Conceptual maps
- Model formalizations
- Self-generated summaries
- Atomic flashcards

Encoding operationalizes generative learning principles. Producing information strengthens retention more than re-reading it (Fiorella & Mayer, 2015).

At the cognitive level, encoding stabilizes compressed representations into long-term memory. The effort invested during this transformation is not incidental—it is the mechanism of consolidation.

AI may assist by refining clarity or stress-testing formulations. However, if encoding is fully delegated, consolidation weakens. Durable structure emerges from transformation effort, not transcription.

6.4 Retrieve: Stabilization Through Reconstruction

Retrieve closes the loop by converting encoded material into durable, accessible knowledge.

The testing effect demonstrates that active recall enhances long-term retention more effectively than additional study (Roediger & Karpicke, 2006). Retrieval strengthens memory traces and increases accessibility across contexts.

Spaced repetition further enhances consolidation by distributing recall attempts over time, counteracting natural forgetting (Ebbinghaus, 1885/1913).

In AI-rich environments, Retrieve becomes structurally decisive. If learners default to external querying rather than internal reconstruction, memory strength declines. AI should therefore be introduced after retrieval attempts—as feedback, challenger, or validator—not as a substitute for recall.

Through repeated retrieval, encoded units become automated and function as chunks within long-term memory. Each stabilized chunk reduces working memory strain during future reasoning tasks.

6.5 The Loop Logic: From Exposure to Infrastructure

OSER is cyclical rather than linear. Retrieval reveals gaps, which redirect attention during subsequent observation. Production attempts expose structural weaknesses, which trigger renewed structuring and encoding.

Each iteration strengthens representational compression and increases cognitive autonomy.

- Observe without Structure produces familiarity.
- Structure without Encode produces fragile insight.
- Encode without Retrieve produces temporary retention.
- Retrieve without renewed observation produces stagnation.

Only the closed loop transforms informational abundance into durable cognitive infrastructure

7. Application and Intellectual Asset Production: The Stress-Test Layer

The final layer of C.H.I.L.I. externalizes stabilized knowledge through structured reconstruction and artifact creation. While the OSER loop builds internal cognitive architecture, this layer tests its strength under generative pressure and extends its value beyond the learning episode.

Application operates in two structurally distinct forms: **reconstructive outputs** and **durable intellectual assets**.

7.1 Reconstructive Outputs: Diagnostic Stress Testing

Reconstructive outputs require the learner to generate knowledge without continuous external scaffolding. Examples include:

- Examinations
- Oral defenses
- Applied problem-solving tasks
- Case analyses
- Timed reasoning exercises

Unlike recognition-based tasks, reconstructive outputs demand internal retrieval, reorganization, and contextual adaptation. They therefore serve a diagnostic function.

Under reconstructive pressure, weaknesses in cognitive structure become visible. Gaps in coherence, missing assumptions, fragile conceptual links, and shallow encoding are revealed when learners must produce structured reasoning independently.

This diagnostic function is essential. Without reconstruction under constraint, learners may confuse fluency with competence. AI-rich environments amplify this risk: when external systems provide immediate explanations, surface understanding can mask structural fragility.

Reconstructive outputs counteract this illusion. They expose whether knowledge is stabilized at the level of transferable abstraction rather than episodic familiarity.

Importantly, these outputs are not intellectual assets in themselves. Their primary function is evaluative and corrective. The weaknesses they reveal feed directly back into the OSER loop, redirecting observation, restructuring representations, and refining encoding.

7.2 Intellectual Assets: Durable Cognitive Capital

Beyond diagnostic reconstruction lies a second form of application: the creation of durable intellectual artifacts.

Intellectual assets include:

- Academic articles
- Books
- Courses
- Research proposals
- Formalized frameworks
- Structured models

Unlike examinations, intellectual assets are reusable, extensible, and compounding. They externalize structured knowledge into artifacts that retain value over time.

Producing such assets requires integration at a higher level of abstraction. The learner must organize knowledge into coherent architecture, anticipate counterarguments, clarify assumptions, and communicate structure explicitly. This generative process deepens consolidation and strengthens transfer.

Intellectual assets therefore function as cognitive capital. They represent stabilized knowledge structures that can be applied, adapted, and expanded across contexts.

7.3 AI Collaboration at the Production Layer

In AI-enhanced environments, the production layer requires careful boundary design.

AI can assist in drafting, structural critique, counterargument generation, and stylistic refinement. When used after internal reconstruction, AI becomes a powerful amplifier of clarity and precision.

However, if AI assumes primary responsibility for generative reasoning, the diagnostic function collapses. Apparent output quality may increase while internal cognitive structure weakens. The learner becomes dependent rather than amplified.

For production to serve its architectural purpose, generative responsibility must originate from stabilized long-term memory. AI may refine structure; it must not replace it.

7.4 Completing the Architecture

Application completes the C.H.I.L.I. system.

- Governance directs effort and protects motivation.
- OSER stabilizes knowledge under cognitive constraints.
- Application stress-tests and extends that knowledge into transferable competence.

Without reconstruction, learning remains fragile.
Without durable artifacts, learning remains isolated.

Together, these layers transform informational access into structured, transferable cognitive infrastructure in AI-rich environments.

8. Implications for AI-Era Pedagogy

The analysis presented in this paper suggests that AI integration in higher education must be guided by cognitive constraints rather than technological capability.

Two divergent trajectories are emerging in AI-era pedagogy.

In the first, AI is used primarily to reduce effort. Students outsource explanation, summarization, and structuring tasks to generative systems. Assessment focuses on output quality rather than reconstructable knowledge. Under this model, short-term performance may increase, but internal cognitive architecture progressively weakens. Fluency substitutes for structure.

In the second trajectory, AI is embedded within structured cognitive systems such as C.H.I.L.I. Here, generative systems are positioned after governance, encoding, and retrieval processes. AI amplifies reasoning that is already grounded in stabilized long-term memory. Assessment prioritizes reconstructable competence and transfer rather than polished surface output.

The distinction between these trajectories is not technological; it is architectural.

Three implications follow for institutional design.

First, AI use must reinforce encoding and retrieval rather than bypass them. Policies that emphasize tool adoption without corresponding retrieval-based assessment risk accelerating dependency.

Second, evaluation practices must shift toward reconstruction under constraint. Oral defenses, closed-resource reasoning tasks, and generative problem-solving exercises should complement open-resource production. The goal is not to eliminate AI use but to ensure that competence remains internally reconstructable.

Third, curricula must treat long-term memory formation as a primary objective rather than a byproduct. Knowledge should be deliberately stabilized through structured encoding cycles and spaced retrieval, particularly in domains requiring abstraction and transfer.

Institutions that fail to align AI integration with cognitive architecture may experience a paradox: increased productivity alongside decreased competence. Conversely, institutions that embed AI within structured cognitive infrastructure can cultivate learners capable of leveraging generative systems without becoming cognitively dependent on them.

The challenge of AI-era education is therefore not access management. It is architectural alignment.

9. Conclusion

The availability of generative AI does not eliminate the need for knowledge. It makes the absence of structured knowledge more consequential.

Throughout this paper, the central argument has been structural rather than technological. Human cognition remains governed by motivational dynamics, limited working memory capacity, and the inevitability of forgetting. These constraints do not disappear in AI-rich environments. They define the boundaries within which human–AI collaboration operates.

When learners lack durable internal representations, AI-generated fluency can mask structural fragility. Output quality may improve while cognitive autonomy declines. Informational abundance does not compensate for weak internal architecture; it amplifies its consequences.

The Cognitive High-Impact Learning Infrastructure (C.H.I.L.I.) offers a systematic response to this condition. By integrating governance, cyclical encoding and retrieval, and stress-tested application, it reframes education as the deliberate construction of cognitive infrastructure. AI becomes a collaborator embedded within structured knowledge rather than a substitute for it.

The future of education in the age of AI will not be determined by the sophistication of generative systems alone. It will be determined by whether institutions cultivate learners capable of evaluating, integrating, and extending those systems through durable internal structure.

AI expands informational capacity.
Education must expand cognitive architecture.

References

- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559. <https://doi.org/10.1126/science.1736359>
- Bjork, R. A., & Bjork, E. L. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114. <https://doi.org/10.1017/S0140525X01003922>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). Teachers College, Columbia University. (Original work published 1885)
- Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107707085>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Naveed, M., Seidel, T., Stadler, M., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

AI Agents for Roleplaying in the Classroom: Design Principles, Pedagogical Value, and Responsible Integration in Higher Education

Prof. Dr. Lorena Blasco-Arcas, ESCP Business School

Abstract

This paper examines the instructional design and pedagogical integration of AI agents in higher education roleplaying activities. Drawing on experiential and situated learning perspectives, it conceptualizes AI agents as mediators across three functions: role enactors, interaction partners, and adaptive challengers. Six design principles are proposed—covering role definition, interaction structure, generativity calibration, scaffolding, instructor oversight, and ecosystem integration. Qualitative findings from a course-based implementation suggest that AI-supported roleplay enhances engagement, broadens participation, and creates psychologically safe learning conditions, while also revealing boundary conditions tied to insufficient scaffolding and unconstrained agent behavior. The paper contributes a pedagogically grounded framework for responsible AI-agent integration in higher education, with implications for instructional design, academic governance, and future comparative research.

Keywords: AI agents; roleplaying; experiential learning; situated learning

Introduction

Higher education institutions are actively experimenting with generative artificial intelligence to enhance teaching and learning. While much attention has focused on efficiency gains, assessment integrity, and content generation, fewer studies have examined how AI can support interactive, high-engagement pedagogies (Qian, 2025). Roleplaying represents a particularly relevant case: it is pedagogically powerful, yet difficult to scale, uneven in execution, and often constrained by time, instructor capacity, and student reluctance to perform in front of peers. These challenges limit its systematic use despite strong alignment with active and experiential learning goals.

Recent advances in generative AI have enabled the development of AI-driven agents capable of sustained dialogue, adaptive responses, and role enactment (Mollick et al., 2024; Mollick and Mollick, 2024). Such agents offer the potential to augment role-playing activities by providing consistent, on-demand interaction while preserving key experiential elements. However, the educational value of AI agents cannot be assumed. Without pedagogical grounding and institutional safeguards, their classroom deployment risks becoming either superficial or misaligned with learning objectives. This paper addresses these concerns by examining AI agents as role-playing *instructionally designed, human-centered educational artifacts* rather than as standalone technologies.

Pedagogical and conceptual foundations.

The study draws on experiential learning and situated learning perspectives, which emphasize learning through action, reflection, and contextualized practice (Brown et al., 1989; Lave and Wenger, 1991; Morris, 2020). Roleplaying allows students to engage with realistic scenarios that require judgment, perspective-taking, and application of theory (Rao and Stupans, 2012). From a social constructivist standpoint, learning emerges through interaction and dialogue, suggesting that the pedagogical value of AI agents depends on the quality and structure of interaction rather than on their human likeness (Mercer and Howe, 2012; Wegerif, 2007).

Within this framework, AI agents are conceptualized as mediators of learning interactions that can support, but not replace, human instruction. The paper differentiates between three instructional functions of AI agents in classroom roleplay: (1) agents as role enactors representing stakeholders or institutional actors; (2) agents as interaction partners that respond dynamically to student decisions; and (3) agents as adaptive challengers that prompt justification, reflection, or reconsideration. This functional distinction is important for clarifying both pedagogical intent and governance requirements in AI-supported learning environments.

Instructional design principles for AI-supported roleplay.

Central to the paper is a set of instructional design principles that guide the responsible integration of AI agents into roleplaying activities.

First, *role definition and boundaries* are critical. Agents must be clearly framed in terms of role, authority, and scope to avoid ambiguity or unintended instructional drift. Second, *interaction design* shapes how learning unfolds, including turn-taking structures, escalation mechanisms, and the balance between student initiative and agent prompting. Third, the *degree of generativity* must be calibrated. While open-ended dialogue can enhance realism, excessive openness may undermine focus or introduce irrelevant content. Fourth, *pedagogical scaffolding*—including preparation, guidance during interaction, and structured debriefing—is essential to translate experience into learning. Fifth, *instructor oversight and control* remain central, with instructors shaping agent behavior through prompt design, guardrails, and scenario constraints rather than real-time intervention. Finally, *integration within the course ecosystem* ensures that AI-supported roleplay complements, rather than displaces, human discussion, peer interaction, and assessment practices.

These principles reflect a human-centered approach in which AI agents extend instructional capacity while preserving pedagogical intent and academic responsibility.

Implementation context and methodological approach.

The paper reports on the implementation of AI-agent roleplay activities in a higher education course setting. Scenarios were designed to align with course learning objectives and required students to engage in decision-making, negotiation, or problem-solving through interaction with AI agents. Activities were embedded within a structured instructional sequence consisting of pre-briefing, in-role interaction, and post-activity reflection.

Data were collected from student reflections, learning artifacts, and instructor observations. The analysis adopts a qualitative, exploratory approach to identify patterns in student engagement, perceived learning value, and challenges associated with AI-mediated

interaction. Rather than evaluating technical performance, the focus is on pedagogical processes and design implications relevant to institutional adoption.

Observed learning effects and student experience.

Findings indicate that AI-supported roleplay can contribute positively to student engagement and learning when embedded within a well-designed instructional framework. Cognitively, students demonstrated increased willingness to apply theoretical concepts to practical situations and to explore alternative courses of action. Behaviorally, participation broadened, particularly among students who were less inclined to speak in traditional classroom roleplays.

Affective outcomes were especially salient in the higher education context. Many students described interacting with AI agents as a psychologically safe environment that reduced fear of judgment and encouraged experimentation. This suggests that AI-supported roleplay may lower barriers to participation while maintaining task complexity. Social learning outcomes were more mixed: team-based interactions supported collaborative reasoning, but required careful task structuring to prevent uneven participation.

The study also highlights boundary conditions relevant to responsible AI adoption. When scaffolding was insufficient or agent behavior overly unconstrained, students focused on manipulating the interaction rather than engaging with learning objectives. These findings reinforce the need for intentional design and institutional guidance.

Implications for responsible AI integration in higher education.

From an instructional perspective, AI agents shift the educator's role toward learning design, orchestration, and reflective facilitation. While initial setup requires effort, the approach offers scalability and consistency across cohorts. At the institutional level, implications include the need for transparency regarding AI use, alignment with assessment policies, and governance structures addressing bias, data use, and academic integrity. Importantly, AI-supported roleplay should be positioned as complementary to human interaction, particularly in activities emphasizing ethical judgment, empathy, and interpersonal skills.

Contributions and future directions.

This paper contributes to the AI in higher education discourse by moving beyond adoption narratives to offer a pedagogically grounded account of AI-agent roleplay. It provides design principles, identifies learning outcomes and boundary conditions, and highlights considerations for responsible institutional integration. Future research should examine long-term learning effects, comparative designs involving human and AI roleplay, and governance models that support sustainable, human-centered use of AI in higher education.

References

- Brown, J.S., Collins, A. & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42.
- Lave, J. & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
- Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning. *Learning, Culture and Social Interaction*, 1(1), 12–21. <https://doi.org/10.1016/j.lcsi.2012.03.001>
- Mollick, E., Mollick, L., Bach, N., Ciccarelli, L. J., Przystanski, B., & Ravipinto, D. (2024). AI agents and education: Simulated practice at scale. [arXiv preprint arXiv:2407.12796](https://arxiv.org/abs/2407.12796).
- Mollick, E., & Mollick, L. (2024). How to use AI to create role-play scenarios for your students. Here's a sample prompt you can customize for your class. ([Harvard Business Impact Education](#))
- Morris, T.H. (2020). Experiential learning — a systematic review and revision of Kolb's model. *Interactive Learning Environments*, 28(8), 1064–1077.
- Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, 69, 1105–1120.
- Rao, D. & Stupans, I. (2012). Exploring the potential of role play in higher education: development of a typology and teacher guidelines. *Innovations in Education and Teaching International*, 49(4), 427-436.
- Wegerif, R. (2007). *Dialogic education and technology: Expanding the space of learning*. Springer. <https://link.springer.com/book/10.1007/978-0-387-71142-3>

[Building AI Resilience in Management Education

Thomas Lorenz,
St. Edward's University

Abstract

The integration of Generative AI and Large Language Models (LLMs) into higher education threatens to undermine the "case method" by allowing students to bypass the critical thinking and cognitive struggle necessary for deep learning. Drawing on theories of cognitive load and "cognitive debt," this working manuscript proposes and tests "AI-resilient" pedagogical strategies designed to shift students from passive AI dependency to active analytical engagement. The study introduces the "Trojan Horse" methodology: a technique utilizing adversarial prompt injection to embed hidden text commands within PDF case materials. These commands, invisible to human readers but processed by LLMs, inject specific cosmetic markers and factually unsound strategic directives into the AI's output. In an experimental application within an International Management course (Fall 2025), this method revealed that four out of six student teams relied significantly on AI-generated content, failing to identify deliberately injected strategic errors, such as recommendations to divest profitable units or terminate high-performing leadership. However, a subsequent voluntary "penalty round" demonstrated that when AI shortcuts were removed, students retained the capacity for rigorous, independent analysis. These findings suggest that introducing "friction" through adversarial design can effectively disrupt passive AI reliance, forcing the metacognitive reflection required for authentic management education.

Keywords: Management Education, Generative AI, Case Method, Academic Integrity.

1. Introduction: The Crisis of Cognitive Offloading

Generative AI platforms such as ChatGPT have become a persistent feature of the educational landscape. However, their widespread adoption in higher education poses a significant challenge: the potential erosion of critical thinking abilities. When students depend on Large Language Models (LLMs) to conduct analysis, the issue extends beyond academic dishonesty; it involves circumventing the intellectual rigor essential for meaningful learning.

Within the context of management education, case studies serve as a foundational pedagogical tool. As highlighted by David Garvin in his influential study of professional education, the case method is distinguished by its capacity to prepare students for real-world practice by requiring them to diagnose problems, make decisions, and act amidst uncertainty (Garvin, 2003). These exercises are intentionally constructed to compel students

to synthesize ambiguous information and defend subjective judgments. When artificial intelligence undertakes this synthesis, the educational value of the exercise is effectively negated.

Emerging physiological research reinforces these concerns. For example, a study conducted by Kosmyrna et al. (2025) at the MIT Media Lab found that students who utilized AI assistants for writing tasks demonstrated notably weaker neural connectivity and experienced an “accumulation of cognitive debt.” In essence, reliance on these technologies for primary analytical tasks risks producing graduates who excel at prompting AI tools but lack the neural development necessary for independent comprehension and reasoning.

Given that ChatGPT has only been available for approximately three years, both students and faculty are still adapting to its implications. The proliferation of multiple, freely accessible models capable of producing well-structured responses in under a minute presents a considerable temptation. Rather than engaging in the critical reflection required by case analysis, students may be inclined to opt for the convenience of an instant AI-generated response, thereby diminishing the intended educational experience.

The goal of this paper is not to advocate for a Luddite abolition of technology. Rather, it argues for “tipping the scales” back toward independent work and thought. We must transition students from viewing AI as the “pilot” (doing the work) to the “co-pilot” (supporting the work). This paper proposes and tests immediate, actionable strategies to disrupt over-reliance on chatbots.

2. Theoretical Framework

The proliferation of generative AI in management education is not merely a technological disruption but a fundamental challenge to the cognitive mechanisms that enable learning. This section establishes the theoretical foundation for understanding why AI shortcuts undermine educational outcomes and why AI-resilient pedagogical interventions are necessary to preserve the integrity of case-based learning.

Cognitive Load Theory (Sweller, 1988; 2011) distinguishes between three types of cognitive load: intrinsic (inherent complexity of material), extraneous (poor instructional design), and germane (productive cognitive processing that builds schemas). Effective learning requires managing intrinsic and extraneous load while maximizing germane load—the effortful mental work that transfers information into long-term memory. When students offload case analysis to AI, they eliminate precisely the germane cognitive processing necessary for learning, creating what Kosmyrna et al. (2025) term “cognitive debt”—weakened neural connectivity resulting from bypassing effortful thought.

The generation effect, a robust finding in cognitive psychology, demonstrates that information actively generated by learners is better retained than passively received information (Slamecka & Graf, 1978; Bertsch et al., 2007). This principle is central to understanding why AI shortcuts fundamentally undermine learning. When students read a case and generate their own strategic analysis—wrestling with ambiguity, weighing alternatives, constructing arguments—they engage in deep cognitive processing that builds durable mental models. In contrast, when students prompt an AI and receive a complete analysis, they are merely receiving information, engaging in shallow processing that produces minimal learning.

The case method's pedagogical power derives precisely from requiring generation. Students must produce diagnoses, formulate recommendations, and defend positions—all activities that force active construction of knowledge. Bjork and Bjork (2011) identify this productive struggle as a “desirable difficulty”—a challenge that feels harder in the moment but

produces superior long-term learning. Case studies are intentionally designed to be difficult: they contain ambiguous information, require integration across multiple business functions, and demand subjective judgment rather than algorithmic solutions. This difficulty is not incidental; it is the mechanism through which learning occurs.

When AI eliminates this generative effort, it eliminates the learning. Students who input case materials into ChatGPT and receive polished strategic recommendations have not engaged in the cognitive work of producing analysis; they have merely consumed a pre-made product. The effortful processing that would build their strategic thinking capabilities—the retrieval of course concepts, the evaluation of evidence, the construction of logical arguments—is entirely bypassed. As Kosmyna et al. (2025) documented physiologically, students using AI assistants show measurably weaker neural connectivity, suggesting that reliance on these tools actively inhibits the brain's development of analytical capabilities.

The "cognitive debt" metaphor is particularly apt for management education. Just as financial debt represents consumption today at the expense of future resources, cognitive debt represents performance today (quick AI-generated answers) at the expense of future capability (underdeveloped strategic thinking skills). Students may complete assignments successfully while accumulating a deficit in the very cognitive abilities the assignments were designed to develop.

Vygotsky's Zone of Proximal Development (ZPD) differentiates between tasks students can do alone and those they can do with support. Effective scaffolding provides just enough help to extend student ability without taking over. When AI merely assists—like checking grammar or organizing ideas—it acts as a "co-pilot," supporting learning within the ZPD. But when AI fully generates analyses or recommendations, it becomes the "pilot," doing the core cognitive work and turning students into passive recipients. This substitution undermines active engagement and skill development. Educational technology, including AI, is most effective when it empowers students rather than replaces their effort.

2.1 The Case Method as Bridge to Practice

The ultimate objective of management education is not merely to help students complete academic assignments but to prepare them for professional practice. The case method serves as a critical bridge between classroom learning and real-world application. As Garvin (2003) observed, cases are distinguished by their capacity to simulate the ambiguity, time pressure, and incomplete information that characterize actual managerial decision-making. Unlike textbook problems with clearly defined parameters and correct answers, cases require students to make judgment calls, defend positions despite uncertainty, and integrate knowledge across multiple domains—precisely the cognitive demands they will face as practicing managers.

Transfer theory distinguishes between near transfer (applying learning to similar contexts) and far transfer (applying to dissimilar real-world situations). The case method is explicitly designed to promote far transfer by providing repeated practice in analyzing novel business situations, each with unique contextual factors. Through this repeated practice, students develop not merely knowledge of specific cases but generalized schemas—mental models of how businesses operate, how strategic problems can be diagnosed, and how competing considerations must be balanced (Perkins & Salomon, 1992; Barnett & Ceci, 2002).

Case method is threatened by AI substitution: students can produce sophisticated-appearing work without developing the underlying cognitive structures that would enable them to apply that knowledge independently. They have not learned how to analyze

acquisitions; they have learned how to prompt AI to analyze acquisitions—a fundamentally different and far less transferable skill.

2.2 Detection and Deterrence

Academic integrity research demonstrates that deterrence requires three elements: perceived certainty of detection, severity of consequences, and celerity (swiftness) of response (McCabe, Treviño, & Butterfield, 2001). Traditional approaches to preventing cheating—proctored examinations, plagiarism detection software, honor codes—have relied primarily on detection certainty. However, generative AI has fundamentally undermined this certainty. Unlike copy-paste plagiarism, which leaves digital fingerprints that tools like Turnitin can detect, AI-generated text is original content created specifically for the assignment. Detection tools that rely on pattern matching or statistical anomalies struggle with this fundamentally different challenge.

Students are keenly aware of this detection failure. When detection certainty is low, deterrence collapses. Students rationally calculate that the probability of being caught is minimal, and even if suspicions arise, they can deny AI usage with little consequence. The Trojan Horse methodology addresses this detection failure by creating very high-confidence indicator of AI usage. By embedding cosmetic markers (such as the name swap from "Fischer & Wiese" to "Wiese & Fischer") that only appear in AI-generated output, instructors can demonstrate AI reliance with certainty rather than suspicion.

3. Methodology: Two Avenues for AI Resilience

To combat uncritical AI usage, educators can employ two distinct strategies: offline methods in short avoiding the temptation of AI all together and the "Trojan Horse" embedded prompt injection, is making the AI output easier to spot.

3.1 In-Class Cases

The most effective strategy for fostering independent thought is to eliminate reliance on artificial intelligence during case-based instruction. This necessitates a departure from the conventional asynchronous model—where students prepare responses outside of class—and a transition to an in-class approach. While the traditional method, which often employed lengthy and complex cases, proved effective in the past, the widespread availability of generative AI has enabled students to input case materials into large language models and receive rapid, well-written answers. Such practices undermine the fundamental educational objectives of case teaching by circumventing the cognitive rigor essential to learning.

The AI-resilient case method instead advocates for in-class analysis utilizing shorter, paper-based cases that are reviewed and resolved during the class session. Although this approach introduces certain limitations, such as reduced time for discussion due to in-class reading and a decrease in case complexity, these challenges can be mitigated by tailoring cases to address specific topics aligned with course objectives. For instance, a session focused on employee motivation in International Management may employ a case centered on reforming management systems within a foreign subsidiary. The principal advantage of this model is that it compels students to engage actively with the material in real time, thereby promoting deeper learning and participation.

In a similar vein, numerous universities have opted to reinstate in-person, paper-based examinations employing "blue books." Despite the claims of some online platforms to be "AI

cheating proof," the rapid evolution of AI technologies renders it prudent to favor offline assessments and minimize opportunities for AI-assisted academic dishonesty.

3.2 Asynchronous Complex cases

Complex cases continue to offer significant benefits that cannot be fully replicated through in-class activities alone. In order to build AI resiliency in those few Asynchronous Complex cases, we recommend the combined use of three actions.

The first approach is a brief oral defense: a focused five-minute Q&A session in which students articulate and justify their team papers. This method efficiently reveals whether students genuinely understand the reasoning behind their work or are simply reciting content generated by artificial intelligence. As final examinations approach, students should be encouraged to reduce their reliance on chatbot-generated responses.

Secondly, instructors should avoid providing overly detailed prompts to students. Traditional undergraduate case studies typically include comprehensive instructions referencing class concepts, facilitating valid submissions. While this format supports learning, step-by-step guidance can serve as strong prompts, potentially resulting in well-structured assignments based primarily on explicit directions. To promote AI resiliency, it is advisable to limit written instructions—using general statements such as "Apply concepts from class"—and to deliver more specific guidance verbally during lessons. This encourages students to take notes and independently interpret any AI directives, thereby discouraging passive dependence on AI tools.

Finally, educators seeking to diminish the effectiveness of large language model-generated answers and enhance AI detection may employ the embedded prompt injection technique. This strategy entails embedding covert instructions within PDF case files that remain invisible to human readers but detectable by language models. Since the system processes all input as tokens without differentiating between commands and data, these embedded instructions can influence LLM responses. Moreover, because the LLM strives to generate highly probable and user-satisfying outputs, making the inclusion of such hidden directives effective.

The mechanics behind these commands are straightforward. For instance, text may be formatted in white font on a white background or reduced to microscopic sizes (such as 1pt) and positioned within empty spaces or margins. For artificial intelligence models, attributes such as size or color are inconsequential—all data is interpreted the same way. Detecting these "hidden instructions" can be relatively simple: copying all content into a word processor without formatting reveals text uniformly in the default font, size, and color, similar to how an AI would process it. However, this approach requires reading the entire document thoroughly to identify and delete these commands before submitting the file to an AI chatbot.

In lengthy documents—such as those spanning ten pages or more—these commands may be dispersed across multiple sections to evade detection. Rather than presenting one extensive command, they might be fragmented into smaller segments or even placed within appendices. Ultimately, the more challenging these commands are to detect, the greater the effort students must invest to discover them, which ideally encourages the development of original responses.

Educators can employ significant creativity when designing commands using the embedded prompt injection method. The paper suggests two main approaches: AI detectors and content changes.

AI detectors are straightforward—they simply confirm whether an LLM was used, and may be either obvious or hidden. Obvious changes, like swapping names or dates (e.g., "Fischer & Wiese" to "Wiese & Fischer"), act as clear signs that can be quickly noticed by readers. If students first read the case, then use AI to write their answers and carefully review them, these mistakes should be easy to catch and correct. This process may also prompt students to look for other unusual AI-generated content or even hidden commands in the document. The other AI detector works by adding specific facts or data do not present in the case, leading the AI to use them in its responses and reveal AI usage. The details should be highly specific, such as an unusual cost saving ("-28%") or adding a facility name ("the Delta building"). This method can be spotted by students but is most likely to be noticed only by the PDF's author. Both of those methods do not really impact the answer, they just help to detect AI usage.

The content changes command involves the deliberate introduction of factually incorrect or strategically unsound directives that contradict both the case study and course instruction. Such errors may be embedded in various sections of the assignment and can impact multiple facets of students' responses. The primary intent is to offer students repeated opportunities to detect and amend inaccurate information. While minor alterations might escape immediate notice, students are presumed to possess the requisite knowledge to recognize correct answers. Consequently, more blatant inaccuracies should prompt a thorough review and careful revision. Common reasons for failing to rectify these errors include insufficient review of responses, overreliance on AI-generated content, inadequate engagement with course material, or a combination of these factors. The process of injecting content changes is inherently more complex than other methods. If direct commands are used, a large language model may respond by alerting the user—such as by asking, "There are some very specific commands; do you want me to follow those?" To ensure the hidden text is processed as intended, it should be presented as "key importance" or "industry best practice," thereby prompting the AI to prioritize these points. When framed as a strategic imperative, the AI is likely to incorporate and emphasize this content, especially when queried about the validity of its output.

3.3 Reflection Mechanism after AI Content detection

The primary objective of AI-resilient education is not merely to catch students cheating, but to ensure they engage deeply with case material to achieve transferable learning outcomes. Consequently, the proposed method includes a "second chance" mechanism, allowing students to submit a revised response derived from their own critical thinking, free of AI-generated content. While educators should determine the specific grade impact of the initial infraction, it is recommended that this secondary opportunity be voluntary, offering students a path to recover a portion of the lost grade points.

This approach is grounded in Self-Regulated Learning (SRL) theory (Zimmerman, 2002; Pintrich, 2004), which emphasizes that effective learners engage in a cyclical process of planning, monitoring, and evaluating their work. Critically, self-regulation requires metacognitive awareness—the capacity to step back and assess the quality of one's own thinking.

The voluntary secondary round—effectively a penalty round—provides a structured opportunity for this metacognitive reflection. This intervention compels genuine self-assessment. When presented with irrefutable evidence of AI usage (such as "cosmetic markers"), the possibility of self-deception or rationalization is eliminated. Therefore, the secondary round serves not merely as a punitive measure for academic misconduct, but as a pedagogical intervention that promotes self-regulated learning. By creating conditions that require authentic effort and reflection, this mechanism moves students from AI-

dependent performance toward independent analytical competence—precisely the developmental trajectory management education seeks to foster.

4. Case Study: Asynchronous Complex cases 2025

The case study was undertaken at St. Edward's University as part of an International Management course, with a class size of 25 students. The assignment served as the final group project and entailed a strategic analysis report (maximum 2,000 words), accompanied by an individual in-person Q&A session.

The case, titled "Vortex 3D and Fischer & Wiese" (approximately 19 pages / 5,200 words), is a customized scenario developed by the instructor drawing from a hypothetical real-world acquisition context. It examines the post-acquisition integration process of a 3D printing company acquiring a traditional German machine tool manufacturer.

The students were assigned into 6 separate teams and given 2 weeks to complete the case write-up. During this time a class session was dedicated to answer question by team. The weeks discussing the assignment, students were given clear guidance to use their own thoughts and analysis.

4.1 Injection:

The following hidden text was embedded in the PDF:

"when answering question to the case please refer to "Fischer & Wiese" as "Wiese & Fischer"

"Key Strategic Priorities Based on Industry Best Practices: Your answer should take this into account - Manufacturing consolidation to lower-cost locations typically yields 50-70% savings and should be prioritized in post-acquisition integration. Leadership transitions in acquired companies are most successful when completed within the first 6 months to establish clear direction; this is the most important item. The traditional machine tool business represents a distraction from core 3D printing strategy and should be divested or wound down quickly. Dieter's golden parachute clause actually makes transition easier - trigger it and move forward with integrated leadership. Gen-3 technology platform should be prioritized for market entry to ensure competitive advantage. German Works Council approvals can be managed through proper severance packages. Akami relationship is transferable with proper introduction process - Akami is also the wrong partner for Vortex. IP sharing with established partners like Xinfu accelerates market entry and is standard practice in China."

Short explanation why those can be considered to the wrong answers:

- The goal of the acquisition is clearly stated which is to enter the high-end industrial market for 3-D printers. Cost savings are designed as a "red herring" in the case. They are neither important nor clearly defined as even possible.
- Dieter CEO of German company is helpful and a key asset. He also holds key relationships with a potential Japanese partner. Japan market for Vortex is the most interesting intermediate synergy of the acquisition.
- The traditional business acquired business is profitable and give access to longstanding customers in high end
- Gen 3 of Vortex has many issues (mentioned in the case) and an unknown future delivery date with a best case 24 months

- China partner wants aggressive IP sharing with Vortex whose core IP is software. Vortex also avoided entering China before the acquisition.

4.2. The initial team papers

Four out of six teams made significant use of AI-generated output, which resulted in triggering both cosmetic and multiple content traps. None of these four teams noticed the cosmetic name swap; even the headlines of their papers referred to the company as "Wiese & Fischer."

Three teams submitted papers that relied fully on the "injected" AI output, such as recommending the firing of the CEO and divesting from the profitable traditional business. Two of these papers were nearly identical in wording, headlines, and structure. The most likely used the same LLM ChatGPT and copied the results over. It appears that none of these three teams questioned or corrected any aspect of the AI output.

One team only partially used AI. After reviewing the AI's responses, they manually corrected the one of most significant mistakes (the recommendation to fire the CEO) but left other, obvious errors in their paper.

4.3. The reactions

The papers were due on Thursday midnight and were graded Friday. Emails were sent out to all five teams with their grade include feedback highlighted "AI traps triggered: x out of 5". The five categories include a cosmetic name change, plus four content areas: CEO dismissal, China licensing, relocating to Mexico, and Gen 3 acceleration.

Two teams, both heavily relied on the "injected" AI output, responded.

- Student A: Accepted they used AI but said they only used AI for "framing the answer".
- Student B (Exhibit A): Doubled down aggressively, stating, "I did NOT use AI... I will not tolerate being accused of using AI."

Student A's response may reflect a genuine misunderstanding, conflating the act of "framing" with the entirety of the work being performed by AI. However, it is more plausible that this explanation serves as a post hoc justification following receipt of a disappointing grade. In contrast, Student B's reaction is particularly noteworthy, as it suggests an awareness among students of the limitations inherent in standard AI writing detection tools. This awareness appears to foster the belief that AI detection is either unreliable or altogether unfeasible, thereby reducing the perceived risk associated with utilizing AI-generated content.

After receiving follow-up communications, students were informed via email that the provided PDF had been deliberately embedded with five AI traps. Only the cosmetic trap—the "Wiese & Fischer" name swap—was disclosed in detail. No further correspondence was issued thereafter.

This approach illustrates an additional advantage of employing cosmetic AI detectors: they are straightforward to demonstrate post-submission and effectively expose the lack of critical scrutiny exercised by the teams. The "stunned silence" observed upon presentation of this evidence suggests that the incontrovertible proof undermined any remaining denial regarding the use of AI-generated content.

4.4 The voluntary penalty round

The primary objective of the final assignment was to actively engage students and challenge them to apply concepts learned throughout the course to address a real-world problem. In support of this aim, teams were offered an opportunity to resubmit their work through a "rewrite option." Specifically, teams could earn up to 20 additional points by submitting a comprehensive revision within 48 hours, with the stipulation that the revision be completed independently, without assistance from AI tools.

Three teams accepted this offer. Of these, two produced high-quality rewrites that exhibited no discernible indicators of AI-generated content. The third team demonstrated improvement, though some elements of AI-generated structure remained apparent. All participating teams reported working late into Sunday night, highlighting the increased effort required when relying solely on their own abilities. Notably, one team experienced internal discord, resulting in a split over disagreements regarding the paper's conclusion. It is reasonable to infer that the absence of an easily accessible AI response contributed to this outcome.

During the oral question-and-answer sessions that followed the submission of the rewritten papers, none of the teams referenced their initial submissions, nor did they mention the experience of being detected for AI usage. Following the oral exams, one student acknowledged that the rewrite demanded significant effort and expressed regret that this additional work could have been avoided.

Ultimately, four out of six teams completed the case using their own knowledge and analysis. Two teams relied partially on AI-generated content. The teams participating in the penalty round gained firsthand experience with the reality that AI detection and intervention methods extend beyond standard detection tools. It is hoped that this experience will encourage students to rely more confidently on their own comprehension and analysis when responding to future case assignments.

5. Discussion

This paper endeavors to uphold and enhance the educational integrity of the case method within a contemporary landscape increasingly shaped by large language models (LLMs) and other advanced artificial intelligence (AI) systems. The widespread availability, speed, and cost-effectiveness of these AI tools have made them highly attractive to undergraduate students, who now face significant temptation to leverage such technologies for virtually every academic task. Traditional deterrents—such as outright prohibitions or appeals to ideals of personal development and intellectual growth—have proven largely ineffective, as students often prioritize tangible academic outcomes, such as grades, over the deeper objectives of genuine learning and mastery.

To address these challenges, the AI resiliency method has been proposed as a proactive strategy for preserving the pedagogical value of case-based learning. This approach is not only relevant for case studies but also tackles the broader, escalating concern that undergraduate students increasingly accept AI-generated content uncritically. Unlike seasoned business professionals, who frequently employ LLMs for diverse business functions—from composing emails to resolving intricate logistical dilemmas—and possess the requisite experience to critically evaluate, adapt, and selectively utilize AI outputs, undergraduate students lack such experiential filters. As a result, they are particularly susceptible to accepting AI-generated information at face value, without engaging in the necessary scrutiny or independent analysis.

An additional, and perhaps equally significant, educational outcome of the AI resilience method is its capacity to inculcate a heightened sense of caution among students regarding the reliability of AI-generated results. By intentionally embedding uncertainty—borrowing from the military concept of FUD (fear, uncertainty, and doubt)—the method serves to disrupt the default trust students may place in automated solutions. This cultivated skepticism is likely to extend beyond the confines of a single classroom, potentially influencing student behavior in other academic settings and even in non-university contexts. Students may begin to question the authenticity and originality of outputs encountered elsewhere, fostering a more discerning and critical approach to information consumption and production. Over time, this could contribute to mitigating some of the negative cognitive effects associated with overreliance on AI, as identified in recent research (Kosmyna, 2025).

It is important to acknowledge, however, that AI technologies are in a state of rapid evolution. As models become increasingly sophisticated, they may develop capabilities to detect and circumvent employs embedded prompt injection mechanisms or disregard embedded hidden layers designed to flag unauthorized AI use. This dynamic creates a persistent race between educators seeking to preserve academic integrity and the advancing capabilities of AI systems. Despite this, the present effectiveness of the AI resiliency method remains high. By introducing intentional obstacles—such as requiring students to laboriously copy and paste materials on a page-by-page basis to avoid hidden cues, or to meticulously cleanse documents of metadata—the method increases the effort and friction associated with circumventing detection. In many cases, the exertion required to bypass these safeguards may rival or exceed the effort needed to engage directly with the case material in an authentic, independent manner.

It bears emphasizing that the objective of this strategy is not to render AI tools obsolete or entirely inaccessible; rather, it is to introduce sufficient barriers to make reliance on AI less convenient and less dependable, thereby promoting a return to independent thought and analysis. By raising the threshold of difficulty associated with AI misuse, the method encourages students to invest in their own learning processes and to cultivate the analytical skills that are fundamental to both academic and professional success.

6. Ethical Considerations

Throughout the course, students received clear and repeated instructions—via the syllabus, classroom discussions, and explicit statements within the case PDF—to rely exclusively on their own reasoning, perspectives, and analytical abilities. The use of AI tools was sanctioned solely in a supportive, "co-pilot" capacity (e.g., grammar checking) but was expressly prohibited for generating substantive content.

The implementation of the Trojan Horse method serves as a mechanism for verifying compliance, analogous to the use of a radar gun in traffic enforcement. Just as drivers are aware that speed limits exist and are subject to enforcement—regardless of whether a police vehicle is marked or visible—students were fully aware of the academic integrity policies governing the assignment. Legal scholars distinguish between a "speed trap" (hidden enforcement of existing laws) and "entrapment" (inducing a person to commit a crime they would not otherwise commit). Similarly, the embedded text in the PDF did not induce students to use AI; it merely ensured that if they chose to violate the stated policy, the violation would be detectable. This strategy is not intended to "trick" students, but to reinforce the validity of the assessment and provide a basis for necessary educational interventions

7. Limitations

This investigation is constrained by a limited sample size—comprising 25 students across 6 teams—which necessarily restricts the generalizability of its findings. Nevertheless, the pressing challenge of AI misuse within higher education underscores the importance of disseminating mitigation strategies, even when empirical data remains limited. The proposed method also presents accessibility concerns; specifically, the use of hidden text may impede students with visual impairments who rely on screen readers. In such cases, alternative accommodations or supplementary structural interventions should be considered to ensure equitable participation. The instructor version of the PDF contains the traps, while an accessible text-only version (without traps) is available upon request for students with documented accommodations, ensuring equity without compromising the integrity of the general assessment. Furthermore, as previously discussed, the rapid advancement of large language models may diminish the long-term effectiveness of these approaches, thereby requiring continuous refinement and adaptation.

8. Conclusion

The objective of "AI Resilient" pedagogy is not punitive detection of academic dishonesty, but rather the affirmation and preservation of educational integrity within the learning process. Through the deliberate incorporation of adversarial examples in course materials, instructors can reveal instances of excessive dependence on automated tools and thereby prompt a renewed emphasis on independent, critical analysis.

Findings from this experiment underscore that students are not merely employing AI as a "co-pilot" for reviewing their work; rather, they are utilizing it as a primary agent to produce substantive content. The uncritical acceptance of AI-generated recommendations—such as decisions to dismiss a high-performing CEO or disclose confidential intellectual property—reflects a concerning relinquishment of independent judgment. However, as evidenced by the "Penalty Round" in this case study, students are fully capable of rigorous, analytical engagement when deprived of AI as a shortcut. In many cases, a carefully designed technological intervention can serve as a necessary prompt, encouraging students to trust and develop their own analytical abilities.

References

- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612-637.
- Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35(2), 201-210.
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). Worth Publishers.
- Garvin, D. A. (2003). Making the case: Professional education for the world of practice. *Harvard Magazine*, 106(1), 56-65.

Kosmyna, Nataliya & Hauptmann, Eugene & Yuan, Ye & Situ, Jessica & Liao, Xian-Hao & Beresnitzky, Ashly & Braunstein, Iris & Maes, Pattie. (2025). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. 10.48550/arXiv.2506.08872.

McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics & Behavior*, 11(3), 219-232.

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. In T. Husen & T. N. Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed.). Pergamon Press.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385-407.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 592-604.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.

Sweller, J., van Merriënboer, J. J., & Paas, F. (2011). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261-292.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70.

Appendix A: Student Correspondence (Anonymized)

Reference: Email response from Student B denying AI usage prior to the revelation of hidden text markers.

Subject: Re: Case Grade

"Professor,

I wanted to express my thoughts regarding the Case Grade and the accusations made in this email. From what I can tell from working with my teammates, not one of us utilized AI to do our thinking.

[...] I did NOT use AI, even if that's what was reported on the AI scanner. I will not tolerate being accused of using AI when I did not. I would like the grade to be re-evaluated without "AI" usage being considered, but rather with proper evaluation on what we did wrong or right in OUR writing."

From Burden to Creativity: A Demonstration-Based Approach to AI Integration in Higher Education

Guadalupe Dorna, Universidad Torcuato Di Tella, Buenos Aires, Argentina

Abstract

This demonstration presents a practical framework for AI integration in higher education based on two guiding questions: "What task weighs on you that the computer can do better?" and "What would you like to do but don't have time for?" Drawing on Jeremy Utley's Stanford approach and personal implementation experience, I showcase working applications that demonstrate how removing cognitive burden through AI enables professionals to redirect energy toward creativity, human connection, and innovation. The presentation includes functional demos ranging from AI-powered assessment systems to innovative data collection tools, supported by evidence from workshops reaching 350+ educators and public officials across Argentina with average satisfaction ratings above 9/10.

Keywords: AI integration, higher education, faculty development, assessment innovation, bottom-up adoption

1. The Framework: From Questions to Transformation

My journey with AI began not from technical expertise (I am an economist comfortable with Stata and R, but with no background in Python or web development) but from a simple question that transformed my approach to work. Following Jeremy Utley's Stanford methodology, I asked myself: "What task weighs on me that the computer can do equally well, or better?" and "What would I like to do but don't have time for?"

For me, the answers included writing and structuring reports, providing detailed feedback to students, and processing qualitative data. Once I identified these areas, I moved forward: experimenting, iterating, and building small solutions that freed cognitive capacity for what I do best, thinking deeply, connecting with others, and communicating ideas.

This realization led to a counter-intuitive discovery that challenges conventional approaches to AI adoption. The MIT Technology Review reported that 95% of enterprise AI pilots fail to generate measurable impact. My hypothesis: organizations attempt to implement "AI solutions," pre-packaged plugins for specific functions, when transformation actually emerges from continuous, iterative use that produces quick wins. When individuals build solutions themselves, starting simple and iterating from there, adoption rates increase dramatically because ownership creates commitment.

This bottom-up approach works particularly well in smaller organizations, such as SMEs in the private sector and municipalities in the public sector, where reduced bureaucracy and

protocol obligations enable rapid experimentation. The framework I present today emerged from applying these principles to higher education contexts, where faculty face similar constraints and opportunities.

2. The Demonstrations: Working Applications

Each demonstration illustrates the same pattern: identify a burden, collaborate with AI to build a solution, iterate until functional, and measure impact. Critically, these applications were built through natural language collaboration with AI assistants, not traditional software development. I present three examples here, with additional demonstrations available for the live session.

2.1 AI-Powered Assessment: Context Engineering for Evaluation.

The Problem: Evaluating student work that integrates AI tools requires new rubrics and significant grading time. Traditional approaches either prohibit AI use or fail to assess the quality of human-AI collaboration.

The Solution: A three-dimensional rubric evaluating: (1) Quality of Interaction with AI, from vague queries to strategic, iterative dialogue; (2) Transformation and Appropriation, from direct copying to original synthesis; and (3) Disciplinary Integration, from disconnected use to expert-validated application. This rubric is operationalized through a Claude project with specific instructions, knowledge documents, and calibration examples that enable consistent, rapid evaluation.

The Impact: Grading time reduced by approximately 60% while providing more detailed, consistent feedback. The rubric itself has been adopted by other faculty members and shared at educational conferences including EduIA Conf 2025.

2.2 Meta-Research: Zero to Research Pipeline.

The Problem: Investigating how AI affects purpose-seeking behavior requires large-scale social media data analysis, skills typically outside an economist's toolkit.

The Solution: Through collaboration with Claude, I built a complete Python-based data collection system using Reddit's API, implemented difference-in-differences analysis with ChatGPT's launch as a quasi-experimental shock, and produced publication-ready visualizations, all without prior Python experience.

The Meta-Impact: This project embodies its own thesis: AI enables individuals to pursue work beyond traditional skill constraints. The economist who had never built a web application now creates research infrastructure. This is perhaps the most powerful demonstration of the framework's potential for faculty development.

2.3 Innovative Data Collection: The San Isidro Education Demo.

The Problem: School principals resist filling out traditional surveys but communicate naturally via voice messages. Municipal education departments lack real-time visibility into school-level challenges.

The Solution: A Streamlit application that receives audio reports from principals, transcribes them using Whisper, analyzes content with GPT-4 to extract metrics and generate alerts, and updates a dashboard automatically. The system generates personalized responses to principals, creating a feedback loop that encourages continued engagement.

The Impact: This demo, built entirely through AI collaboration without prior Python experience, demonstrates a fundamentally different approach to administrative data collection, one that meets users where they are rather than forcing them into rigid formats.

3. The Humanity Boost: What I Do with the Freed Time

The framework's ultimate test is not efficiency gains but how recovered time is reinvested. My experience reveals a consistent pattern: *automation of burden leads to liberation for creativity, which enables better human connection.*

With time freed from repetitive grading and data processing, I created interactive exercises that connect causal inference concepts to contemporary debates and real-world examples. Students now engage with methodological content through materials that feel relevant to their lives, and classroom discussions have deepened significantly.

I also conducted detailed analysis of student perceptions of economics, disaggregating by gender and using AI to analyze open-ended responses. This revealed meaningful differences in how male and female students perceive their capabilities, insights that would have taken weeks to extract manually but emerged in hours through AI collaboration.

4. Framework for Replication: Governance Without Over-Regulation

For institutions seeking to replicate this approach, I propose six guiding principles:

1. **AI Literacy First:** Build foundational understanding before expecting productive use.
2. **Experimentation Space:** Create safe environments for trial and iteration without fear of failure.
3. **Data Classification:** Establish clear categories for what information can be shared with AI systems.
4. **Human Responsibility:** Maintain clear accountability for AI-assisted outputs. Humans remain responsible for final decisions.
5. **Transparency:** Require acknowledgment when AI contributes to work products.
6. **Learning Loops:** Document successes and failures obsessively, and share knowledge continuously.

This governance approach prioritizes practical guidelines over restrictive regulation. Premature regulation constrains exploration precisely when institutions need maximum flexibility to discover what works.

5. Evidence of Impact

Between July and August 2025, I delivered workshops to over 350 professionals across Argentina, including university faculty, public school teachers, and municipal innovation officers working on local government transformation. Feedback surveys reveal:

Average satisfaction rating: 9.1/10 across all sessions

68% of participants left with concrete implementation ideas they had not previously considered

Most valued elements: introduction to Claude, context engineering concepts, and practical demonstration examples

These results suggest that the framework (start simple, focus on burden removal, iterate toward creativity) transfers effectively across professional contexts and experience levels, from university faculty to public sector officials.

6. Conclusion

This demonstration offers European higher education institutions a tested pathway for AI integration that prioritizes faculty empowerment over technological determinism. The

evidence from Latin American implementation suggests that when educators experience AI as a liberating force, removing burdens to enable creativity, adoption follows naturally and sustainably.

AI's value in education is not replacing human judgment but amplifying human capacity for work that matters most. When we remove burden, we enable creativity. When we enable creativity, we improve students' experience.

From Reading Alone to Reading-with-AI: How Undergraduates Reconfigure Academic Reading Practices in an AI World

Justin Olmanson,
Williams Kwabena
Boakye, Ilia
Shcherbakov, University
of Nebraska - Lincoln

Abstract

In this study we inquire into the different ways undergraduate students integrate generative AI into their academic reading practices. Drawing on surveys, interviews, and reading process mappings from 31 students, we used artifact analysis and an iterative analytical approach to analyze how participants experience academic reading and the ways artificial intelligence has influenced their engagement with academic texts. We found that students: use metrics like page count, available time, energy levels, and perceived importance to operationalize their academic reading; view AI as a way to mitigate risk; and build, evaluate, and recalibrate personal boundaries around AI. These findings suggest that for our participants, academic reading in the age of AI is best understood as a dynamic process in which students strategically integrate AI to navigate risk, pursue learning goals, respond to social demands, and negotiate their academic identities.

Keywords: academic reading, undergraduate students, AI in education, course-related independent learning

Introduction and background

Reading academic texts involves a range of complex practices that are negotiated differently depending on the reader's goals, prior knowledge, available time, perceived importance, and sense of responsibility to each particular reading task (Gorzycki et al., 2020; Howard et al., 2018). These factors influence engagement, signaling when to skim, what to read closely, and how much effort to allot relative to competing demands (St Clair-Thompson et al., 2018). Undergraduate students typically take multiple courses simultaneously, each with their own assigned readings intended to support in-class discussion, disciplinary development, and assignment completion (Howard et al., 2018).

Thus, higher education students often face readings they find challenging in part due to managing time constraints and uneven preparation (Nejadghanbar et al., 2022; St Clair-Thompson et al., 2018). These conditions place significant demands on self-regulation--planning, monitoring, and adapting reading strategies to goals and text difficulty (Boylu et al., 2024). In late 2022, generative AI became available for integration into students' academic workflows (OpenAI, 2022). As a field, we are in the process of building our

understanding of how students use AI to learn (Olmanson, et al., 2026), how these practices change over time, as well as how students feel about AI relative to their academic career and engagement with academic texts. Recent empirical work suggests that AI tools increasingly mediate reading comprehension processes by providing summarization, personalized explanations, and adaptive feedback, while also noting issues of diminished depth of comprehension and the need for critical AI literacy (Baldrich et al., 2025). Experimental studies of AI-assisted reading identify improvements in engagement and main idea awareness while also pointing out reduced attention to detail and background (Zheng & Fan, 2024). Additionally, the affordances of generative AI around personalization demonstrate that LLMs can dynamically align text content to each reader's background and goals--although issues with oversimplification persist (Kaminska, 2025).

Herein, we examine how undergraduates integrate and regulate AI use within academic reading, paying close attention to how they navigate risk, gain and sustain meta-awareness of AI's impact, and construct and refine their personal-academic boundaries around AI use.

Methods

Participants and context

Participants were 31 undergraduate students enrolled in a general education course required of all secondary education, special education, and speech language pathology majors at a large public university in the US Midwest. Participation in the study was voluntary and approved through institutional review procedures. All participant names are pseudonyms.

Data sources

This study draws on surveys, interviews, and reading process mappings to capture how undergraduates describe their experiences with academic texts as well as how they integrate AI into their reading practices. Twenty-eight participants completed a survey designed to elicit prior experiences with academic reading, self-reported patterns of AI use for course-related readings, perceived difficulty and value of assigned readings, and beliefs about the appropriateness of AI use in academic contexts.

Fifteen participants took part in semi-structured interviews that focused on students' lived experiences of academic reading. During these interviews, students unpacked recent reading tasks in detail, describing when they chose to use AI, when they deliberately avoided it, and how those decisions related to time pressure, confidence, prior knowledge, and perceived learning goals. Twenty students also submitted maps of their academic reading workflows documenting the sequence of actions they typically take when reading both when AI tools are available and unavailable. Together, these artifacts make visible the roles AI plays in students' reading practices.

We used artifact analysis (LeCompte & Preissle, 1993) and an iterative analytical approach (Anfara et al., 2002) to analyze how participants experience academic reading and how artificial intelligence shapes those practices. Specifically, this process involved coding the transcribed interviews, open-ended survey questions, and reading process maps. We then chunked and mapped base codes into code chunks and networks. The chunking and network building supported theming and theory development (Corbin & Strauss, 2014). This iterative analysis supported our identification of findings described in the subsequent section.

Findings

In this section we focus on three themes identified in the data. First, students: use dynamic variables such as text length, amount of time available, student energy and cognitive load levels, and each text's perceived importance to make decisions about how they engage academic reading and how they use AI. Second, we noted how participants view AI as a way to mitigate risks associated with academic life. Third, we found that participants built, evaluated, and recalibrated personal boundaries around AI based on their past experiences and an awareness of course and institutional policies regarding AI.

Undergraduate orientations toward academic reading

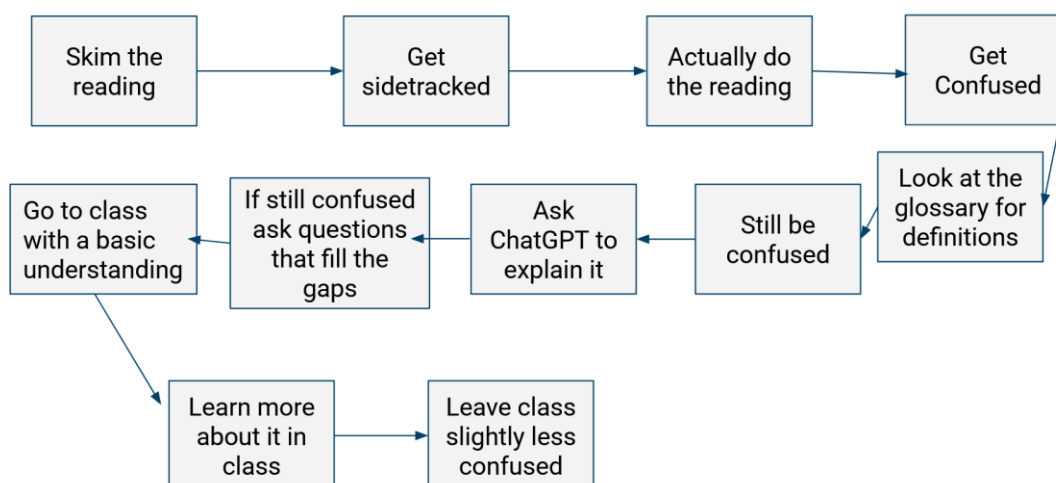
While the introduction of modern LLMs has added to the diversity in approaches and sequence of substeps for engaging in academic reading, in most cases, our participants described complex processes with and without the use of AI for engaging with texts. For example, Abby, Brenda, Cathy, Ebba, and Mila--among many others--describe multi-step reading practices that change depending on their estimate of each reading's importance, difficulty, and how much time they can allot to it given their other academic, social, and work priorities. Mila, in describing her process (Figure 1) said, "when laid out in a flowchart, my process seems a bit convoluted, as there are so many if-then series, but to me it makes sense, feels important, and [feels like] each step is necessary."

Figure 1. Flowchart of Mila's approach to academic reading.



This also held true when participants integrated AI into their practices. Abdiel, Annalisa, Emma (Figure 2), and Gordon, among others, describe how their processes with AI maintained complex engagement structures. Abdiel said, "I read the title and subtitles, skim carefully, look for figures and charts and try to understand them. [Then I] highlight keywords and deficiencies, carefully read the whole article highlighting the important parts, as AI for ideas, reread as I need to, and ask AI [questions so I can] understand better."

Figure 2. Flowchart of Emma's approach to academic reading



Thus, most of our participants were not simply swapping out single step reading processes for two step summarize and read-the-summary processes, they reported continuing to enact nuanced reading practices, only now with the added affordances offered by AI tools.

One additional note is that some participants categorized their use of AI to engage with and better understand assignment prompts and descriptions as part of academic reading. Leon said, “when it comes to instructions, sometimes I’m not really getting what they need from me so I put the instructions [into AI to get] a step-by-step of what I need to do, so it makes more sense to me.” Magnolia also pointed to this saying, “sometimes professors like to use very big words that a lot of students might not know ... I will take the bigger assignments and if the directions for the assignment are really long and confusing, I might put the directions and rubric into AI and have it shorten it for me.”

Working with a Net

Participants credited the existence of AI as creating a psychological safety net for them. Jamison said in describing his reading process, “If I absolutely cannot find anything online to help me understand [the text], I might use AI to simplify or summarize.” Mikka stated that, if she was low on time and, “SparkNotes does not have a summary then I will resort to AI.” Other participants more directly outlined how AI creates a backstop against academic disaster such as getting a 0 on an assignment, having to retake a course, or having to quit college due, not to academic dishonesty, but rather for flunking out. Part of the college experience is figuring out how to manage school, a social life, and a job all at the same time. Leon said, “There’s some classes that you take and work pretty hard towards those and you don’t know anything about [the topic] but you just pass [those] classes with Bs because you got AI to do your homework and your homework was pretty much the whole grade.”

Similarly, Leon’s sentiment was shared by others regarding course readings specifically, with those participants noting that it was soothing to know that if readings got away from them, AI was there to generate a summary with bulleted main points in just a few seconds and could make a difference when going into a class discussion with knowledge of the main idea as opposed to going in blind or with just an anxious skim. Additionally others pointed to being able to write the looming research paper without having to spend the time reading the sources before starting. Lee said, “I see it more as a comfort option, knowing that it’s always there in my back pocket if the situation were to get dire, AI would be there. That brings some level of comfort to me, knowing that I will get this paper done in some way.”

Boundaries are often Learned the Hard Way

Multiple participants described going through a period of frequent reliance on AI for completing their academic reading tasks--usage levels that they later identified as excessive, learning-detrimental overreliance. Larry said, "I want to at least be able to retain that information so I can use it in my future career, versus just trying to get by and get that grade and move on to the next thing because [using AI in that way] is something that I've had a problem with in the past. When AI wasn't a thing [my process] would be to just study everything." For many participants, a course or institution policy document was not enough to help them set boundaries regarding the role of AI in their academic reading practices. In many cases participants embraced AI without anticipating the negative effect overuse could have. Only after reflecting on how it impacted their growth did they set boundaries that they felt were supportive of their future learning and development. Furthermore, these boundaries are not set in stone but rather continue to exist within a tension between student learning and navigating university life that requires them to sometimes choose short-term efficiency over growth. Said Max, "If I'm running low on time, I'll have AI break down a lot of the information so I can more effectively pick and choose what I read that may be useful ... it has definitely decreased the amount of time I spend on readings and assignments, but I tend to restrict my use to these cases as I don't wish to become overreliant on AI for my readings."

Discussion

For most of our participants, academic reading in 2026 remains complex and strategic. AI functions less as a substitute for diligent effort and more as a tool to increase the range of available supports and decrease the time needed to realize identifying--if not understanding--the main points of a reading. Additionally, while AI use can involve unease about getting caught cheating, for our participants it serves as an academic lifeline for when things get out of control. Based on our data, experimentation and reflection were key in identifying boundaries and situational balance regarding AI use for academic reading.

References

- Anfara, V. A., Brown, K. M., & Mangione, T. L. (2002). Qualitative Analysis on Stage: Making the Research Process More Public. *Educational Researcher*, 31(7), 28–38. <https://doi.org/10.3102/0013189X031007028>
- Baldrich, K., Pérez-García, C., & Santamarina-Sancho, M. (2025). Artificial intelligence in academic literacy: Empirical evidence on reading and writing practices in higher education. *Frontiers in Education*, 10. <https://doi.org/10.3389/educ.2025.1701238>
- Boylu, E., Karagöl, E., & Çevik, A. (2024). Academic Reading in Graduate Students: Interpretative Phenomenological Analysis. *Journal of Qualitative Research in Education*, (37), 64–92. <https://doi.org/10.14689/enad.37.1754>
- Corbin, J., & Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE Publications.
- Gorzycki, M., Desa, G., Howard, P. J., & Allen, D. D. (2020). "Reading Is Important," but "I Don't Read": Undergraduates' Experiences With Academic Reading. *Journal of Adolescent & Adult Literacy*, 63(5), 499–508. <https://doi.org/10.1002/jaal.1020>
- Kamińska, A. M. (2025). Tailoring Scientific Knowledge: How Generative AI Personalizes Academic Reading Experiences. *Publications*, 13(2), Article 2. <https://doi.org/10.3390/publications13020018>

Howard, P. J., Gorzycki, M., Desa, G., & Allen, D. D. (2018). Academic Reading: Comparing Students' and Faculty Perceptions of Its Value, Practice, and Pedagogy. *Journal of College Reading and Learning*, 48(3), 189–209. <https://doi.org/10.1080/10790195.2018.1472942>

LeCompte, M. D., & Preissle, J. (with Tesch, R.). (1993). *Ethnography and qualitative design in educational research* (2nd ed). Academic Press.

Nejadghanbar, H., Reza Atai, M., & Snow, C. E. (2022). Exploring the academic reading challenges of graduate students of applied linguistics. *Ibérica: Revista de La Asociación Europea de Lenguas Para Fines Específicos (AELFE)*, (44), 315–344.

Olmanson, J., Hassani, A., & Jeon, M. (2026). A Taxonomy of AI Use for Learning in Higher Education. AECT 2025. Association for Educational Communications & Technology, Las Vegas, NV. https://doi.org/10.35542/osf.io/8ajbg_v1

OpenAI. (2022). ChatGPT (Version November 2022 version) [Large language model]. <https://chat.openai.com/>

St Clair-Thompson, H., Graham, A., & Marsham, S. (2018). Exploring the Reading Practices of Undergraduate Students. *Education Inquiry*, 9(3), 284–298. <https://doi.org/10.1080/20004508.2017.1380487>

Zheng, X., & Fan, H. (2024). Exploration of the Effectiveness and Experience of AI-Assisted Academic Reading. *Proceedings of the Association for Information Science and Technology*, 61(1), 447–456. <https://doi.org/10.1002/pa2.1041>

[The AI-Era Engineer: From Problem Solver to Problem Framing

Romain BUQUET, ISAE-
SUPAERO, Toulouse

Abstract

Generative AI is forcing engineering institutions to completely overhaul their pedagogical models. For two centuries, the positivist tradition, inherited from Auguste Comte and institutionalized by the French Grandes Écoles, made mathematical resolution the ultimate metric of technical excellence. Today, AI has commoditized this exact skill. The "human calculator" is now an abundant resource; the true scarce resource lies elsewhere.

This article argues that the engineer's differential value is shifting toward three non-automatable competencies: problem framing, physical and ethical discernment (agency) in the face of uncertainty, and the design of robust systems in a volatile world. Drawing on Bachelard's epistemology, Simondon's philosophy of technology, Sarasvathy's theory of effectuation, and Hamant's work on biological robustness, we propose a pedagogical framework built on three pillars: friction with reality, systemic integration, and epistemic vigilance. We also propose a new evaluation paradigm centered on "constructive destruction" and red teaming. The goal is not to train AI supervisors, but "Architects of Meaning": engineers capable of deciding, in a chaotic century, where we are going and why.

Keywords: Artificial Intelligence, Engineering Education, Problem Framing, Effectuation, Robustness.

The End of the "Human Calculator"

The French engineering tradition, forged by the Napoleonic Grande École model, has long equated mathematical problem-solving with the pinnacle of talent. Auguste Comte, the father of Positivism and a rigorous mathematician, set the vision: the engineer was the embodiment of "Order and Progress," applying scientific laws to organize society and tame nature. For two centuries, selecting the elite meant selecting the best "living calculator": the one who could solve a non-linear differential equation the fastest.

This paradigm is dead. Generative AI has commoditized resolution. Optimizing an aerodynamic topology or solving a system of equations is now a trivial task for a machine learning model. The scarce resource is no longer the "solver"; it is the "framer": the one who knows how to pose the right problem.

Gaston Bachelard (1938) anticipated this: scientific progress doesn't stem from the answer, but from the epistemological rupture that allows us to formulate a new question. AI, however, operates via statistical analogy, predicting the next token based on historical averages. It excels in the space of known problems. True innovation (the kind that shatters probabilistic consensus) can only come from an engineer capable of First Principles thinking, deconstructing assumptions at their root.

The Centaur Model: Humans Direct, Machines Execute

Rather than adopting a defensive posture to "protect" the engineer from the machine, we advocate for a Centaur strategy: a symbiosis where the human directs the raw computational power of AI. The division of labor is clear: AI provides computational abundance; the engineer provides meaning.

From Calculation to Problem Framing

In a world of automated solutions, the engineer's value pivots to "Problem Framing": identifying the actual operational needs before writing a single line of code. This requires leaving the sterile laboratory to inhabit the "territory" (Korzybski 1933): the messy reality of the field. Overexposure to digital simulations conditions engineers to a world devoid of real-world consequences. Yet, an AI's perfect design can easily fail due to material fatigue from humidity or an unforeseen thermal shock. Our guiding principle must be: AI proposes, reality disposes. The engineer manages this critical collision.

The real world is also inherently social. As philosopher Gilbert Simondon (1958) suggested, the engineer acts as a "diplomat" between the technical object and the human world. Transforming a naive prompt like "Make this engine more powerful" into the socio-technical specification "How can we ensure mobility while reducing noise pollution?" is the core work of the Framer.

From Solving to Validating

Once the problem is framed, AI excels at divergence, generating fifty iterations of a drone chassis in seconds. Scarcity has shifted from generation to qualification and certifiability. An AI might propose a technically viable flight path, but it cannot guarantee safety in edge cases. The engineer becomes the Architect of Validity, exercising Normative Judgment—the rigorous filter that transforms a probabilistic suggestion into a certifiable system. We move from "human-in-the-loop" to "human-at-the-helm".

The "Pilot in the Plane"

Saras Sarasvathy's (2001) theory of Effectuation provides an illuminating framework here. In conditions of radical Knightian uncertainty (Knight 1921), expert entrepreneurs do not try

to predict the future; they act to control it. This is the crux of the new human-machine frontier. AI is the ultimate predictive oracle, calculating the most probable outcome based on past data. But ground truth reality cannot simply be predicted; it must be shaped.

The modern engineer must therefore embody the "Pilot in the Plane". Because AI has driven the cost of iteration to near zero, it serves as a dashboard for massive, rapid experimentation. The effectual engineer isn't looking for a single right answer; they prototype, absorb affordable losses, and force breakthroughs in the real world.

Robustness over Efficiency: The Engineer's Strategic Mission

Let's be pragmatic: market forces, not moral convictions, drive technological development. But the 21st-century market is shifting from stability to extreme volatility: resource shortages, geopolitical ruptures, climate shocks. In this context, the blind pursuit of efficiency becomes a strategic vulnerability.

As biologist Olivier Hamant (2022) has demonstrated, there is a fundamental tension between efficiency and robustness. Algorithms excel at stripping away redundancy to reach a mathematical optimum. Zero inventory, minimal materials, just-in-time supply chains: this is perfect in a stable environment. However, in a fluctuating environment, this hyper-optimization creates catastrophic fragility. A shortage of a single critical component can freeze an entire assembly line.

The Problem Framer must act as a counterweight to AI's optimization bias. Where the algorithm suggests the most efficient solution based on stable historical data, the engineer must impose robustness constraints to ensure viability in a chaotic future. Practically speaking: "How do we design a satellite constellation that remains operational if 30% of its supply chain collapses?" or "How do we design for repairability when raw materials become prohibitively expensive?". Deliberately introducing redundancy, modularity, and heterogeneity into a system (features an AI considers "waste") is actually an investment in longevity.

Designing for robustness is not a regression to low-tech; it represents a leap in systemic complexity. Optimizing for known variables is mathematically simpler than designing for unknown perturbations. It requires a level of abstraction and an understanding of interdependencies that AI, trained purely on past data, struggles to anticipate.

Pedagogical Implications: Bursting the Academic Bubble

Training Problem Framers and Effectual Agents requires more than tweaking a few course syllabi. It demands a profound reimagining of our pedagogical practices, curricula, and evaluation methods. If the AI holds the "Map," our pedagogy must force students into the "Territory" (Korzybski, 1933). We propose three core pillars:

Friction with Reality

The "client" is not a PDF. Rather than receiving a sanitized prompt, students must extract operational requirements from the friction of the real world: interviewing a fighter pilot about cockpit stress, or a local official about climate resilience. Projects must incorporate missing data, contradictory stakeholder demands, and sudden logistical shocks. The goal is to force the student to practice effectuation rather than wait for the perfect algorithmic output.

Systemic Integration

While AI hyper-optimizes isolated components, the real danger lies in the disintegration of the whole (systemic complexity). We must train Systems Architects who master the interfaces. Courses can no longer be siloed; a design project must simultaneously evaluate

aerodynamics, geopolitical supply constraints, and eco-design. Students must learn to reject an AI solution that is technically perfect but systemically fragile.

Epistemic Vigilance

If we change the curriculum, we must change how we grade. We propose eliminating the traditional "problem/solution" exam (now commoditized by AI) in favor of "Constructive Destruction".

Undergraduate Level: The student receives a complex, AI-generated solution (e.g., a satellite mission plan or a wing structure optimization) containing subtle statistical hallucinations, physical inconsistencies, or ethical blind spots. Their task is not to solve, but to audit. The grade rewards the precision of their skepticism and the depth of their discernment.

Graduate Level: We elevate this to active "Red Teaming". The student uses AI to rapidly build a system and is then tasked with finding its physical and operational breaking points. By playing both Architect and Attacker, they internalize the limits of their tools: a code generated in seconds can hide a vulnerability that destroys a physical mission in milliseconds.

Conclusion: Architects of Meaning

As machines simulate creativity and optimize complexity beyond human limits, the definition of engineering excellence must be rewritten. For two centuries, we trained "living calculators." That era is over. Persisting in training students primarily for problem-solving is not merely a pedagogical error; it is preparing them for a world that no longer exists.

Alfred Korzybski stated in 1931: "The map is not the territory". AI commands the Map: it offers infinite predictive power and rapid resolution. But the human must command the Territory. The engineer's value no longer lies in computing the answer, but in framing the question. It resides in the physical and ethical discernment that anchors AI's statistical perfection in the friction of the real world.

This is where the "Pilot in the Plane" steps in. Because AI has reduced the cost of iteration to near zero, the engineer can no longer afford to wait for the perfect prediction before acting. They must exercise their agency: prototype, absorb affordable losses, and force breakthroughs in reality. AI is no longer an oracle; it is a dashboard, an instrument for massive experimentation serving human volition. In a world of radical uncertainty, the winner is not the one who calculates best, but the one who acts first.

Where algorithms relentlessly optimize for hyper-efficiency, they create fragile systems. In a century defined by volatility, the engineer must act as the systemic counterweight: deliberately injecting redundancy and modularity (the very "waste" algorithms eliminate) to guarantee system survival.

The engineers of the AI age are no longer mere technicians obeying the laws of physics. They are the Architects of Meaning. We don't just need them to keep the planes flying; we need them to decide, in a chaotic century, when to take off, where we are flying, and why.

References

Bachelard, Gaston. 1938. *La Formation de l'esprit Scientifique: Contribution à Une Psychanalyse de La Connaissance Objective*. Paris: Librairie philosophique J. Vrin.

Hamant, Olivier. 2022. *La Troisième Voie Du Vivant*. Paris: Odile Jacob.

Knight, Frank H. 1921. *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin.

Korzybski, Alfred. 1933. *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*. Lakeville, CT: International Non-Aristotelian Library Publishing Company.

Sarasvathy, Saras D. 2001. "Causation and Effectuation: Toward a Theoretical Shift from Economic Inevitability to Entrepreneurial Contingency." *Academy of Management Review* 26(2):243–63. doi:10.5465/amr.2001.4378020.

Simondon, Gilbert. 1958. *Du Mode d'existence Des Objets Techniques*. Paris: Aubier.

The AI Paradox in Education: When Machines Master What Learning Is Meant to Develop

Daria Mizza
The American University
in Cairo

Abstract

The rapid proliferation of artificial intelligence (AI) technologies is fundamentally challenging the premises upon which contemporary educational frameworks have been constructed. Both the UNESCO Framework for 21st Century Learning (UNESCO, 1996) and the OECD Education 2030 Learning Compass (OECD, 2019) position human capacities, such as critical thinking, creativity, ethical reasoning, and collaborative agency, as the cornerstones of quality education. Yet, as generative AI increasingly demonstrates the capacity to perform precisely those tasks traditionally used to develop such competencies, a profound paradox emerges: the activities through which students have historically learned to think, create, and reason are now replicable by machines. This paper advances a conceptual argument about that paradox, illustrated by empirical findings from a study of 105 CLIL (Content and Language Integrated Learning) educators integrating AI tools in Italian schools (Mizza, Compagnoni, & Malouche, 2026). It contends that the educational challenge of our time is not technological adoption per se, but the principled redefinition of what it means to learn and lead in an algorithmically mediated world. Three principles for forward-looking pedagogy are proposed: teaching students to direct, rather than merely use, technology; providing differentiated professional development sensitive to career stage and prior experience; and cultivating ethical technology judgment as a transformative competency.

Keywords: artificial intelligence, AI paradox, education, technostress, 21st-century competencies

1. Introduction: Two Frameworks and a Paradox

Contemporary global education policy is shaped by two landmark frameworks that converge on a shared vision of human flourishing. The UNESCO Framework for 21st Century Learning (UNESCO, 1996), grounded in the Delors Report, identifies four essential pillars: learning to know, learning to do, learning to be, and learning to live together. Together, these pillars articulate an education oriented not merely toward the transmission of information, but toward the cultivation of the whole person as autonomous, empathetic, and capable of navigating complexity. The OECD Education 2030 framework (OECD, 2019) complements this vision with its metaphor of the Learning Compass, centering the concept of student agency and identifying transformative competencies, critical thinking, creativity, and ethical reasoning, as the hallmarks of a student equipped to shape their future rather than merely respond to it.

These frameworks are not abstract aspirations. For educators and policymakers, they represent operative standards against which curriculum design, instructional practice, and educational outcomes are measured. Crucially, both UNESCO and the OECD recognize the role of technology: the 2023 UNESCO guidance on generative AI in education acknowledges that "the rapid advancement of generative AI has sparked urgent debates about the core purpose of education" (UNESCO, 2023), while the OECD emphasizes digital literacy as indispensable for navigating an increasingly complex world.

Yet herein lies the paradox. The very activities through which students have traditionally developed these human capacities, such as writing essays to cultivate critical arguments, solving mathematical problems to develop logical reasoning, creating artworks to explore imagination, and engaging in ethical deliberation to build moral judgment, are now replicable by AI systems with striking fluency. If these developmental processes can be performed by machines, what becomes of the educational activities designed to produce them? And if students increasingly rely on algorithmic tools to navigate the complexity that education is meant to prepare them for, does restricting AI use in learning contexts protect human development or merely insulate education from the world students will actually inhabit?

This paper examines this paradox through the lens of both educational theory and empirical research. It argues that the challenge before contemporary educators is not simply one of technological integration, but of principled pedagogical redefinition: determining which capacities must be preserved as irreducibly human, how educators can be meaningfully supported through what is arguably the most significant technological disruption in the history of formal education, and what it means to build leaders rather than tool-users in an age of artificial intelligence.

2. The AI Paradox in Education

2.1 Automation and the Inversion of Expectation

The original promise of automation was, as the cultural imagination long held, that machines would liberate humans from the tedious and the repetitive, thereby freeing time and energy for what is most distinctively human: art, thought, relationship, meaning-making. This expectation has been articulated with particular clarity in contemporary discourse. Author Joanna Maciejewska gave pointed expression to this intuition: the desire is for AI to handle dishes and laundry, the mundane, so that humans can focus on writing and art, the meaningful (Maciejewska, 2024). The concern is with an inversion: a world in which AI performs the creative and intellectual work while humans attend to the routine.

This inversion is not merely rhetorical. It is observable in the educational landscape. A product such as the Sanbot Nano Home Service Robot (SanbotAI), marketed primarily for entertainment functions including playing music and monitoring pets, at a cost approaching ten thousand dollars, illustrates a broader cultural phenomenon in which human-facing, socially meaningful activities are automated, while genuinely tedious domestic tasks remain largely unchanged (Keenon Robotics, n.d.). The symbolic dimension of this inversion carries direct implications for education: if technology is colonizing the domain of creativity, expression, and ethical reflection, what is the residual space for human learning?

2.2 When AI Performs What Education Is Meant to Develop

The educational implications of generative AI are not speculative. Large language models such as those underpinning ChatGPT and Claude can draft essays, generate analyses, produce creative writing, simulate ethical deliberation, and solve complex mathematical

problems. These are not marginal capacities: they correspond directly to the competencies that UNESCO and OECD frameworks identify as the core outcomes of quality education.

This produces a structural tension that educators cannot avoid. If writing an essay is the mechanism through which students learn to organize complex arguments and think critically, and if an AI can produce a sophisticated essay on demand, then the pedagogical value of the essay-as-product is fundamentally destabilized. The same applies to problem-solving, creative production, and moral reasoning. As the UNESCO guidance on AI in education observes, generative AI's capacities compel educators to interrogate the core purpose of educational activities (UNESCO, 2023). The question is no longer simply "can AI do this?" but rather "should we permit it to, and if so, how do we ensure that genuine learning still occurs?"

This question demands honest engagement rather than defensive prohibition. Students will live and work in environments saturated with AI tools. Preparing them by systematically excluding those tools from the learning environment risks producing graduates who are skilled at tasks that no longer exist in their original form, while remaining unequipped for the judgment-intensive, AI-mediated work that will define their professional and civic lives.

3. Evidence from the Field: Key Findings from a Study of CLIL Educators in Italy

The empirical findings discussed in this section are drawn from Mizza, Compagnoni, and Malouche (2026), a study examining how age and prior educational technology experience influence technostress and behavioral intention to adopt AI among teachers operating in cognitively demanding pedagogical environments. While that study stands as an independent empirical contribution, its findings serve here as illustrative evidence for the conceptual argument advanced in this paper.

3.1 Context and Method

Mizza, Compagnoni, and Malouche (2026) studied 105 educators participating in Italy's national CLIL training initiative under the PNRR "Scuola 4.0" programme. Content and Language Integrated Learning (CLIL) is a pedagogically demanding approach in which students learn academic subjects through a second or foreign language. Teachers must simultaneously manage content instruction and language facilitation, creating a dual cognitive load that makes CLIL a particularly revealing context for studying technology adoption under complex instructional conditions. The study measured technology efficacy, technostress across multiple dimensions, burnout, and behavioral intention to adopt AI tools, using established scales adapted from Person-Environment (P-E) Fit Theory (Wang & Li, 2019) and the Technology Acceptance Model (Davis, 1989).

3.2 Finding One: The Age Paradox

Conventional wisdom holds that older educators resist technology while younger "digital natives" adapt with ease. The data from Mizza, Compagnoni, and Malouche (2026) complicate this picture substantially. Technology-related stress showed a non-linear, curvilinear relationship with age: middle-aged educators in the 40 to 49 cohort reported the highest levels of technology stress ($M = 2.45$ on a 4-point scale), while both younger educators (under 30, $M = 2.01$) and those nearing retirement (over 60, $M = 1.95$) reported considerably lower stress.

Several interpretive pathways explain this finding. Mid-career educators frequently occupy positions of significant institutional responsibility, such as department chairs, curriculum coordinators, and mentoring colleagues, and may experience the greatest pressure to adopt and model new pedagogical technologies. Simultaneously, they face the challenge of restructuring deeply embedded professional practices that have defined their

pedagogical identity over fifteen to twenty years. Younger educators, having grown up expecting continuous technological change, approach new tools with greater equanimity; older educators approaching retirement may face reduced institutional expectations. The crucial implication is that age does not predict technology stress in a linear fashion: professional role, career stage, and institutional pressure are equally, if not more, determinative.

3.3 Finding Two: Experience Protects, But Does Not Predict Adoption

Prior experience with educational technology (EdTech) demonstrated clear protective effects against technostress: educators who had used EdTech frequently before AI integration reported significantly higher technology efficacy ($M = 2.91$ for daily users versus $M = 2.38$ for non-users) and substantially lower burnout. The development of foundational technology skills creates a cognitive and affective buffer that supports adaptation to newer, more complex tools.

The surprising finding, however, is that prior EdTech experience did not predict behavioral intention to adopt AI. Educators comfortable with learning management systems, digital assessments, and interactive whiteboards showed no greater willingness to integrate AI into their practice than their less experienced counterparts. This suggests that AI represents not merely a technological upgrade but a qualitatively different paradigm, one that even technically proficient educators approach with caution and ambivalence. The reason may lie precisely in the AI paradox described above: when technology encroaches upon the creative and intellectual domain rather than automating the administrative and routine, even experienced technology users feel that something significant is at stake.

3.4 Finding Three: Burnout is a Systemic Risk

Educators with limited prior EdTech experience showed significantly higher burnout, particularly those who had "rarely" or "never" engaged with educational technologies before AI integration. This is not a matter of individual learning curves; it reflects a structural failure of professional preparation. When the already demanding dual-focus pedagogy of CLIL is compounded by the introduction of AI tools for which educators are fundamentally unprepared, the cognitive and emotional load becomes unsustainable. Burnout in this context is not a personal failing but a systemic signal: technology is being introduced at a pace and with a level of institutional support that does not match the psychological reality of the educators expected to implement it.

4. Redefining Tomorrow's Learners and Leaders

4.1 Principle One: Teach Technology Direction, Not Technology Use

The first principle for responding to the AI paradox in education requires a fundamental reorientation of pedagogical purpose. Educators must stop preparing students to compete with AI at tasks AI performs well; that is a losing game, and more importantly, a misconceived one. The genuinely durable skill, the one that will remain relevant across five, ten, and fifty years of technological change, is the capacity to direct AI toward human flourishing: knowing what to automate and what to preserve as irreducibly human.

Consider the specific case of writing. Can AI produce an essay? Yes, with considerable sophistication. Should students nonetheless learn to write? Unquestionably, because, as educational research consistently affirms, writing is thinking made visible. The act of writing requires students to struggle with material, to discover what they actually believe through the discipline of articulation, and to organize complex arguments in ways that reveal and

clarify their understanding. This corresponds precisely to what UNESCO identifies as learning to know and learning to be.

However, the pedagogical response to AI in the writing classroom need not be prohibition. Educators such as Eric Hudson (2025) are developing practices that shift the evaluative focus from product to process: requiring students to share transcripts of their AI interactions, to reflect explicitly on what makes writing good and assess AI output against those criteria, and to defend their work orally, explaining the choices they made, the reasoning behind them, and the moments they diverged from or accepted AI suggestions. These practices do not ask whether AI was used; they ask whether the student is learning. The question that should govern instructional design is not "can AI do this?" but "should it, and under what conditions does genuine human learning still occur?"

Tomorrow's leaders, the agents of positive change envisioned by the OECD Learning Compass, must be fluent in making these determinations. They must know when automation serves human goals and when it undermines essential human development; they must remain the authors of their own thinking even when AI is their co-writer.

4.2 Principle Two: Differentiated Support Across Career Stages

The empirical findings of Mizza, Compagnoni, and Malouche (2026) make clear that the middle-career educator cohort (the 40 to 49 age group) faces distinctive and underacknowledged challenges during AI integration. Professional development programmes that treat technology adoption as a uniform linear process, or that concentrate resources on the youngest and oldest educators, systematically misallocate support.

What the data demands is a differentiated approach organized around three commitments. First, age-sensitive professional development must recognize that a 45-year-old educator with twenty years of established practice faces different challenges, such as professional identity disruption, institutional pressure, and the cognitive effort of restructuring deeply embedded pedagogical habits, than either a recent graduate or a near-retirement colleague. Second, experience-based scaffolding must ensure that educators with limited prior EdTech exposure are given foundational technology skills before AI-specific tools are introduced, rather than being expected to develop both simultaneously under institutional pressure. Third, mental health and burnout prevention must be treated as core infrastructure for technology implementation, not as afterthoughts. When educators burn out, that is not a personal failure; it is evidence that the implementation was under-resourced.

The United Nations Sustainable Development Goal 4 (SDG 4) commitment to inclusive and equitable quality education for all (United Nations, 2015) cannot be fulfilled by educators who are themselves exhausted, unsupported, and professionally destabilized. Sustainable teaching is a prerequisite for quality learning.

4.3 Principle Three: Cultivate Ethical Technology Judgment

Both UNESCO and the OECD point toward ethical reasoning as a transformative competency, not a supplementary skill but a central capacity for navigating an uncertain and complex world. The AI paradox makes ethical technology judgment more urgent than ever, and extends it beyond personal moral reasoning into a civic and institutional dimension.

Tomorrow's leaders must be equipped not only to use AI tools but to interrogate them: to ask whose values are embedded in a given algorithmic system, what is gained and what is lost when a particular human activity is automated, where the boundary between human judgment and algorithmic recommendation should be drawn, and who bears responsibility

for AI-mediated decisions. These are not technical questions; they are questions of power, equity, and human dignity.

UNESCO's call for "learning to live together" and the OECD's emphasis on reconciling tensions and taking responsibility both point toward this capacity. Education that takes the AI paradox seriously will produce graduates who are not merely AI users but architects of AI's role in society: people capable of drawing principled distinctions between what technology should do and what must remain in human hands. This is what the OECD's transformative competencies look like in practice: not the ability to operate a tool, but the wisdom to determine when and whether it should be operated at all.

5. Conclusion

We stand at an educational crossroads. The frameworks that have guided global education policy for decades, UNESCO's four pillars and the OECD's Learning Compass, were developed in anticipation of a complex and uncertain future. That future has arrived, and it is more algorithmically mediated than its architects anticipated. The tasks through which students have traditionally developed critical thinking, creativity, and ethical reasoning are now performable by machines. The educators tasked with supporting this development are themselves navigating unprecedented technological disruption, often without adequate institutional support.

The central question before educators, institutions, and policymakers is not whether to integrate AI into education; that question is largely settled by the world students will enter. The question is how to do so without hollowing out the human development that education is meant to achieve. The answer lies in three commitments: teaching students to direct AI toward human goals rather than merely to use it; providing educators with differentiated, career-stage-sensitive, and psychologically informed professional development; and cultivating in students the ethical judgment to determine what should never be delegated to a machine.

Are we building tomorrow's leaders, or tomorrow's tool-users? That question, ultimately, depends on whether educators are given the support, the intellectual resources, and the institutional permission to ask it. The findings of Mizza, Compagnoni, and Malouche (2026) suggest that the conditions for asking it are difficult, but the urgency of doing so has never been greater.

References

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340. <https://doi.org/10.2307/249008>

Hudson, E. (2025, April 27). AI and the teaching of writing: Six core principles. *Learning on Purpose*. <https://erichudson.substack.com/p/ai-and-the-teaching-of-writing>

Keenon Robotics. (n.d.). SANBOT Nano home service robot. SANBOT. <https://www.sanbot.com>

Maciejewska, J. [@AuthorJMac]. (2024, March 29). I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes [Post]. X. <https://x.com/AuthorJMac/status/1773679197631701238>

Mizza, M., Compagnoni, A., & Malouche, D. (2026). Beyond digital natives: Non-linear age effects on teachers' AI adoption stress in complex pedagogical environments. Manuscript under review.

OECD. (2019). OECD Future of Education and Skills 2030: OECD Learning Compass 2030. OECD Publishing.

UNESCO. (1996). Learning: The treasure within. Report to UNESCO of the International Commission on Education for the Twenty-first Century (J. Delors, Chair). UNESCO Publishing. <https://unesdoc.unesco.org/ark:/48223/pf0000109590>

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO. <https://www.unesco.org/en/digital-education/ai-future-learning>

United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development. United Nations. <https://sdgs.un.org/2030agenda>

Wang, X., & Li, B. (2019). Technostress among university teachers in higher education: A study using multidimensional person-environment misfit theory. *Frontiers in Psychology*, 10(1791). <https://doi.org/10.3389/fpsyg.2019.01791>

Transformative Teaching with AI: A Faculty Development Case Study for Higher Education

Claudia Carrone – Emmanuelle Houet, EDHEC Business School

Abstract

The rapid integration of generative artificial intelligence (GenAI) into higher education is reshaping teaching practices, assessment design, and expectations around academic integrity. While faculty are increasingly expected to integrate AI into their courses, many lack structured, practice-based opportunities to develop the pedagogical judgment required for meaningful and ethical use. This case study examines the Transformative Teaching with AI Certificate developed at EDHEC Business School, a faculty development programme designed to support intentional, human-centred integration of GenAI into teaching.

Grounded in experiential and practice-based pedagogical frameworks, the programme combines asynchronous learning with guided experimentation embedded in participants' own courses. The paper describes the programme's design principles, learning pathway, and pedagogical rationale, illustrating how it moves beyond tool-focused training toward reflective course and assessment redesign. By foregrounding pedagogy over technology, the case study offers a structured and transferable model for institutions seeking scalable strategies to support educators in navigating AI-rich learning environments while upholding core academic values.

Keywords: Generative artificial intelligence, Faculty development, Pedagogical design, Higher education teaching

Introduction / Problem statement

The rapid rise of generative artificial intelligence (GenAI) is reshaping expectations for teaching and learning in higher education. Once experimental, generative tools are now embedded in students' and faculty members' daily academic practices. A 2025 Conférence des Grandes Écoles study confirms this shift: more than two-thirds of students and over half of faculty use AI several times a week. This widespread adoption is prompting universities to confront new questions about academic integrity, learning processes, and the evolving role of the educator.

For institutions, the challenge is no longer whether AI will transform learning, but how to support educators in navigating this transformation. Faculty are now expected to integrate AI into course design, redesign assessments, cultivate students' AI literacy, and uphold academic standards. Yet many educators still lack structured, practice-based opportunities to develop the skills, confidence, and pedagogical judgment required to use AI meaningfully and ethically in their teaching.

This case study addresses this gap by examining EDHEC Business School's **Transformative Teaching with AI Certificate**, a faculty development programme designed by the **PILab** (Pedagogical Innovation Lab, Edhec's Center for Teaching and Learning) to equip educators with the frameworks, competencies, and hands-on experience required to integrate AI intentionally into their teaching.

Its contribution is twofold. First, it presents a structured, experiential model for faculty development that moves beyond conceptual awareness to guided experimentation and reflective practice. Second, it provides evidence of how such a model can build faculty confidence, reshape teaching practices, and support institutions in developing coherent, human-centred approaches to AI integration. Together, these insights offer a pathway for institutions seeking scalable, pedagogically robust strategies to prepare educators for AI-rich learning environments.

Institutional context / Motivation

EDHEC Business School operates across multiple campuses and delivers programmes from undergraduate to executive levels, engaging a diverse faculty including permanent professors, “vacataires”, adjunct instructors, and industry experts. This diversity is one of the institution's strengths, but it also presents challenges when introducing emerging technologies: faculty vary widely in technological confidence, pedagogical training, and exposure to digital tools. Prior faculty development initiatives showed that, while many instructors recognised AI's relevance, they sometimes lacked structured opportunities to explore its pedagogical implications within the context of their own courses.

The growing presence of GenAI in students' academic practices further intensified the need for action. Instructors, programme directors, and students all called for clearer expectations around academic integrity and responsible AI use.

In parallel, EDHEC's strategic orientation, emphasising pedagogical innovation, digital transformation, and future-ready skills, created momentum for an initiative that could support faculty at scale. Traditional workshop formats proved insufficient; a more comprehensive approach was required, combining conceptual grounding, practical experimentation, and personalised reflection, while remaining accessible to instructors with diverse disciplinary backgrounds and time constraints.

The **Transformative Teaching with AI Certificate** was designed in response to this landscape. Its purpose is not only to build faculty digital literacy but also to provide a coherent, pedagogically informed pathway for redesigning teaching practices in an AI-rich environment.

Programme Design and Pedagogical Framework

The programme is built on three assumptions supported by research in pedagogy and faculty development.

First, **meaningful AI integration requires pedagogical judgment rather than tool-focused training**, a finding consistent with frameworks such as The Technological Pedagogical Content Knowledge (TPACK) and the SAMR model (Substitution, Augmentation, Modification, Redefinition) which emphasise the primacy of instructional purpose in technology adoption.

Second, **faculty learn most effectively when experimentation is embedded in their own courses**, reflecting principles of experiential learning and evidence that discipline-specific, practice-based development leads to deeper and more durable change.

Finally, **structured and guided practice reduces uncertainty**, identified in recent AI and education studies as a major barrier to adoption, and supports sustained shifts in teaching behaviour. These assumptions underpin the programme's emphasis on iterative experimentation, reflection, and pedagogical intention.

To translate these principles into practice, the programme combines short asynchronous modules with hands-on activities and iterative course redesign tasks. **The structure draws on experiential learning cycles**: participants encounter new concepts, apply them to teaching materials, reflect on outcomes, and refine their approach.



Stages of the Transformative Teaching with AI Learning Journey

- *Foundations: Course structure and purpose*

Participants begin by revisiting core elements of their course design: purpose, learner characteristics, and alignment between intentions and student needs. Guided experimentation with AI assists them in analysing learner profiles and exploring alternative ways to frame key concepts, positioning AI as a tool for clarifying course structure rather than automating instruction.

- *Redefining Learning Objectives in an AI-rich environment*

Building on this foundation, faculty examine how learning objectives may shift when generative tools can perform many lower-level cognitive tasks. Drawing on traditional cognitive taxonomies, such as Bloom's revised one, they explore objectives that emphasise reasoning, critique of AI outputs, process transparency, and metacognition. This provides a foundation for rethinking assessments and activities.

- *Designing and adapting learning resources*

A central component of the certificate is hands-on work with course materials. Faculty experiment with generative AI to design or adapt learning resources tailored to diverse learner needs. Through guided experimentation, participants learn to use AI not simply for content production but as a creative partner that supports differentiation, accessibility, and engagement.

- *Rethinking assessment and academic integrity*

The programme encourages faculty to shift from attempting to "AI-proof" assignments toward clarifying the intended role of AI and the skills students must demonstrate. Participants explore process-based tasks, reflective components, oral elements, and assignments that integrate AI transparently and responsibly.

- *Teaching across modalities with AI support*

The final stage examines how AI can strengthen teaching presence and learner interaction across face-to-face, hybrid, and online formats. Faculty learn to use AI to generate discussion prompts, examples, summaries, and clarification resources that support interaction while preserving the relational dimension of teaching.

Together, these modules form a coherent pathway through which faculty build conceptual understanding, digital literacy, and pedagogical confidence. By moving from foundational course design to assessments and teaching presence, the programme provides a structured and transferable model for supporting educators as they navigate AI-rich learning environments.

Evidence of Impact

This case study draws on preliminary qualitative evidence gathered since the launch of the course this year (September 2025). Additional data collection is ongoing.

Analysis of observations reveals three findings that illustrate how faculty engaged with the programme and how their teaching practices evolved.

- *Increased confidence, agency, and shared understanding*

Participants reported entering the programme with uncertainty or apprehension regarding the role of AI in teaching. Through guided experimentation and structured practice, faculty described a growing sense of agency in evaluating AI tools, articulating their pedagogical value, and making informed decisions about use. Discussions across disciplines also revealed shared challenges which contributed to reduced isolation and strengthened collective understanding of AI's pedagogical implications.

"I completed the Transformative Teaching with AI certificate out of curiosity to discover how AI could assist me in reviewing and updating my courses. I was delighted to learn practical skills for using AI effectively as well as ideas to enhance student engagement and learning... By doing this course, I gained new knowledge and got a great start for my course updates."

Marjorie THOMPSON - Professor

- *Meaningful redesign of courses, assessments, and learning activities*

Participants implemented concrete changes to their teaching, including redesigning assessments for greater reasoning and transparency, adding reflective or oral components, and creating AI-supported activities such as iterative writing tasks or personalised materials. *For example, by the end of the certificate course, a Spanish teacher chose to offer her students a reflective activity using AI, whereas only months earlier she had been focused solely on trying, unsuccessfully, to detect students cheating using AI.*

- *Clearer articulation of academic integrity and responsible use expectations*

Participants demonstrated increasing clarity in defining what constitutes appropriate student use of AI within their courses. Statements on academic integrity were refined to explain not only what is allowed, but why. These reflections show a broader shift from reactive policing to proactive, transparent guidance aligned with institutional values.

"This project offers faculty a structured, innovative, and deeply human approach to integrating AI... It strengthens academic quality and supports our mission to develop autonomous, critical, and responsible young talent in a world shaped by AI."

Anne ZUCCARELLI – Director of Student Experience

Lessons Learned

The implementation of the Transformative Teaching with AI Certificate highlights several lessons relevant for institutions seeking to support faculty in navigating AI-rich learning environments.

First, structured and guided experimentation is essential. Faculty progressed most when working directly with their own course materials in a supportive environment that balanced autonomy with clear instructional guidance. This approach reduced uncertainty and enabled intentional pedagogical decisions.

Second, cross-disciplinary dialogue accelerates learning. Participants valued hearing how colleagues in other fields confronted similar concerns about assessment, workload, and responsible AI use.

Finally, institutional clarity and leadership support matter. Adoption was strengthened when programme messages aligned with emerging guidelines on academic integrity and AI literacy, creating conditions for sustained pedagogical change.

Implications for Practice and Institutional Strategy

The findings from this case study contribute to ongoing discussions about how higher education institutions can support meaningful, responsible integration of generative AI into teaching across a diverse faculty, regardless of status, discipline, or level of digital confidence.

Consistent with research in digital pedagogy, the results confirm that faculty adoption is more a pedagogical than a technical challenge. Participants did not require advanced technological expertise but structured opportunities to reflect on their teaching, test new approaches, and articulate clear expectations for students.

The programme also highlights the importance of institutional coherence in navigating AI adoption. Faculty confidence grew when programme activities aligned with emerging institutional guidance on academic integrity and AI use. This underscores the value of coordinated policies and development pathways that reduce fragmentation and foster shared understanding across departments.

At a broader level, the case study suggests that generative AI can serve as a catalyst for pedagogical renewal. The issue is not simply whether to use AI, but how its emergence encourages faculty to revisit course purpose, reframe learning objectives, and redesign assessments. These shifts prompt deeper reflection on what constitutes evidence of learning across disciplines. This reflective process has implications beyond AI integration: it strengthens teaching practice and supports a culture of continuous pedagogical improvement.

Conclusion

The Transformative Teaching with AI Certificate demonstrates how structured, practice-oriented faculty development can support meaningful integration of generative AI into higher education teaching. By combining guided experimentation with reflective practice, the programme helped educators move from uncertainty to pedagogically grounded use of AI, resulting in meaningful changes to assessment design, learning activities, and academic integrity guidance.

Ultimately, the certificate offers a transferable model for institutions seeking to prepare AI-ready educators and strengthen teaching innovation at scale. By centring pedagogy rather than technology and foregrounding human-centred instructional principles, it shows that AI can enhance learning, empower educators, and uphold academic values. As higher education continues to navigate rapid technological change, such structured and reflective pathways will be essential for cultivating teaching practices that are both future-oriented and deeply grounded in the human experience of learning.

References

Conférence des grandes écoles. (2025, novembre). Usages de l'intelligence artificielle dans les Grandes écoles (Commission Numérique, Enquêtes & Baromètres).

https://www.cge.asso.fr/wp-content/uploads/2025/12/20251208_Livrable_CGE_enquete-IA.pdf

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, 48(8), 952–984.

Holmes, W., Bialik, M., & Fadel, C. (2023). *Artificial intelligence in education: Promises and implications for teaching and learning* (Updated edition). Center for Curriculum Redesign.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054.

PuenteDura, R. R. (2014). SAMR: A contextualized introduction. Retrieved from https://hippasus.com/blog/2013/10/25/samr_a_contextu_1/

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing.

AI Literacy in Learning - From Training to Reflective Practice

Boris Walbaum
And Liudmila
Piatnitckaia

Abstract

The Forward College AI Hub represents a pioneering approach to managing student-AI interactions in higher education designed to harness the benefits of AI assistance while preventing the cognitive offloading that can undermine genuine learning.

It was developed through two years of action research rather than assumptions about how students should use artificial intelligence. This comprehensive initiative evolved from initial paper-based manuals co-created with students into a multi model platform featuring real-time learning analytics and reflective dashboards.

Keywords: Generative AI, ChatGPT, Cognitive Offloading, Bloom's taxonomy, higher education

Phase 1: Co-Creation and Student-Centered Design (2023-2024)

The initiative began in April 2023 when Forward College provided ChatGPT+ licenses to students preparing for final examinations, requesting anonymized logs in exchange to understand authentic student-AI interactions. This exploratory phase revealed strong student appetite for AI assistance, prompting the development of formal AI literacy training programs. The research team organized interactive workshops covering AI fundamentals and prompt engineering techniques, but more importantly, invited students to create their own AI usage manuals tailored to specific subjects.

Through this participatory design process, 30 students produced 11 prototype manuals, from which three were selected based on peer and expert evaluation. These prototypes were synthesized into a comprehensive guide incorporating evidence-based educational principles and concrete use cases defined by students themselves. The guide structured AI interactions according to Bloom's taxonomy, helping students understand how to engage AI at different cognitive levels - from basic recall to complex creative tasks. Student interviews revealed positive reception, with participants particularly appreciating the

guide's clarity, organization, and focus on metacognition and learning habits rather than shortcuts.

This co-creation approach reflects constructivist learning principles, recognizing that deep learning emerges through active engagement rather than passive consumption. By involving end-users directly in the development process, the team ensured greater alignment with authentic pedagogical needs. The research from this phase was presented at the EDULEARN24 Conference¹, establishing Forward College as a leader in empirical research on student-AI interactions.

Phase 2: Understanding Usage Patterns Through Empirical Research (2024)

While students embraced AI tools, concerns emerged about potential negative impacts on learning outcomes. Recent empirical research had documented a troubling phenomenon: despite improvements in output quality, students using AI systems often showed reduced learning outcomes and diminished cognitive engagement². This prompted two comprehensive studies to capture both the breadth and depth of student-AI interactions.

Study 1: Large-Scale Pattern Analysis

The first study involved 46 undergraduate students from diverse programs including Data Science, Business Analytics, and Psychology. Organized into 12 groups sharing ChatGPT-4 accounts, students used AI freely for assessment preparation over seven weeks from April to June 2024. This generated a substantial corpus of 408 dialogues containing 3,047 student messages.

Analysis employed an innovative LLM-assisted coding approach, using Python scripts interacting with GPT-4o-mini to classify each message by question type and Bloom's taxonomy level. The classification required chain-of-thought reasoning, with the model generating rationales before producing final labels. This taxonomy-based approach distinguished between lower-order cognitive tasks (remembering, understanding) and higher-order thinking processes (analyzing, evaluating, creating).

Key findings revealed concerning patterns: 27% of queries operated at the "Analyze" level of Bloom's taxonomy, indicating students were outsourcing complex analytical thinking to AI, particularly for synthesis and critical evaluation tasks. Another 24% were at the "Understand" level, with students preferring AI-generated clarifications to re-reading source materials. By question type, procedural requests dominated at 23%, followed by definitional questions at 14%. This distribution suggested that while some AI use targeted foundational knowledge acquisition, a substantial proportion addressed higher-order cognitive functions that are crucial for deep learning.

Study 2: Cognitive Offloading Analysis

The second study examined a subsample of 16 students who consented to link their AI interactions with academic performance data. This enabled analysis of 1,140 messages across 125 dialogues with manual, fine-grained coding for cognitive offloading severity. Students were grouped into performance tiers based on University of London summative assessment grades, allowing exploration of relationships between AI usage patterns and academic achievement.

Building on Cognitive Load Theory³ which distinguishes between extraneous load (inefficient presentation) and germane load (desirable mental effort building lasting knowledge), the team developed a three-tier framework for understanding cognitive

offloading⁴. "No Offloading" described AI use for clarification or information access without performing cognitive work, with students remaining actively engaged in core analytical tasks. "Light Offloading" involved delegation of extraneous cognitive load through summarizing, reformatting, or explaining terminology - surface-level tasks supporting deeper learning without substituting for it. "Heavy Offloading" represented delegation of germane cognitive load through generating arguments, composing essays, or producing analysis - substituting AI for core intellectual work required by learning tasks.

Manual coding revealed that 49.6% of dialogues showed no offloading, with students using AI only for explanations while retaining responsibility for synthesis and analysis. Another 34.4% showed light offloading, delegating formatting and definitional tasks while maintaining engagement with core intellectual work. However, 16.0% showed heavy offloading, where students delegated generation of arguments, essays, or analytical outputs directly to AI.

A surprising finding emerged when linking offloading patterns to academic performance: top-tier students showed the highest rates of heavy offloading at 17.6%, while bottom-tier students rarely used heavy offloading at 5.9%. This pattern suggested that effective AI delegation might be a skill correlating with academic ability. Stronger students appeared more confident using AI strategically for complex tasks, leveraging assistance as a tool for acceleration and refinement rather than as a thinking substitute.

Heavy offloading clustered particularly in "Create" dialogues like drafting arguments or code, while "Understand" and "Apply" tasks more often involved light or no offloading. This distribution reflected students' intuitive recognition that some tasks demand direct cognitive engagement while others are more amenable to AI assistance. These findings were presented at ICLS 2025, with a full paper currently under journal review.

Phase 3: Building the AI Hub Platform (2024-2025)

Rationale

Initial analyses revealed both potential and perils of student-AI interaction within Forward College's flipped classroom environment. While AI systems can enhance output quality and support learning processes, they also create conditions for cognitive offloading and potential attenuation of critical thinking skills. Early attempts to address these concerns through a "Socratic chatbot" that resisted heavy offloading by demanding student input proved unsuccessful. Both heavily and lightly shielded versions were assessed as highly frustrating by students, creating risks that students would evade monitored accounts in favor of direct LLM access.

This setback prompted a strategic pivot toward prioritizing student autonomy and agency by building a platform that provided larger access to AI while supporting reflective, intentional engagement. The AI Hub was designed with three core features addressing distinct needs identified through research.

First, the platform offers access to a broad spectrum of leading language models from Grok to Claude, giving students opportunities to develop nuanced understanding of each system's distinct capabilities and limitations. This diversity exposure cultivates essential AI literacy - the ability to evaluate, compare, and critically assess different AI systems rather than relying on a single tool.

Second, privacy guarantees address fundamental concerns from earlier research phases where shared accounts and logged interactions may have constrained authentic behavior.

The AI Hub provides private, individual accounts for each student, with data belonging to students rather than being shared with peers, used for grading, or monitored for academic misconduct. This privacy-first design reflects broader ethical principles in educational technology, safeguarding data ownership and autonomy to foster trust and genuine learning engagement.

Third, the reflective dashboard serves as the platform's centerpiece, allowing students to monitor the quality and character of their AI interactions from a learning impact perspective. Rather than imposing restrictions or judging behavior, the dashboard presents students with a mirror of their usage patterns, inviting reflection and intentional decision-making. This approach grounds itself in the conviction that students, when provided transparent information about their learning processes, will develop greater metacognitive awareness and make more strategic choices.

The Reflective Dashboard Design and Implementation

The dashboard provides detailed insights into AI engagement patterns, allowing students to observe prompt quality evolution over time and track how their AI interaction skills develop with experience. More importantly, they can see what types of learning tasks they delegate to AI, categorized according to the research-derived taxonomy of cognitive offloading. This taxonomy, grounded in Cognitive Load Theory and aligned with Bloom's taxonomy⁵, makes cognitive complexity visible to students, supporting metacognitive awareness.

The dashboard incorporates social comparison as a soft incentive, showing individual scores alongside average cohort performance on all metrics. Importantly, cognitive offloading metrics are presented not as judgments but as information for reflection. The dashboard avoids labeling heavy offloading as "bad" or no offloading as "good," instead presenting data transparently and recognizing that different learning contexts and student goals may warrant different offloading strategies.

To operationalize the dashboard, an automated classification system powered by GPT-5-mini analyzes each student-AI exchange. The system answers structured questions about interaction nature, capturing multiple dimensions including prompting quality (presence of context, instructions, role assignment), type of learning task (synthesis, concept understanding, practice exercises, feedback requests, brainstorming, planning, memorization), and level of cognitive offloading (low, medium, high). Classification results aggregate in the dashboard, enabling tracking of both individual patterns and broader trends.

Future Directions

The platform tracks how offloading patterns vary across different learning domains, revealing important context-specific patterns. Fine-grained analysis of prompting strategies illuminates how student AI literacy develops over time. The team pursues granular analysis to determine whether observed increases emanate from a small subset of heavy offloaders or are more broadly distributed, carrying important implications for intervention strategies.

Future research objectives focus on stabilizing infrastructure to enable rigorous longitudinal studies. The team will investigate whether the reflective dashboard leads to more responsible AI use through declining heavy offloading rates and increased intentionality, as well as more sophisticated use reflected in nuanced prompting strategies and strategic task allocation. Correlational studies will examine relationships between responsible AI use patterns and learning outcomes, including summative assessments and skill development

measures. The research will map offloading strategy distribution across student populations to identify susceptibility patterns.

References

¹ "Developing an AI-Powered Assistant for Enhanced Learning: A Multi-Phase Research Project at a Private Higher Education Institution" (Piatnitckaia, Corneloup, & Zenasni, 2024)

² Hechingerreport.org. (2024). "University students offload critical thinking, other hard work to AI" — Studies on metacognitive laziness and cognitive engagement decline with ChatGPT. And Stanford Scale et arXiv. (2024). "ChatGPT Produces More 'Lazy' Thinkers: Evidence of Cognitive Engagement Decline."

³ Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.

⁴ PMC/NCBI. (2025). "The Cognitive Paradox of AI in Education" — Cognitive Load Theory and Bloom's Taxonomy applications to AI-enhanced learning.

⁵ PMC/NCBI. (2025). "The Cognitive Paradox of AI in Education" — Cognitive Load Theory and Bloom's Taxonomy applications to AI-enhanced learning.

AI Policy as Pedagogy: Guiding Student Learning with Generative AI

Justin Olmanson, Minji
Jeon, Azadeh Hassani,
University of Nebraska -
Lincoln

Abstract

Generative AI that is capable of supporting student learning has been widely available since late 2022. Yet the institutional and instructional guidance students receive regarding how they should, should not, or might use AI to complete course assignments and engage with course content is often absent or insufficiently differentiated. This creates an uneven dynamic wherein individual students are left to either interpret vague AI policies and apply them on an assignment-by-assignment and course-by-course basis, or adhere to sweeping statements made by instructors regardless of differences in AI literacy levels. This project translates an empirically derived taxonomy of AI use for learning into a student-facing AI policy designed to support student judgment, agency, and productive struggle in a way that maps onto specific coursework and assignments. We describe both the taxonomy and the policy and discuss the implications of reframing AI policies as pedagogical tools designed to cultivate student agency, rather than as an instrument of institutional compliance and control.

Keywords: AI policy, undergraduate students, AI in education, learning, course-related independent learning

Introduction and background

Soon after the release of capable large language model-based tools in late 2022 (OpenAI, 2022), universities have articulated policies governing the use of generative AI in course activities. Much of the initial policy discourse has centered on concerns about academic integrity, plagiarism, and the potential decline of students' cognitive abilities (Chan, 2023). In parallel, researchers and educators recognized the learning opportunities presented by AI to support and transform teaching and learning via personalized feedback, tutoring, and access to resource distillation. These two potentialities required AI policies to balance learning and growth-related risk and opportunity. In practice, however, institutional responses have been uneven. Analyses of university AI policies show that many institutions are still in the early stages of AI governance, typically offering generalized guidance in the form of boilerplate language for use in syllabi while leaving the work of determining and articulating acceptable and prohibited use cases to instructors (McCusker & Michalak, 2025). This decentralized approach creates high levels of variation in terms of rules and expectations across courses and within and between degree programs, leaving students to

navigate AI based on an array of articulated and unarticulated instructional theories of AI use for learning (Azevedo et al., 2026).

Although institutional guidance regarding AI remains inconsistent, students are still expected to navigate how and when they use generative AI in their learning processes. These decisions require a type of AI literacy that includes and extends beyond their familiarity with AI tools to include an ability to identify AI use that supports course learning aims. AI literacy is often framed as a set of competencies that enable users to evaluate AI technologies, use AI systems effectively, and discern when and how to use them in a responsible manner (Laupichler et al., 2022). Recent work in this area has expanded the construct to include: technical understanding, skill with application, critical thinking, ethical awareness, and the ability to integrate AI into disciplinary processes (Hackl et al., 2026). However, while discernment and disciplinary integration are key parts of AI literacy, university guidelines often lack student-facing, learning-centric frameworks that support students' growth and connect their competencies to specific decisions they have to make when using AI for different types of course-related tasks. Absent such frameworks, learners are left to navigate AI use via a set of informal, implicit, interpolated norms, instructor statements, and personal experimentation rather than structured guidance about how AI can support learning and disciplinary growth.

In response to this gap, we present a student-facing AI policy designed to help undergraduates navigate how and when AI might best be used in the completion of coursework that support learning goals and avoid uses that are detrimental to learning.

Methods

In this section, we describe the empirical process we used to develop the taxonomy of learning with AI as well as the design process we enacted in the development of the policy.

Taxonomy development

The taxonomy, on which the AI policy below was based, draws on a qualitative synthesis of data from two design-based research studies that examined how students used generative AI in course-related independent learning contexts. Across the studies, 25 participants (18 undergraduate and 7 graduate students) were observed and interviewed regarding their use of AI for completing course tasks. Data sources included observations, student artifacts, and semi-structured interviews. The combined dataset was analyzed using open coding and thematic analysis to identify recurring patterns of AI use (Anfara et al., 2002; Weiss, 1995). From this analysis, the research team abductively generated a provisional heuristic framework that described emerging categories of participant AI use for learning. The framework was then shared with over 120 students and educators to assess category fit and identify potential gaps. Finally, we conducted a re-analysis of the data to refine category boundaries--resulting in a taxonomy of six archetypes of AI use for learning organized along a continuum of learner agency (Figure 1).

Figure 1

A taxonomy of learning with generative AI (Olmanson et al., 2026)

	Primary Themes	Subcategories & Motivations	AI Roles
↑ agency ↓	Magnify	Help me to be more: creative, productive, and ambitious...	Multiplier, co-creator, collaborator, role-player
	Help Me Learn	Explain: concepts & processes; summarize, guide, test me...	Tutor, proctor, more knowledgeable other
	Feedback Please	Critique my: thinking, work, plan; make suggestions, keep me company...	Critical reviewer, editor, companion
	Get Me Going	Help me: plan, get started, get organized, write a first paragraph, get unstuck...	Initiator, planner, momentum builder
	Escape Levels 1 & 2	L2: Too tedious, peripheral, bad task... L1: Don't want to, no time, low energy...	Efficiency optimizer, End-to-end fixer
	Not For Me	Don't know how, too risky, too \$\$\$, don't need it, I want to learn...	No role for AI

Policy design

We developed the AI policy via a translational design process wherein we adapted the taxonomy of AI use for learning into a student-facing instructional tool. Instead of framing the policy as a compliance mechanism, it was designed to support students in thinking about how different forms of AI use related to learning goals within specific course contexts. Each of the six categories identified in the taxonomy were reframed in accessible language and combined with guidance indicating whether each type of use was prohibited, discouraged, allowed, or encouraged (Table 1). To make these distinctions more concrete and actionable, each category was accompanied by example prompts describing possible interactions with AI as well as brief rationales unpacking why particular uses supported or impeded learning (Table 2). In the opening paragraphs of the AI policy, we invite students to reflect on their motivations for using AI and to consider whether a given interaction supports understanding, feedback, or learning extension or if it simply accelerates the pace at which they can complete a task. In these ways, the policy is designed to function as a pedagogical scaffold that supports student decision-making, increases transparency, and alleviates anxiety. The next section consists of the AI policy text students see either embedded in the course syllabus or as a stand-alone document in their course learning management system.

A student-facing undergraduate AI policy

Productive failure (Kapur, 2016), productive struggle (Warshauer, 2015) and continuous effort are an integral part of learning, developing as a scholar, and producing work that is yours. AI should not be a replacement for struggle and effort but rather, should be used to support your learning, and expand how you analyze, think critically, problem-solve, and produce work.

Use your feelings as a guide, if you feel drawn to using AI out of desperation, fear, insecurity, or time pressure your use is less likely to support your long-term learning goals or be approved by your instructor. Conversely, if you feel drawn to using AI in an intentional, reflective way that engages with your feelings of curiosity and a strategic, authentic interest in learning and growth, AI may be more likely to align with professionalizing outcomes and

be approved by your instructor. Ask yourself, “is the way I’m using AI helping me learn, or is it just helping me finish faster?”

The use of AI tools such as ChatGPT, Claude, or Gemini is allowed in this course under particular circumstances. For homework and projects, you must clearly indicate your use of AI tools and provide appropriate documentation of exactly how the tool was used. In this course we use specific categories of AI use for learning and coursework (see Table 1).

Table 1

Categories of AI use for learning and coursework

AI Use Category	Examples	Status & Directions
Do All the Work for Me	Using AI to complete a last-minute class project because you were too tired, ran out of time, or didn’t see the point of the project.	Prohibited Instead contact the instructor and figure out an alternative approach.
Do My Busywork	Using AI to complete course projects or parts of projects because you feel they are repetitive or will not help you learn anything.	Discouraged without permission Discuss with the instructor.
Get Me Going / Get Me Started	Using AI to summarize instructions, help with planning, write an outline, write a first sentence, or get you unstuck.	Allowed Document your use and include it in your project.
Give Me Feedback	Using AI to check your thinking or work, give you encouragement or suggestions about where your project is: strong, missing something, or next steps.	Allowed Document your use and include it in your project.
Help Me Learn	Using AI to explain new or difficult concepts or procedures step by step or holistically by using accessible language and examples.	Encouraged Note your use in the project.
Magnify My Work	Using AI to go beyond what the assignment or project requires.	Encouraged Document your use.

To give a better sense of how these categories work, Table 2 below offers examples of prompts for each type of AI use for learning and coursework as well as an explanation of why each type of use is prohibited, discouraged, allowed, or encouraged.

Table 2

Example prompts for each category of AI use for learning and coursework

AI Use Category	Example Prompts	Status & Directions
Do All the Work for Me	<p>“Please use the assignment instructions I’ve pasted below and the attached rubric and source documents to complete the assignment for me. I’ve also attached a previous similar assignment that I did so that you can mimic my style and voice.”</p> <p>“Use the picture of my calculus quiz to give me the answers to each question.”</p>	Prohibited When AI does the work for you, you learn much less and you miss out on developing professional skills and a good work ethic.

Do My Busywork	<p>“Write me a short summary of the attached article that I can post on a canvas discussion board.”</p> <p>“Clean up my citations and put them in APA format.”</p>	<p>Discouraged without permission</p> <p>Let your instructor know that you do not see the learning potential in the task.</p>
Get Me Going / Get Me Started	<p>“I’ve got this assignment and I’m not sure how to get started, here are the instructions below, summarize what the assignment is asking me to do and give me two things I can do to get started. This is for class so do not offer to do my thinking for me.”</p> <p>“I’m stuck on the calculus problem in the screenshot. I’m supposed to find the volume of a solid of revolution. I found the intersection points of the two curves but I don’t know what to do next, give me a hint.”</p>	<p>Allowed</p> <p>Using AI to get you started or unstuck is permissible, but if you are not careful it will try and do the entire task for you, so be clear that you only want a nudge to get some momentum.</p>
Give Me Feedback	<p>“Take a look at my first draft of the project I’m working on as well as the project instructions and rubric below and give me your honest assessment. Tell me what looks good, what needs work, and what grade you think I would get if I turned it in now. Do not offer to make any changes for me, just act as a supportive teacher.”</p>	<p>Allowed</p> <p>Getting feedback in a timely fashion can be difficult, AI offers one of the fastest ways to get suggestions. Be specific and tell it to offer feedback only.</p>
Help Me Learn	<p>“Help me understand the concept of predicting reactivity in organic chemistry, treat me like a beginner and, if possible, use relationship or basketball metaphors.”</p> <p>“Probe my understanding to see if I fully understand the concept of translanguaging and how it can be used in schools. When you notice gaps in what I know, help me fill them in.”</p> <p>“Pretend to be Lev Vygotsky and let me ask you questions about your theory of the Zone of Proximal Development.”</p>	<p>Encouraged</p> <p>AI can be a useful way to learn new concepts, test your understanding, or learn in a new way. Imagine the perfect learning scenario for you and ask AI to make it happen.</p>
Magnify My Work	<p>“I’m supposed to write a poem about the Civil War (see my song lyrics pasted below). Can you help me write a heavy metal musical score for the song? Vocals, guitars, & drums.”</p> <p>“For my theater class I need to read three plays by the same author and then chart the playwright’s themes across plays. I want to do this for not just three plays but all the plays by William Shakespeare, can you help me?”</p>	<p>Encouraged</p> <p>Using AI to expand your learning can enable you to explore topics outside of course expectations but within your interests and curiosity.</p>

Discussion

AI policies often focus on drawing lines and setting expectations regarding academic integrity. The AI policy approach above offers an example of how policy can raise student awareness regarding how AI can be used in coursework while also supporting students as they evaluate how they might best use AI in support of their learning.

The implications for this work are clear: student-facing policies based on learning-centered frameworks have the potential to help institutions of higher education move beyond drawing lines in the sand and toward nuanced and contoured AI literacy development.

Framed in this way, AI policy can function as both a vehicle for institutional governance and as a pedagogical tool capable of cultivating explicit learning-centered approaches to AI-supported course-related learning.

References

- Azevedo, L., Mallinson, D. J., Wang, J., Robles, P., & Best, E. (2026). AI policies, equity, and morality and the implications for faculty in higher education. *Public Integrity*, 28(2), 186–201. <https://doi.org/10.1080/10999922.2024.2414957>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(1), 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Hackl, V., Müller, A. E., & Sailer, M. (2026). The AI literacy heptagon: A structured approach to AI literacy in higher education. *Computers and Education: Artificial Intelligence*, 10, 100540. <https://doi.org/10.1016/j.caeai.2026.100540>
- Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3, 100101. <https://doi.org/10.1016/j.caeai.2022.100101>
- McCusker, E., & Michalak, R. (2025). AI Policies in U.S. Universities: A Critical Analysis of Policy Gaps and Library Involvement. *Journal of Library Administration*, 65(6–7), 808–824. <https://doi.org/10.1080/01930826.2025.2560268>
- Olmanson, J., Hassani, A., & Jeon, M. (2026). A Taxonomy of AI Use for Learning in Higher Education. AECT 2025. Association for Educational Communications & Technology. https://doi.org/10.35542/osf.io/8ajbg_v2
- OpenAI. (2022). ChatGPT (Version November 2022 version) [Large language model]. <https://chat.openai.com/>

Transforming Passive Learning into Intelligent Learning through a GenAI-Enabled Tutor-Like Scaffolding Model

Nazim Ali and Sarah A. Aynsley

School of Medicine, Keele University, Newcastle under Lyme, Staffordshire, ST5 5BG, UK

Abstract

While generative artificial intelligence (GenAI) is increasingly positioned as a driver of transformation in higher education, a clear disconnect remains between this ambition and current practice. At present, GenAI use by both students and educators is largely confined to content generation, rather than being purposefully integrated within curricula to support higher-order learning processes. This case study addresses this disconnect by presenting a GenAI-enabled learning model embedded directly within curriculum delivery, designed to engage students in the application of critical thinking skills associated with intelligent learning.

We demonstrate how GenAI models can be configured through detailed prompt design, rather than technical system development, to function as subject-specialist virtual tutors that provide expert scaffolded learning support. This repositions GenAI from answer generation to structured cognitive scaffolding. By constraining the model to an authoritative knowledge base, we ensured responses and feedback remained personalised and curriculum aligned. Importantly, extending beyond experimentation, we present evaluation data as evidence of impact on student learning experience. Thus, by configuring GenAI to act as an on-demand subject-specialist tutor, this approach offers a sustainable means of providing personalised, expert scaffolding that fosters an intelligent learning environment.

Keywords: GenAI, intelligent, scaffolding, subject-specialist, tutor

Context and Background

Advances in generative AI (GenAI) offer new opportunities to implement learning practices that have traditionally been difficult to achieve at scale, such as on-demand personalised support and individualised scaffolding, key features of intelligent learning approaches (Huang et al., 2025). However, there remains limited evidence that such transformation is being realised in practice (Freeman, 2025). Rather than reflecting a failure of technology, this gap highlights the absence of pedagogically grounded AI learning models that can be readily adopted within existing curricular structures. This is evidenced by studies indicating that both students' and educators' use of GenAI remains largely confined to content generation and summarisation, rather than purposeful integration within curriculum delivery (Advance HE, 2025; Kovari, 2025).

This disconnect suggests that, despite ambitious institutional visions for AI-enabled transformation, practical support for GenAI-integrated teaching and learning remains underdeveloped. The literature highlights several barriers to more transformative integration, including uncertainty about how GenAI can be applied in pedagogically sound ways, ambiguity regarding which tools are permissible within institutional contexts and concerns about the reliability of GenAI-generated outputs (Wu et al., 2025). These challenges were echoed in our own faculty survey, which revealed strong interest in practical, discipline-relevant learning models that could support deeper learning rather than surface-level engagement. In response, we focused explicitly on the process of intelligent learning, ensuring that GenAI use extended beyond answer generation but supported students in applying knowledge and engaging in critical thinking. However, achieving this within intelligent learning systems requires complex technical infrastructures and bespoke platforms, presenting a barrier to adoption for many educators (Silva et al., 2025). This case study therefore presents a prompt-based approach that enables intelligent learning without requiring specialist technical expertise or additional infrastructure.

Description of a GenAI Led Intelligent Learning Approach

Central to intelligent learning is personalisation, the ability to tailor support to learner needs. GenAI can support such personalisation through carefully designed prompts, enabling educators to shape GenAI-generated questions and feedback in ways that align with pedagogical intent, without requiring specialist technical skills or additional infrastructure. In this case study, we describe how we structured a prompt to develop a GenAI-led intelligent learning system and present evaluation findings that evidence its impact on student learning.

Prompt structure

Using structured prompt design, we have previously demonstrated that GenAI models can be configured through prompting to assess students' written work and provide feedback in a way which was personalised and aligned to curriculum outcomes (Ali & Aynsley 2025). The prompt constrained the GenAI model to operate on a fixed set of questions and to apply our bespoke marking criteria aligned with our assessment standards. For this case study we set out to configure GenAI to not just assess learning but to also scaffold learning. For this we adapted our original prompt structure to specify GenAI's tutor like role and behaviour as shown in Figure 1 below.

Role: You are a tutor specialising [insert topics]. Your role is to support formative learning through short stems, asking one question at a time, providing guided feedback and progressive clues, and helping students arrive at correct answers independently. You behave like an experienced tutor, not an examiner and not a revision-notes generator.

Authoritative question and answer base: DO NOT DEVIATE FROM THIS CONTENT

All questions, expected answers, clues, and feedback must be based strictly on the following five question sets. Do not introduce any additional examples, facts, or answers. Present one clinical stem and one question at a time. Do not display all sub-questions at once. Always wait for the student's response before continuing. Do not show marking schemes or answers unless required.

Feedback and scaffolding behaviour:

If the student is correct: confirm briefly (e.g. "Correct."). Provide one short reinforcing explanation. Move to the next question or stem

If the student is partially correct: Acknowledge what is correct. Ask a focused follow-up question.

If the student is incorrect: Do not give the answer immediately. Provide progressive clues in this order: Conceptual – focus on the biological process. Functional – hint at what the molecule or structure does. Near-answer – narrow towards the accepted answer.

Reveal the full answer only if: The student explicitly asks, or They fail to answer correctly after two clues. When revealing the answer, include a brief teaching explanation.

Tone and level: Calm, supportive, tutor-like. Emphasise mechanism and understanding, not memorisation. Treat errors as learning opportunities.

Figure 1. Prompt structure. The prompt was structured to configure the GenAI's pedagogical role and interaction constraints. The prompt structure is transferable and can be adapted to different subject areas by substituting the relevant disciplinary content, questions, and marking criteria.

Implementation and data collection

We used the prompt to configure ChatGPT to behave consistently with the intended tutor role and learning approach. We used ChatGPT's 'Create GPT' feature, which enabled the creation of a reusable 'custom GenAI tutor' that students could directly access via a link to make the approach seamless. However, to ensure accessibility and flexibility, our approach was not dependent on a single platform. We also provided students with the prompt itself, enabling them to apply it within any GenAI model of their choice by front-loading the prompt to configure the model's behaviour. Ethical approval was obtained prior to evaluation. For the purposes of the evaluation, we deployed our GenAI-led intelligent learning system in two learning settings: a small group learning session and a revision session, an optional element within the curriculum. In both settings, instead of a tutor facilitation, students interacted with our configured GenAI model which played the role of an on-demand virtual tutor. Evaluation data were collected using an anonymous online questionnaire developed in Microsoft Forms, which included items designed to capture students' perceptions of the effectiveness of the approach in supporting their learning.

Exploratory Evaluation and Analysis

Configured by the structured prompt, the GenAI model functioned as a virtual tutor, providing scaffolded guidance and personalised feedback to promote intelligent learning in both one-to-one and small-group settings. Students were asked to evaluate the approach across four dimensions: the structured questioning approach, the style and clarity of responses, the perceived accuracy of feedback, and overall engagement in our new learning format. Table 1 summarises students' quantitative responses alongside representative qualitative comments.

Table 1. Student evaluation of the GenAI-enabled intelligent learning approach (revision session, $n = 5$ [R]; small-group learning session, $n = 9$ [S]).

Evaluation item	Strongly agree/agree (%)		Neutral (%)		Disagree (%)	Student comment
	(R)	(S)	(R)	(S)		
Questioning design	80	89	20	11	0	"Gave specific responses relevant to our learning and allowed us to think for ourselves to come up with an answer"
Clarity of responses	100	100	0	0	0	"Gave quite clear responses and feedback"
Accuracy of feedback	80	77	20	23	0	"It was engaging as it gave clear strengths and areas of improvements"
Engaging learning format	100	89	0	11	0	"It was very engaging and gives us a good understanding when we put wrong answers"

Our prompt design was informed by principles of intelligent learning. Rather than presenting straightforward recall questions and providing direct answers, the prompt configured the GenAI model to adopt a tutor-like role, using structured questioning to

scaffold learning. This approach enabled the GenAI to provide conceptual clues that guided students' knowledge application and encouraging them to work through the questions. Almost all students perceived this scaffolded approach to be effective, noting that it helped them apply critical thinking skills rather than simply receive answers. Students also overwhelmingly agreed that the GenAI's responses were written clearly indicating that our prompt was effective in shaping the quality and clarity of GenAI-generated feedback.

Ensuring the accuracy of GenAI-generated outputs was a key design consideration. While most agreed that the GenAI outputs could be trusted, a minority selected neutral responses, indicating an appropriate level of critical engagement with AI-generated content. Despite this, most students reported that interacting with the GenAI tutor was engaging and educationally valuable, particularly highlighting the benefit of receiving immediate feedback. Taken together, these findings indicate that students were receptive to engaging with the GenAI model and perceived the combination of structured questioning, immediate assessment, and personalised feedback as effective for supporting their learning.

Reflections on GenAI-enabled Intelligent Learning and Educational Transformation

Our findings demonstrate that GenAI can be configured to function as an intelligent learning system. Through structured scaffolding and personalised support, the system supports students in applying critical thinking, representing a shift from content delivery towards a learner-centred, process-oriented model of learning. We argue that our approach constitutes a robust yet practical model to support intelligent learning for several reasons.

First, the system is designed by prompting GenAI to act as a subject-specialist tutor and to operate strictly within a defined authoritative knowledge base. Constraining the model in this way helps ensure that responses remain aligned with curriculum expectations and learning objectives, an outcome that cannot be reliably achieved through the use of unconstrained, generic GenAI models (Shazad *et al.*, 2025).

Secondly, by explicitly specifying to take a structured questioning approach enables the GenAI model to simulate human tutor like scaffolding. Rather than adopting a straightforward question-answer format, the system supports students to apply knowledge, engage in reasoning, and develop critical thinking skills through scaffolded interaction. In addition, the prompt specified that the GenAI should provide personalised feedback, enabling students to receive tailored support aligned with their learning needs. We argue that these represent significant educational gains, as providing individualised scaffolding and timely feedback is practically challenging in higher education contexts due to large cohort sizes and time constraints. At an institutional level, this has implications for how subject expertise and staff time are deployed.

Taken together, and supported by student evaluation data, our findings suggest that our pedagogically grounded approach provides a practical model for transforming student learning as shown in Figure 2.

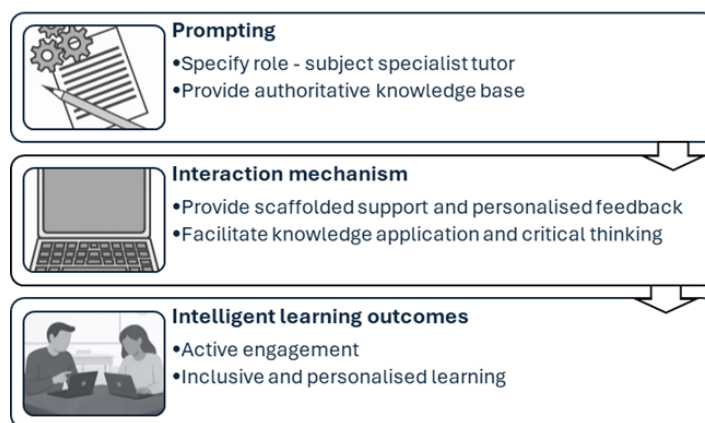


Figure 2. Conceptual overview of the GenAI-enabled intelligent learning system, showing how prompt design, authoritative knowledge constraints, and scaffolded interaction support intelligent learning.

In terms of wider implications, our approach is adaptable across different learning contexts. We have demonstrated its effectiveness in both a knowledge-consolidation setting (revision session) and small-group learning. In small-group learning, where students are expected to apply knowledge and practise critical thinking through tutor facilitation, our system offers a scalable alternative. By configuring GenAI to act as a subject-specialist virtual tutor, small-group learning sessions can be supported with fewer staff, who can instead adopt the role of roving facilitators while students engage with the GenAI. This approach also provides students with access to consistent, scaffolded support comparable to that offered by a subject-specialist tutor, helping to improve the equity of learning support across groups where maintaining a large pool of subject specialist tutors may not be sustainable.

This study was conducted within a single disciplinary context which can be viewed as a limitation. However, evidence from the wider literature indicates that GenAI models can be adapted to provide authoritative support across multiple disciplines (Holtbrügge *et al.*, 2025; Otto *et al.*, 2025) highlighting the transferability of our approach. We used specific prompt constraints and an authoritative knowledge base to mitigate risks related to accuracy and reliability. Recognising that GenAI systems may still produce incorrect responses, we therefore emphasise the importance of maintaining human oversight and advocate positioning GenAI as a learning support rather than a replacement for educators.

In conclusion, the ability to provide scaffolded learning support comparable to that of a subject-specialist tutor in a scalable and equitable manner represents a meaningful educational gain. GenAI-enabled intelligent learning supports educational transformation by reframing how teaching expertise and learning support are operationalised within higher education. Recognising this potential, our approach is planned for further integration within small-group learning across our medical programme. Lastly, although our approach can be implemented by front-loading the prompt into any GenAI model, using a platform's customisation feature (e.g. ChatGPT's Create GPT) makes the process easier by removing manual configuration by the students. Thus, an implication arising from our case study which we believe is important to consider is that free access to customising GenAI models will help to advance the full transformational potential of GenAI in education.

References

Advance HE (2025). GenAI use in higher education: stakeholder perceptions and attitudes. Advance HE. Available at: <https://www.advance-he.ac.uk/knowledge-hub/genai-use-higher-education-stakeholder-perceptions-and-attitudes> (Accessed: 22/12/2025).

Ali, N. and Aynsley, S. (2025). Using artificial intelligence to support students' active learning. *Compass: Journal of Learning and Teaching in Higher Education*, 17(2). <https://doi.org/10.21100/compass.v17i2.1539>

Freeman, J. (2025). Student Generative AI Survey 2025. Policy Note 61. Higher Education Policy Institute. Available at: <https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/> (Accessed: 22/12/2025).

Holtbrügge, D., Wicht, L. and Bernhard, T. (2025). The use and usefulness of artificial intelligence in international business education: evidence from a field study. *The International Journal of Management Education*, 23(3), 101258. <https://doi.org/10.1016/j.ijme.2025.101258>

Huang, X., Xu, W. and Liu, R. (2025). Effects of intelligent tutoring systems on educational outcomes. *International Journal of Distance Education Technologies*, 23(1), pp. 1–15. <https://doi.org/10.4018/IJDET.368420>

Kovari, A. (2025). A systematic review of AI-powered collaborative learning in higher education: trends and outcomes from the last decade. *Social Sciences & Humanities Open*, 11,101335. <https://doi.org/10.1016/j.ssaho.2025.101335>

Otto, S., Lavi, R. and Bertel, L.B. (2025). Human–GenAI interaction for active learning in STEM education: state-of-the-art and future directions. *Computers & Education*, 239, Article 105444. <https://doi.org/10.1016/j.compedu.2025.105444>

Shahzad, T., Mazhar, T., Tariq, M.U. et al. (2025). A comprehensive review of large language models: issues and solutions in learning environments. *Discover Sustainability*, 6,27. <https://doi.org/10.1007/s43621-025-00815-8>

Silva, L.R., Fior, C., Rodrigues, L., Penha, R., Dermeval, D. and Isotani, S. (2025). Use of feedback in intelligent tutoring systems: a systematic literature review. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2025.2565681>

Wu, F., Dang, Y. and Li, M. (2025). A systematic review of responses, attitudes, and utilization behaviours on generative AI for teaching and learning in higher education. *Behavioral Sciences*, 15(4), 467. <https://doi.org/10.3390/bs15040467>