4th Symposium on AI Opportunities and Challenges (SAIOC 2026)

After the bubble, a more mature appreciation of AI?

Booklet of Keynote Speaker Outlines and Presentation Abstracts

Contents

4th Symposium on AI Opportunities and Challenges (SAIOC 2026)

After the bubble, a more mature appreciation of AI?

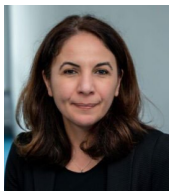
Keynote outlines and speaker bios

AI Beyond the Hype: Key Learnings and Emerging Opportunities

Artificial Intelligence is often portrayed as a recent technological revolution, driven by the intense interest and excitement surrounding its latest developments. Yet, it was first debated more than seventy years ago. Many of the foundational concepts underpinning modern AI such as neural networks, machine learning, and intelligent systems were developed decades ago. What distinguishes the current era is not the invention of AI itself, but the convergence of unprecedented computational power, advances in hardware, access to vast quantities of digital data, and the availability of scalable cloud infrastructure. Together, these developments have accelerated AI research and enabled capabilities that were previously impractical or impossible to achieve.

This keynote will explore AI beyond the hype, placing recent breakthroughs within their historical and technological context. Drawing on research and real-world applications across academia and industry, the talk will reflect on key lessons learned from the evolution of AI from early foundations to the rapid emergence of modern deep learning methods and generative AI. It will examine both the achievements and limitations of contemporary AI, highlighting what AI can realistically deliver through effective human–AI collaboration and identifying areas where significant challenges remain.

The keynote will also address critical issues surrounding data quality, transparency, ethics, governance, and the skills required to develop and deploy AI responsibly. It will also consider how researchers, developers and organisations can work together to maximise AI's potential while ensuring that innovation remains trustworthy, inclusive, and human-centred.



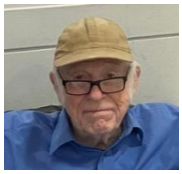
Professor Ella Pereira is a Professor of Computing at Edge Hill University, UK. Her research focuses on key enabling technologies, ranging from distributed services to the Internet of Things (IoT), including cloud, fog, and edge computing. Her work applies artificial intelligence and data analytics to health and other interdisciplinary application areas.

Education for life in the AI century – *using a cognitive antagonist*

The history of technology has largely been a story of the improvement of productivity through the improvement in speed or convenience or both. Of course, some would argue that medical technology although it is arguable that the relief and prevention of pain is a great convenience to humankind. But more generally productivity has been that main objective. Perhaps that is why we have become so concerned about the impact of AI as a productivity tool in addressing established educational processes that we have overlooked how AI might actually influence the very nature of education itself. What is needed is a much deeper look at what we want to achieve through education and how this could be used to allow individuals to have a more fulfilled life and to contribute more positively to a better society. When AI is placed under this microscope, we come to a number of quite different conclusions as to how it should be used in education and what we can expect its result to be from its participation in promoting learning.

There are real opportunities to use AI as a cognitive antagonist to trigger learning at a different level. If used correctly AI can provide the learner with challenges, prompts and auditing forcing the human mind to do the heavy lifting of synthesis and evaluation which is needed for real learning.

Maybe the use of AI in education can help us understand technology in a greater context that simple productivity.



Dr Dan Remenyi, an Extraordinary Professor at the University of the Western Cape has had a lifelong interest in AI having had an academic paper urging caution about its development published nearly 30 years ago. During his long career he has been a Visiting, Honorary or Extraordinary Professor at 10 different universities in the UK, Ireland, South Africa and Sweden. His work has been published in some 60 papers and 30 books. He hold a BSocSc, an MBA and a PhD.

Mind Meets Machine: Human Judgment and Responsibility in an AI-Shaped Society

As artificial intelligence becomes increasingly embedded in everyday life, public discourse often focuses on what AI can do: automate tasks, generate content, and optimize decisions. This keynote shifts attention toward a more fundamental question: *how should humans decide* when, why, and to what extent AI ought to be used?

Drawing on research and practice in higher education, online learning, and AI adoption, the keynote examines how individuals and institutions exercise judgment and responsibility while navigating competing priorities in an AI-shaped society, including tensions between efficiency and human development, innovation and responsibility, and automation and human judgment. Through selected real-world examples, it explores how decisions about AI use reflect broader assumptions about learning, work, and human flourishing. The keynote invites the audience to consider what should remain distinctly human, what responsibilities might be shared with intelligent technologies, and how these choices shape the futures we are collectively creating.

Dr. Connie Levina Yuen is an Associate Professor of Emerging Technologies in Open, Digital, and Distance Education at Athabasca University, Canada. Her research focuses on human-centred AI, ethics and governance, and AI-related competencies in higher education and the workplace, with an emphasis on equity, policy, and pedagogical design.

AI-Driven Fraud vs AI-Driven Detection: An Arms Race

Alexandru Adrian Blăgău and Cezar Scarlat
National University of Science and Technology Politehnica Bucharest, Romania
alex.blagau@gmail.com

Digital transformation has enabled significant improvements in organizations' processes, products, and services, delivering benefits such as increased efficiency, cost reduction, enhanced customer experiences, and greater scalability of operations. However, it has also led to the emergence of negative effects that affect both individuals and organizations. As dependence on technology increases, organizations and users become more vulnerable to cyberattacks, while increasingly complex digital infrastructures provide a broader attack surface for malicious actors. Digital transformation has redefined how organizations operate and interact with users, but it has also created a favorable environment for the development of increasingly sophisticated cyber fraud. The digitalization of financial processes, banking services, and e-commerce has led to a surge in online transactions, offering expanded opportunities for attackers to exploit both technological and human vulnerabilities. The adoption of new disruptive technologies, such as the Internet of Things, Artificial Intelligence, Big Data, Cloud Computing, and Blockchain, has significantly changed the way organizations interact with both customers and partners. This transformation of transactions, from physical to digital, brings major benefits, but also involves significant cybersecurity risks, increasing exposure to cyberattacks and fraud. The development of AI capabilities has created opportunities both for detecting and preventing cyber fraud and for committing it. The use of artificial intelligence enables the generation of highly convincing false content, such as deepfake, increasing the success rate of attacks. At the same time, process automation reduces direct human oversight, which can delay the detection of fraudulent activities. Artificial intelligence is taking an increasingly central role in detecting cyber fraud, making it possible to analyze massive volumes of data in real time and identify patterns that the human would easily miss. Machine learning algorithms can recognize anomalies in user behavior, flag suspicious transactions, and continuously refine their accuracy as they incorporate new data. In this paper, the authors analyze the "arms race" between the use of AI in committing cyber fraud and preventing it, highlighting ethical, technological, and social implications.

Keywords: digital transformation, artificial intelligence, cyber security, cyber fraud, prevention

Evidence-Bound Retrieval in a Materials Science Community: A Case Study

Manfred Bornemann and Gerhard Hackl
Intangible Assets Consulting GmbH, Austria
bornemann@ia-consulting.at

Abstract: Generative AI is currently (2026) applied in knowledge-intensive environments. In high-precision domains such as materials science, extensive documentation is available, containing relevant expert contributions. However, these documents follow no uniform conventions in terminology, structure, or detail and lack consistent metadata. As a result, full-text search across such collections often produces result sets that are either too broad to assess efficiently or too narrow to capture relevant evidence, requiring iterative refinement. The practical challenge is the systematic structuring and navigation of existing material.

This paper presents a case study developed within an industrial community of practice in materials science, based on more than 3,000 specialized documents accumulated over 20 years. The objective is to improve access to documented expertise and increase productivity for expert users. The approach uses large language models to identify relevant evidence in unstructured documents and extract it into a structured metadata layer. This process follows principles from information retrieval and applies LLMs as an operational tool for large-scale extraction.

Documents are indexed to allow precise reference to relevant passages, including page-level evidence. The system supports targeted navigation across the document base. Users identify relevant sources more efficiently and access them directly. Conversational interfaces based on large language models are used in an initial orientation phase to support search. Their role is limited to retrieval support and they are not used for problem-solving or as authoritative sources. Users are directed to original documents, including figures and images, where domain-specific interpretation takes place.

Initial validation with experienced users indicates improved retrieval efficiency and higher confidence in result relevance, especially for tasks requiring exact references. The system provides structured access to existing material and supports identification of relevant authors for further expert exchange. Current work examines the use of agents to support ongoing maintenance of the metadata layer and evaluates options for scalable integration in enterprise environments.

The case study shows how large language models can be applied to structure and access domain-specific evidence, focusing on navigation and usability of existing documentation.

Keywords: Evidence-based retrieval, Metadata extraction, Materials science documentation, Industrial community of practice

From Explaining to Engaging: How AI-Guided Tools May Transform Science Communication

Christine Flaßbeck Svenja Krämer, Livia Melanie Malkus, Karin Thier, Paulina Sophie Westing
IU International University of Applied Sciences, Germany
christine.flassbeck@iu.org

Despite increasing calls for accessible and engaging science communication, many researchers still struggle to translate their work into formats that resonate with non-expert audiences. A central challenge lies not in the lack of knowledge about effective communication principles, but in the absence of structured support for applying these principles in practice. As a result, science communication often remains either overly complex or insufficiently tailored to audience needs.

This paper introduces an AI-supported approach that reframes science communication as a guided decision-making process. Rather than using artificial intelligence primarily as a content generator, we propose a structured prompting system that actively supports researchers in making communication-relevant choices—such as audience adaptation, language style, and the balance between accuracy and accessibility. The tool is grounded in communication psychology and designed to function as a "thinking partner" that supports the development of public-facing communication formats.

In addition to outlining the conceptual framework and design principles of the tool, we propose an empirical evaluation strategy to systematically assess its effectiveness. The planned design involves comparing communication outputs developed with and without AI-supported guidance. These outputs will be evaluated by lay audiences in terms of perceived comprehensibility, engagement, and selected indicators of knowledge transfer. This approach allows for a systematic investigation of whether structured AI guidance can improve the quality and impact of science communication.

By shifting the focus from AI as a generator to AI as a facilitator of better communication decisions, this work contributes to a more differentiated understanding of the role of artificial intelligence in science communication. It further outlines a scalable approach for supporting researchers in translating their work for broader audiences. Implications for research, training, and public engagement are discussed.

Keywords: science communication, artificial intelligence, public engagement, decision support, audience understanding

Embedding Generative AI Within an MBA Management Consulting Progran

Ken Grant
Toronto Metropolitan University, Canada
kagrant@torontomu.ca

Management consulting is a professional discipline at the forefront of Generative AI (GenAI) related change. Not only is GenAI threatening the recruitment of individuals in the field, it is also changing the ways that consultants work as well as client expectations in how they do that work. This session will describe how a popular MBA specialization (three linked courses) on developing and practicing management consulting skills has been reinvented to embrace Gen AI. As opposed to resisting or policing the use of AI, the related course outlines state: "This course not only encourages the use of generative AI, it demands it". Over more than a decade this MBA specialization has prepared MBA students for consulting careers. Built around three key objectives (explaining consulting, understanding consulting careers, and doing consulting work) the course is a mixture of active learning, flipped classroom, dynamic classroom activities and self-assessment. It includes executing a significant project for a real client.

The presentation will describe how the course has been adapted to include:

- A pre-course survey of students' GenAI skills
- A combination of 8 mini-lectures, 6 short recorded lectures and 5 in-class workshops addressing GenAI in consulting
- In-class live use of GenAI to investigate key concepts in course content , with out-of-class follow up work
- A Chatbot-driven class case
- A laddered approach to developing AI skills, covering such topics as:
    - Prompt generation
    - Dealing with Gen AI in resume preparation and candidate evaluation
    - Ethical challenges, in general and in the consulting workplace
    - Unfriending your AI
    - Student's developing their own chatbot linked to course content
    - Using GenAI in consulting work
- Specific inclusion of GenAI requirements in major assignments
- Students' post-course reflection on their GenAI experiences in the course

The student feedback suggests that:

- Only a few MBA students considered themselves to have real GenAI competence (and even that appeared rather optimistic)
- Students' GenAI skills were significantly improved through the course work
- Students developed a clearer understanding of the limitations and challenges of effective GenAI use
- Most significantly, perhaps, many students reported that their view of GenAI had changed, Rather than using it as a tool to find answers, they now considered it to be a working partner, saving time, helping explore concepts, developing more structured approaches to data analysis

The session will also address how GenAi was embedded in course assignments/assessments and approaches to grading. Finally, lessons learned and opportunities for improvement will be discussed.

Keywords: Impact on professional work, AI skills development, ethical issues, critical thinking

New Era in Finance: Its Kernel Accelerated by AI

Paul Griffiths and Diego Arguello Ron
EM Normandie Business School, United Kingdom
pgriffiths@em-normandie.co.uk

Are we in an AI bubble?

It is reasonable to differentiate technology bubbles from financial bubbles. When either of these bubbles burst, many people lose money. However, technology bubbles such as the railways and the internet/dotcom leave assets of significant value to society, while financial bubbles such as crypto-assets leave nothing when they crash. There is little doubt that the AI bubble is of the technology type and will leave much value behind but in this paper our focus will not be on what infrastructure will be left when the AI bubble bursts. We will focus on visualising how the whole of finance will be transformed (Griffiths, 2025).

Effect of the AI bubble on finance and financial services
The integration of AI into financial services is reshaping how the industry operates at a fundamental level. Traditional Financial Intermediation Theory provides a useful lens to examine this transformation (Diamond, 1984; Allen, 1990).

Financial Intermediation Theory and the Three Pillars of Finance
Financial intermediation theory posits that intermediaries add value by reducing information asymmetries, pooling resources, and facilitating transactions in a way that markets operating solely on their own might not (Diamond, 1984; Allen, 1990). Within this framework, we identify three fundamental pillars of finance: Expectations (Borio, 2012; Kindleberger & Aliber, 2015); Risks (Markowitz, 1952; Basel Committee on Banking Supervision, 2011; Allen & Santomero, 1998); and Trust (Guiso et al., 2008; Diamond & Dybvig, 1983; Large, 2003).

The purpose of this presentation is to assess whether DeFi will materialise in the foreseeable future. It reinforces the highlighted three pillars of finance within the framework of financial intermediation theory and explores how AI enhances each pillar while ensuring compliance with evolving regulations. We address the question "Will the development of AI and its introduction to finance eliminate the need for financial intermediaries?"

AI's Enhancement of Expectations, Risk, and Trust
AI technologies are increasingly being deployed by banks, fintechs, and regulators to strengthen each pillar: Enhancing Expectations (Mullainathan & Spiess, 2017); enhancing Risk Management (Jorion, 2007; Consumer Financial Protection Bureau, 2019); and enhancing Trust (FINRA, 2020; Guidotti et al., 2018).

We conduct a Systematic Literature Review (SLR) covering 1 January 2008 to 11 February 2025, combining peer-reviewed research with high-authority "grey literature" from global standard setters and supervisors. The findings lead us to respond to the research question by concluding that traditional financial intermediaries will have to transform through the adoption of AI technologies but will not be replaced by the presence of those technologies in so-called DeFi models.

Keywords: AI in finance, Explainable AI, Financial Intermediation Theory, Risk Management, Systemic Literature Review

From Retail Innovation to Responsible AI: The Jean Coutu Experience

Izold Guihur
Université de Moncton, Canada
izold.guihur@umoncton.ca

This case study analyzes the long-term digital transformation of Groupe Jean Coutu and its responsible approach to integrating artificial intelligence (AI) within the highly regulated retail pharmacy sector. Rather than framing AI as a disruptive innovation, the case positions current AI adjacent practices as the continuation of more than five decades of technological and organizational innovation in Canadian pharmacy retail.

Founded in 1969, Jean Coutu differentiated itself early through innovations focused on accessibility, efficiency, and patient experience, including extended opening hours, competitive pricing, and a pharmacist owned franchise model. Over time, the organization invested heavily in centralized information systems, building a robust digital foundation well before AI emerged. A landmark initiative was the launch of Canada's first online prescription renewal service in 1999, signaling an early shift toward digitally enabled patient services.

This digital maturity later enabled the adoption of AI adjacent technologies. Following its acquisition by Metro Inc. in 2018, Jean Coutu gained access to enterprise scale data infrastructure and advanced analytics, supporting further standardization and optimization of pharmacy operations while preserving pharmacist autonomy. Integration with national digital health initiatives, such as PrescribeIT, improved data interoperability and reduced prescription errors, contributing to safer and more reliable intelligent systems.

Jean Coutu has deliberately avoided autonomous AI, instead adopting a human in the loop model in which intelligent systems support—but never replace—professional judgment. Current applications include automated prescription workflows, barcode verification, inventory optimization, smart lockers, and limited robotics pilots. These largely "invisible" AI adjacent technologies enhance efficiency and safety while maintaining clear accountability.

The case also highlights key constraints on AI adoption, including regulatory requirements, health data sensitivity, workforce acceptance, and the need to maintain public trust. Jean Coutu's incremental approach emphasizes privacy by design, transparent governance, and framing AI as a care enabler rather than a substitute for professionals. A comparison with U.S. pharmacy chains underscores the distinctiveness of Jean Coutu's governance first, professionally anchored strategy. The central insight is that effective AI in healthcare retail depends more on strong digital foundations, ethical governance, and alignment with professional practice than on advanced algorithms alone.

Keywords: Responsible AI in Healthcare Retail, Human in the Loop AI Governance, Long Term Digital Transformation, AI Adjacent Technologies, Pharmacist Centered Professional Autonomy

Responsible Human-AI Socialisation

Ali Hessami
Vega Systems, UK
a.g.hessami@ieee.org

This Presentation will cover the latest research outcomes on a Charter for Responsible Human-AI Socialisation. Artificial intelligence systems can be classified along a spectrum of autonomy and generality. On one end are narrow AI systems that provide specific outputs based on bounded inputs, operating as tools to augment human intelligence. On the other end is artificial general intelligence (AGI) and artificial super intelligence (ASI)–AI systems that can match or exceed human-level performance across a wide range of cognitive tasks. These properties make emerging AI technologies a unique and unprecedented development in human history that deserves a Precautionary approach to deployment.

The main focus of the talk is on the Autonomous Decision Making and Algorithmic Learning Systems and the rules for symbiotic relationship amounting to a Charter for Responsible Human-AI Mutual Socialisation.

Keywords: responsible socialisation, AI, Ethics, Charter

Agentic AI Graph Analysis of MITRE ATLAS Mitigations Across ML Ops

Thomas Heverin
The Baldwin School, USA
thomas.heverin@baldwinschool.org

This study presents a novel, student-driven approach to analyzing AI security mitigations by leveraging agentic AI and graph-based algorithms within the MITRE ATLAS framework. Conducted in an advanced high school ethical hacking course, the project explores how Replit-based AI agents can be used to model, analyze, and prioritize mitigation strategies across the machine learning (ML) operations lifecycle. The work bridges applied cybersecurity, artificial intelligence, and network science, demonstrating how emerging tools can support structured reasoning about adversarial AI risks.

Using mitigation data derived from MITRE ATLAS, students constructed a graph representation in which nodes correspond to mitigations and edges capture relationships such as shared coverage of adversarial techniques. This graph was then segmented according to key ML Ops lifecycle phases, including data preparation, machine learning engineering, evaluation, deployment, and monitoring. Replit AI agents were used to automate the execution of ten graph algorithms across each lifecycle phase, enabling scalable and repeatable analysis without requiring extensive manual coding.

The selected algorithms span three primary analytical categories: importance (e.g., degree centrality, closeness centrality), structural weakness (e.g., betweenness centrality, chokepoint analysis, critical node detection), and adversarial decision-making (e.g., k-most vital nodes, shortest path interdiction, sequential interdiction, and defender–attacker–defender modeling). Each algorithm provides a distinct lens for evaluating the role of mitigations within the broader security landscape, from identifying highly connected controls to uncovering critical dependencies and potential points of failure under attack scenarios.

Results reveal that the relative importance and effectiveness of mitigation strategies vary significantly across ML Ops lifecycle phases. Technical cyber mitigations tend to dominate in early and late stages, such as business understanding and deployment, where system-level protections and access controls are most critical. In contrast, technical ML mitigations are most influential during the evaluation phase, where robustness testing and model validation play a central role. Policy-based mitigations emerge as particularly important during the machine learning engineering phase, highlighting the role of governance, standards, and procedural controls in shaping secure development practices. Mixed mitigation strategies are most effective during data preparation, where both technical and policy considerations intersect.

This work demonstrates the feasibility and value of combining agentic AI with graph analytics to support structured analysis of AI security frameworks. By enabling students to engage directly with complex, real-world datasets and advanced analytical techniques, the project also highlights the potential for experiential learning in AI.

Keywords: Agentic AI, Graph Algorithms, MITRE ATLAS, ML Ops Lifecycle, AI Security Mitigations

From Analytics Systems to Knowledge Sharing: Psychological and Workflow Mechanisms

Kristina Jespersen
Aarhus BSS, Aarhus University, Denmark
kjespersen@econ.au.dk

Artificial intelligence (AI) and analytics systems are increasingly embedded in organizational work, promising to enhance knowledge sharing and decision-making. However, prior information systems (IS) research has predominantly conceptualized system success in terms of use and user satisfaction, offering limited insight into how analytics systems translate into meaningful collaborative outcomes. Addressing this gap, this study extends IS success research from use and satisfaction toward collaborative capability outcomes by examining the post-adoptive value realization mechanisms through which analytics systems enable knowledge sharing.

Building on post-adoptive IS use theory and psychological empowerment perspectives, the study conceptualizes analytics value realization as a dual-mechanism process. We argue that analytics systems create value primarily by enhancing users' psychological job value—defined as employees' perceptions of their competence, effectiveness, and contribution in their work roles. This psychological mechanism is complemented by a structural mechanism, workflow integration, which reflects the extent to which analytics use becomes embedded in daily work practices and enables the enactment of system capabilities.

Empirically, the study draws on survey data from 100 users of a standardized business intelligence (BI) interface deployed across multiple organizations and functional domains, providing a natural laboratory for examining post-adoptive processes under consistent technological conditions. Using partial least squares structural equation modeling (PLS-SEM), the results show that BI system quality and trust do not directly influence knowledge-sharing capability. Instead, their effects are primarily mediated through psychological job value, which emerges as the strongest predictor of collaborative capability. Workflow integration plays a complementary role by enabling users to apply analytics in their tasks, thereby supporting—but not substituting—the psychological value mechanism.

The study makes three contributions. First, it extends IS success research by demonstrating that analytics systems create value not merely through use or satisfaction, but through enhancing employees' collaborative capabilities. Second, it identifies psychological job value as a central mechanism in AI-enabled work transformation, highlighting how analytics systems reshape users' perceptions of competence and contribution. Third, it shows that value realization depends on both human-centered and structural processes, emphasizing the importance of supporting users in integrating analytics into their work practices.

Overall, the findings suggest that organizations seeking to leverage AI and analytics should prioritize not only system quality but also the design of user experiences that enhance employees' psychological job value, thereby enabling more effective knowledge sharing.

Keywords: AI, psychological job value, knowledge sharing capability, analytic system quality

Beyond Panacea: Digital Governance, AI, and the Political Economy of Public Benefit in African Development

Alieu Stephen Kafoe
Marymount University, United States
stephenkafoe@yahoo.com

This conceptual and critical analytical review examines who benefits from artificial intelligence (AI) governance frameworks in Sub-Saharan Africa and identifies the institutional conditions under which digital governance reforms generate public value rather than enabling elite capture, rent-seeking, and technologically mediated dependency. Drawing on a critical synthesis of peer-reviewed scholarship published between 2020 and 2025 and institutional evidence from major multilateral and practitioner sources, the study integrates four analytical lenses: decolonial and digital-colonial critique, public administration and policy learning, political economy of procurement and contracting, and Ubuntu-informed normative epistemology. The analysis is further anchored in four theoretical frameworks: polycentric governance, rent-seeking political economy, digital colonialism, and institutional dualism, to explain why AI governance in low-capacity states is structurally prone to capture. The study demonstrates that AI governance outcomes depend less on the formal elegance of national strategy documents than on the interaction between state capacity, procurement incentives, data governance, and informal institutions. Where these conditions are unfavorable, AI governance produces an illusion of efficiency: measurable process improvements coexisting with deepening exclusion, opacity, and the redistribution of informational and contractual power toward political elites, donor intermediaries, and international vendors. The paper's principal conceptual contribution is the AI Governance Outcome Matrix, a two-dimensional diagnostic framework that maps likely governance trajectories across four quadrants defined by state capacity and institutional alignment. The matrix identifies simultaneous reform along both dimensions as the necessary condition for durable movement toward public-value-creating governance. Ubuntu philosophy is engaged as a substantive governance epistemology rather than as rhetorical decoration, providing a normative counter-framework to Western-centric AI ethics models by reconceptualizing data as communal stewardship, algorithmic accountability as collective obligation, and public benefit as relational flourishing. The article concludes with ten actionable recommendations spanning open procurement, infrastructure sequencing, distributed oversight, participatory design, and iterative policy learning, directed at policymakers, development finance institutions, and governance practitioners operating within the 2025-2030 African Union Continental AI Strategy implementation window.

Keywords: AI governance; Africa; public interest; elite capture; rent-seeking; digital colonialism; polycentric governance; Ubuntu; state capacity; procurement; informal institutions

Artificial Intelligence in Financial Reporting and Business Decision-Making: Opportunities, Challenges and Professional Perspectives

Suzi Mikulić[1] and Ryosuke Nakajima[2]
[1]University of Mostar, Suzi Mikulić
[2]Tokyo Business and Language College, Japan
suzi.mikulic@gmail.com

Abstract: The rapid development and application of artificial intelligence (AI) are significantly transforming business processes and financial reporting in companies. In the context of global digitalization and increasing data availability, companies face the need for faster and better decision-making, making financial reporting even more strategically important. As a key source of relevant financial information, it plays an important role in supporting business decisions, while the integration of AI technologies changes the way financial data is created, analysed and interpreted. The application of technologies such as machine learning, big data analysis and robotic process automation (RPA) enables the automation of accounting activities, reduction of errors, increased data accuracy, and greater efficiency and speed in decision-making. In addition, artificial intelligence supports the development of advanced analytics and predictive models that contribute to improved strategic planning.

However, its application raises several issues and practical challenges, including the transparency of algorithms, reliability of results, compliance with legal regulations, data security and ethical responsibility. A particular challenge is the so-called "black box" effect, where users do not have complete insight into how algorithms make decisions, which can affect the level of trust in the system. Furthermore, the need to comply with regulatory requirements and maintain professional judgment in decision-making is becoming increasingly evident.

This paper examines the role of artificial intelligence in financial reporting and business decision-making, focusing on the perceptions of accounting professionals who work with this information daily and are directly affected by changes in accounting practice. The research will employ a quantitative approach using a survey questionnaire administered to accountants in both the private and public sectors. The objective is to assess their attitudes, perceived benefits and risks, and the influence of AI on accounting processes, as well as its effect on the quality of financial reporting and decision-making. Data collection is ongoing, and the findings will be presented at the symposium.

The data will be analysed using descriptive statistics. The paper aims to provide an overview of the actual impact of artificial intelligence on current business practices and highlight the importance of balancing technology with professional expertise.

Keywords: artificial intelligence, financial reporting, business decision-making, accounting, digitalization, risks

Deepfake Detection in the Age of GenAI: A Multimodal Approach to Trust, Security, and Governance

Kasey Miller, Karl Ricanek, Anastasiia Iufereva, Peter Mozelius
University of North Carolina Wilmington, USA
millerkc@uncw.edu

The rapid advancement of generative artificial intelligence has significantly increased the sophistication, accessibility, and impact of deepfakes, creating urgent challenges for cybersecurity, digital trust, and information integrity. As synthetic media becomes increasingly indistinguishable from authentic content, organizations face growing risks related to misinformation, fraud, and identity manipulation.

This research explores emerging approaches to deepfake detection by integrating facial analytics, computer vision, and biometric-based artificial intelligence models. It introduces a multimodal detection framework that combines spatial–temporal facial analysis, audio–visual consistency validation, and embedding-based anomaly detection to identify inconsistencies in synthetic media. The approach emphasizes resilience against evolving adversarial techniques and highlights the importance of moving beyond single-modality detection methods.

In addition to technical considerations, the session examines key challenges, including dataset bias, model generalization, and the ongoing arms race between generative and detection technologies. The presentation also addresses governance and operational implications, focusing on how organizations can integrate deepfake detection into cybersecurity workflows while maintaining ethical and trustworthy use of artificial intelligence.

The discussion contributes to broader conversations on balancing innovation with safeguards, positioning deepfake detection as a critical capability within the evolving landscape of AI opportunities and challenges. Attendees will gain practical insights into current detection strategies, limitations, and future directions for securing digital environments against synthetic media threats.

Keywords: Deepfakes, Deepfake detection, GenAI, Misinformation, Synthetic media

Deepfakes challenging the communication ecosystems - Toward a Policy Framework for Synthetic Media Governance

Peter Mozelius, Anastasiia Iufereva, Kasey Miller, Karl Ricanek
Mid Sweden University, Sweden
Peter.Mozelius@miun.se

The proliferation of deepfakes, enabled by advances in generative artificial intelligence, presents a growing challenge to communication ecosystems, democratic processes, and public trust in digital information. This study examines deepfakes through a communications and policy lens, focusing on how synthetic media disrupts information integrity, amplifies dis/misinformation, and complicates the verification practices of journalists, institutions, and the public. Concurrently, emerging AI-based detection tools have been developed to distinguish synthetic from authentic content—yet this remains a dynamic and unsettled contest. The race between generative and detection technologies is ongoing, and deepfakes have evolved into a multimodal phenomenon that has outpaced existing governance structures. While technical detection methods continue to evolve, gaps remain in governance frameworks, regulatory alignment, and institutional readiness to address the societal impacts of deepfakes.

Drawing on interdisciplinary insights from a scoping review spanning communication studies, information policy, and cybersecurity, this study analyses prevailing approaches to deepfake mitigation and surfaces critical deficiencies in transparency, accountability, and cross-sector coordination. From a comparative global perspective, regulatory responses vary considerably: the California AI Transparency Act and the EU AI Act represent substantive—if imperfect—policy efforts, yet much of the world remains without meaningful guidance or enforceable standards.

Authors propose a policy-oriented framework emphasizing three pillars: (1) standardized disclosure and labeling mechanisms for synthetic media, (2) institutional capacity-building for verification and media literacy, and (3) collaborative governance models involving public agencies, technology platforms, and academia. Moreover, the study has explored the ethical implications of detection technologies, including risks of overreach, bias, and unintended censorship. By situating deepfake detection within broader communication policy debates, this research contributes to a more holistic understanding of how societies can balance innovation with the protection of information integrity and public trust.

Keywords: Deepfakes, Deepfake detection, Deepfake policy, Artificial intelligence, GenAI, Scoping review

If the AI bubble bursts, what will be the difference in higher education and lifelong learning?

Peter Mozelius, Marcia Håkansson Lindqvist, Martha Cleveland-Innes, Jimmy Jaldemark
Mid Sweden University, Sweden
Peter.Mozelius@miun.se

What are the indicators suggesting that Artificial Intelligence has created a socioeconomic activity bubble that may be unsustainable? As pointed out by Floridi (2024) and Youvan (2025), the current AI hype shows several characteristics that resemble earlier tech bubbles such as the Dot-Com Bubble and the Cryptocurrency Bubble. These two bubbles were also centred around a technology with a potential to revolutionise multiple sectors, and at the same time had a tendency to cause tunnel vision. The general pressure to meet high expectations, contradicted by reports on technological challenges and AI slop (Ansari, 2025), raise concerns about sustainability. What impact will the loss of this intense activity have on social, economic, and educational structures worldwide? If the AI bubble bursts there could be drastic consequences across tech industry, job market, and the general economy (Youvan, 2025). Less has been published about how a bursting AI bubble might have a strong impact on different educational contexts. This presentation will review, analyse, and discuss evidence regarding the impact of such a change on the fields of higher education and lifelong learning.

Significant challenges in higher education and lifelong learning will remain. The use of GenAI in these education areas may be immune to the decline of interest, investment, and use of GenAI elsewhere. The need for rethinking assessment and course design (Perkins et al., 2025) related to GenAI use in higher education would remain.There is still a need for concrete guidelines on how educators and student might ethically use generative AI (GenAI) in teaching and learning activities. Regarding lifelong learning, the need for an increase of lifelong learning has increased in modern society, creating new challenges for higher education to achieve lifelong learning. As lifelong learning includes learning throughout the entire life span, it can be seen as a way to strengthen individuals' opportunities to participate actively in society and to develop democratic and personal competencies. Here, the iterative improvement of GenAI tools could support flexibility individualisation, accessibility and individualisation for students. Further, lifelong learning is often emphasised in relation to the labour market's need for a well-educated and competent workforce (Jaldemark; 2021; OECD, 2021) of which competences in tools will be necessary in the future.

If the AI bubble bursts, it may result in an interrupted development of GenAI tools. The existing toolbox works well to at least partly solve many types of traditional higher education assignments for students. The quality of speech synthesis, audio to text conversion, and tailored instructions are today much higher than just a few years ago. Students' use of chatbots as artificial tutors and study buddies will not be ended by a burst AI bubble.

Keywords: Artificial Intelligence, AI, AI bubble, GenAI, Higher education, Lifelong learning

When AI Thinks for Us: Measuring Cognitive Offloading at Work

Calvin Nobles
University of Maryland Global Campus, USA
calvin.nobles@umgc.edu

As artificial intelligence becomes increasingly embedded in everyday work, professionals now rely on large language models, code assistants, and decision-support systems to draft content, summarize information, and support problem-solving. Although these tools can enhance speed and productivity, they may also alter how users allocate attention, memory, verification, and judgment. This presentation introduces a human-centered analytics framework for quantifying cognitive offloading in AI-assisted work, defined as shifting portions of memory, judgment, or problem-solving to an external aid. Cognitive offloading is not inherently harmful, but it becomes concerning when verification declines, revision effort decreases, and unaided performance worsens.

Using a Design Science Research approach, the study developed and evaluated an artifact composed of three integrated elements: a synthetic behavioral dataset, a machine learning workflow, and a Gradio-based interactive interface. The dataset included 1,500 task-session records, 40 variables, and 25 occupations where AI use is plausible or increasingly common. Feature layers captured participant background, work context, task characteristics, AI interaction traces, and human factors outcomes. The analytical pipeline used a RandomForestRegressor to estimate a continuous Cognitive Offloading Index and a RandomForestClassifier to predict a categorical Cognitive Offloading Level.

Results showed substantial variation in AI reliance behavior. The mean Cognitive Offloading Index was 45.84, the average verification rate was 0.42, the average performance drop without AI was 0.30, and the average trust in AI score was 54.45. Occupation-only models performed poorly, indicating that job title is a weak proxy for cognitive offloading. In contrast, full-feature models achieved near-perfect predictive performance, showing that the strongest explanatory signal lies in behavioral and human factors variables. The most important predictors were verification rate, AI use frequency, acceptance rate, performance drop without AI, edit distance, and verification time.

This presentation argues that verification is the key variable separating adaptive augmentation from maladaptive dependency. It demonstrates how explainable machine learning can make cognitive offloading more visible, interpretable, and actionable as AI becomes more deeply embedded in work and education.

Keywords: Cognitive Offloading, AI-assisted Work, Behavioral Analytics, Explainable Machine Learning, Human-centered Design

Algorithmic Bias and Discrimination: Why Technical Fixes Alone Are Not Enough

Calvin Nobles, Darrell Burrell, Marwan Omar
University of Maryland Global Campus, USA
calvin.nobles@umgc.edu

Artificial intelligence systems increasingly shape high-stakes decisions in employment, credit, healthcare, criminal justice, and public services. Although these systems are often promoted as objective, efficient, and scalable, they can also reproduce and intensify existing forms of inequality. This presentation argues that algorithmic bias is not merely a technical defect that can be corrected through better models, cleaner code, or larger datasets. Rather, it is a socio-technical and institutional problem rooted in historical inequities, flawed design assumptions, limited participation, and weak accountability structures.

The presentation examines how algorithmic systems generate discriminatory outcomes through multiple mechanisms, including biased training data, historically unequal source conditions, and design choices that privilege efficiency over justice. It also highlights a more novel concern: AI systems can create new forms of discrimination by grouping individuals into opaque computational categories that do not neatly align with traditional legally protected classes. As a result, existing legal and policy frameworks may be insufficient to identify or remedy emerging harms.

Drawing on examples from employment, healthcare, and financial services, the presentation shows how algorithmic bias can deny opportunity, misallocate resources, and scale inequity across already vulnerable populations. It further contends that technical mitigation strategies, while valuable, remain incomplete when they are not paired with ethical oversight, participatory design, continuous auditing, and institutional accountability. In this view, fairness cannot be treated as a downstream adjustment or compliance exercise. It must be established as a foundational principle in the design, deployment, and governance of AI systems.

The presentation concludes by advancing five pathways for action: data equity and representation, participatory design, legal and institutional accountability, continuous auditing and monitoring, and transparency with explainability. Together, these pathways offer a more comprehensive and just framework for addressing AI inequity. The central claim is clear: technical solutions are necessary but not sufficient. Achieving fairness in the AI era requires systemic intervention across technical, legal, organizational, and social domains.

Keywords: Algorithmic bias, Artificial Intelligence, Ethics, Discrimination, Participatory Design, Accountability

Artificial Intelligence (AI): Between Words and Action

Dan Ophir
Computer Science, Ariel University, Israel
dano@ariel.ac.il

At first, I would touch the origins of AI: self-learning algorithms, acquiring experience, and improving itself after each iteration. There were so-called Expert Systems, receiving their initial input from a human expert, next appeared the Neural Networks and Deep Learning systems.

Nowadays AI might be classified, similarly to human IQ, into the following classes:

a. Verbal Intelligence, which is mainly expressed by human linguistic communication, supported by huge amounts of data, and impressive, in real time its manipulation, which I/O is mainly textual. Its aim is to supply information and to support making decisions, supported by the Data appearing in the world Knowledge Data Bases. The commercial applications are called generally, after the character of human – computer conversations "Chat", such as Chat-GPT. LLM - Large Language Model is the methodology supporting those systems.

b. Executive Intelligence is module with mainly digital I/O – representing communication with hardware devices. These instructions are obtained to control and optimize the device's performance using the classical methods mentioned above, Besides the textual I/O the Executive Intelligence of AI may deal with GUI – for example create dynamic scenarios with supplied heroes. Additional advances domain dealt by AI is programming simple codes in a programming language. However, there are still problems to difficult to be solved by the Executive Intelligence of AI, for example, the CPM problem.

The future challenge might be to examine AI by the standard Wexler measuring IQ test, which can distinguish between the Verbal and Execution Intelligence.

Another performed test is checking what does the AI "think" about the above issue of differentiation between the above to AI abilities: verbal and executive ones. Its answer is hereby:
". Bottom line

- Verbal ability in AI = strong, fluent, often impressive
- Executive ability in AI = still limited, especially in open-ended real-world tasks"

Detailed comparison between the two above aspects will be shown in other formats like: article lecture with the support of examples.

Keywords: Deep Learning IQ, Verbal Intelligence, Executive Intelligence, CPM - Critical Path Method, Wexler IQ Test

When Heavy Metal Meets Generative AI: Theoretical Promise vs. Organizational Reality

Maria Pajuoja, Emilia Wietrak, Matti Pihlajamaa, Sebastian Bouschery
University of Vaasa, Finland
maria.pajuoja@uwasa.fi

Generative Artificial Intelligence (GenAI) is emerging as a general-purpose technology reshaping innovation processes, supporting activities from early-stage discovery to concept prototyping and refinement (Calvino et al., 2025; Füller et al., 2022; Gama & Magistretti, 2025). It is expected to reduce development time and cost while lowering barriers to participation in innovation (Bilgram & Laarmann, 2023; Wang & Wu, 2025). However, early evidence suggests a gap between these expectations and their systematic realization, particularly within traditional manufacturing (Hughes et al., 2026). This shifts research attention to where and under what conditions GenAI can be effectively integrated into new product development (NPD).

We investigated these questions through two in-depth case studies of global manufacturing companies in Finland. Both firms are multinationals depending on R&D for competitive advantage, yet their NPD processes span years, involve millions of euros in investment, and carry high costs of error. Their products consist of hundreds of interdependent components, making expert judgment and historical knowledge essential. We conducted 18 interviews with R&D experts, complemented by observations and documents from a five-year collaboration. Following Gioia et al. (2023), we conducted an inductive analysis to identify aggregate dimensions of GenAI integration.

Our results show that GenAI use is currently concentrated in peripheral, low-risk activities, such as editing language and summarizing reports, with limited impact on core innovation outcomes. We explain this as a consequence of socio-technical misalignment across four conditions: organizational arrangements, task structure, user practices, and technological constraints. Internal company policies restrict tool choice, while performance metrics often prioritize immediate delivery over the experimentation required for AI adoption. Furthermore, cross-functional silos limit the availability of the structured, reusable knowledge that these tools require to be effective.

Our study makes two primary contributions. First, we highlight task structure as a critical boundary condition. Existing studies often focus on standalone innovation tasks where design decisions are easily reversible (Bouschery et al., 2023). We show that integration is far more complex when decisions are tightly interconnected and costly to correct (Thomke, 1998; Ulrich, 1995). Second, we challenge accounts that emphasize individual-level literacy as the primary lever for value (Chui et al., 2023). While such efforts are helpful (Dell'Acqua et al., 2026), they do not resolve structural misalignments. We argue that for traditional manufacturing, successful integration depends less on marginal skill upgrades and more on broad organizational redesign.

Keywords: generative AI, manufacturing, new product development, socio-technical alignment

From Role-Play to Real Impact: A Customisable Bodyswaps AI Simulation Case Study on Driving Soft Skill Development in Employment Relations

Madeleine Pickles
Liverpool John Moores University, England
M.Pickles@ljmu.ac.uk

The study examines the integration of Bodyswaps AI simulations within a postgraduate module focused on redundancy consultation and dispute resolution. A key strength of this approach is the ability to customise scenarios to align with module content, learning outcomes, and industry contexts. Simulations were tailored to reflect realistic employment relations challenges, such as collective consultation and trade union negotiation, ensuring discipline-specific relevance and strong pedagogical alignment. Drawing on student feedback and recorded simulation interactions, the research evaluates how AI supports reflective and experiential learning, interpreted through Kolb's (1984) experiential learning theory and Gibbs' (1988) reflective cycle.

Findings indicate clear improvements in learner confidence and self-awareness. Most participants reported increased confidence in negotiation skills, while qualitative feedback highlighted behavioural insights such as the importance of acknowledging emotions and demonstrating compassion. These outcomes suggest that Bodyswaps simulations effectively develop emotional intelligence, a critical competency in employment relations (Katz and Sosa, 2015).

Compared to traditional approaches, students reported higher engagement and perceived value, describing the simulations as "eye-opening" and "an intelligent role player." This aligns with research showing that immersive learning environments enhance engagement and enable safe practice of real-world skills (Marlow et al., 2017; Stevens, 2024). Additionally, real-time feedback reflects best practice in formative assessment, which is essential for developing interpersonal skills (Shute, 2008). The ability to adapt scenarios further strengthens relevance and inclusivity by allowing educators to respond to changing organisational and legal contexts.

However, limitations were identified. Some participants experienced technical issues, including speech recognition errors and a lack of "humanistic cues." These concerns reflect broader limitations of AI in replicating emotional nuance and highlight the risks of over-reliance on automated systems. The study therefore emphasises the importance of a blended approach, where AI complements rather than replaces human interaction.

Overall, this case study demonstrates how AI simulations can bridge the gap between theory and practice in employment relations education. Students developed both negotiation strategies and essential soft skills such as empathy, active listening, and confidence in managing conflict. This supports wider evidence that future workplaces require a combination of digital and interpersonal competencies (World Economic Forum, 2023). While AI cannot replace human interaction, it can significantly enhance experiential learning when thoughtfully integrated into pedagogical design.

Keywords: AI simulations, experiential learning, soft skills, emotional intelligence, employment relations

Human–AI Bond Formation, Overtrust, and Dependency: A Multifactorial Framework

Angelica Ortiz de Gortari
Centre for the Science of Learning & Technology, Norway
angelica@gametransferphenomena.com

Human–computer interaction has evolved beyond purely instrumental use to include a significant affective dimension, presenting both novel opportunities and challenges. Emotional attachment and dependency on interactive technologies have been documented since the early days of digital companions (e.g., Tamagotchi) and persist in modern forms such as virtual avatars and voice assistants (e.g., Alexa). Recent advances in artificial intelligence, particularly large language models (LLMs) enable context-sensitive, adaptive responses that can appear emotionally attuned. These capabilities can foster a sense of reciprocity and, for some users, the perception of a social relationship.

This presentation outlines a multifactorial framework that illustrates how technological design and psychological processes interact. These dynamics drive people to overtrust AI, misattribute its capacities, and rely on AI for guidance or companionship, which increases their vulnerability to manipulation and dependency. This framework examines relational AI systems, including AI companions, which are technologies designed for ongoing, personalized, and socially responsive interactions that that foster perceived social bonds over time. The framework draws on research in human–computer interaction, media psychology, and cognitive science.

Four interacting dimensions are identified: (1) Technological affordances and design features, including personalisation, immediacy, consistent responses, and emotionally responsive outputs, that reinforce engagement and shape user expectations. (2) Users' perceptions of technology, including self-efficacy, anthropomorphism, and the attribution of social agency, that play a key role in shaping positive attitudes toward AI. (3) Media-induced experiential processes include shifts in self-perception (e.g., the Proteus effect, identification) and presence-related states (e.g., immersion, telepresence, transportation).

These processes, alongside neurocognitive responses to simulated agents, can intensify perceived relational depth and embodiment. (4) Individual cognitive and dispositional traits, including cognitive biases, source monitoring errors, animistic thinking, and fantasy proneness, can increase vulnerability by shaping how AI outputs are interpreted and attributed, especially in contexts where the boundaries between simulation and reality become less distinct.

Mitigating vulnerabilities in AI bond formation requires embedding safeguards into LLMs. For instance, systems can be engineered to acknowledge uncertainty, prompt alternative interpretations, and direct users toward human support when distress or delusional content is detected. Strengthening users' contextual skills, such as the ability to differentiate between human and AI communication and to recognise persuasive or self-referential cues. These efforts can support a more adaptive and healthier human–AI relationship.

Keywords: Human–AI interaction, Large language models (LLMs), Conversational systems, Responsible AI use, AI-attachmen

AI in Educational Leadership Practice

Jussara Reis-Andersson
Mid Sweden University, Sweden
jussara.reis-andersson@miun.se

School leaders play an important role in schools as they are responsible for both pedagogical development and administrative management. They guide and support teachers in improving teaching and learning while also ensuring that the school functions efficiently. At the same time, artificial intelligence (AI) is shaping leadership practices across different sectors, including education. For school leaders, the growing presence of AI brings both challenges and opportunities and highlights the importance of digital competence. Because school leaders are directly involved in everyday decision-making that influences teaching and learning, understanding how AI affects their practice has become increasingly relevant.

The aim of this study is to explore how school leaders perceive the impact of AI on their professional practice. The research question is: How do school leaders perceive the impact of AI on their professional practice as school leaders? Two interviews were conducted with school leaders who lead primary schools from grades F-6. Each lasted about 45 minutes and was conducted online. They were recorded with consent and carefully transcribed. The school leaders are coded SLA and SLB. Thematic analysis (Braun & Clarke, 2021) was used to structure the data, and the Cynefin framework (Snowden & Boone, 2007) inspired the interpretation of complex contexts, particularly regarding the uncertainty in integrating AI.

Analysis identified two main themes. The first theme concerns AI as a tool. AI supports the collection, structuring, and analysis of data, but it is always the school leader who makes the decisions. SLA explained, “I would never use it to answer an email from a parent; that is something I want to do myself.” AI can provide accurate summaries and suggestions, but it can never replace human judgment or understand the softer aspects of leadership. As SLB noted, “AI can help me gather and structure data, but I still have to be critical and check everything,” highlighting the need for oversight and reflection.

The second theme concerns AI’s impact on leadership and focus. AI reduces administrative work and allows school leaders to focus more on pedagogical leadership. It works as a tool and sometimes as a reflective conversational partner, though its influence is limited to hard facts and numbers. SLB described it as “like having an assistant principal to bounce ideas with, but it only works with hard facts.” AI can make administrative tasks more efficient, giving leaders time to concentrate on teaching and learning priorities. SLB explained that “AI might make administrative work easier, giving me more time for pedagogical leadership.”

The conclusion is that, even if AI seems efficient, critical review and experience are still needed to interpret the results correctly. Another important point is that AI can never replace human judgment, but it can make work more efficient and support data-based decision-making.

Keywords: Artificial Intelligence (AI), Data-Driven Decision Making, Digital Competence, Educational Management, School Leadership

DIFA: AI-Assisted Formative Feedback for Scalable Pedagogy

Raja Khurram Shahzad, Peter Mozelius and Erik Ström
Mid Sweden University, Sweden
raja-khurram.shahzad@miun.se

The Draft-Based Iterative Feedback Accelerator (DIFA) is an AI-enhanced extension of the FAMS [1], which is designed to address challenges in providing scalable, high-quality formative feedback in higher education. Formative feedback is crucial for metacognitive development and self-regulated learning, yet its effective implementation is often limited in large cohorts, particularly for neurodiverse learners who need structured and transparent feedback. Moreover, the rise in students' reliance on generative AI raises concerns about deep learning, highlighting the need for assessment designs that promote meaningful cognitive engagement. DIFA transforms AI from a passive tool into an active pedagogical assistant within the feedback loop. It allows students to submit ongoing, incomplete work for evaluation, shifting focus from outcome-oriented grading to a process centered on reflection and continuous improvement. DIFA features reusable feedback fragments and a repository of common errors.

DIFA offers a modern approach to feedback by processing student drafts to generate context-specific responses that instructors can refine. This makes iterative feedback cycles more manageable within existing workloads. It enhances the Feedback and Assessment Management System (FAMS) by using natural language processing (NLP) to automate the categorization of feedback and create descriptive titles, improving organization and retrieval. When predefined feedback is unavailable, a lightweight language model trained on historical data generates relevant and clear responses, transforming the system into an adaptive feedback generator.

The architecture employs a teacher-in-the-loop approach where AI-generated feedback serves as a draft for instructors to review and refine. This ensures that AI enhances, rather than replaces, educators' professional judgment. By integrating AI into organizing and refining feedback, the DIFA system creates a continuous, data-driven feedback cycle that lightens instructors' workloads while delivering timely, personalized responses. Moreover, incorporating AI within a structured assessment framework helps prevent students' misuse of generative AI, positioning it as a tool for genuine learning.

Preliminary results show that this AI-enhanced architecture improves feedback delivery while maintaining the necessary depth and personalization for effective formative assessment. The DIFA system allows for iterative revisions rather than just final outputs, illustrating how explainable AI can enhance educational expertise. Future developments will include a student-facing interface that focuses on adaptive explanations and interactive feedback for neurodiverse learners.
References:
Shahzad, R. K., Ström, E. & Mozelius, P. (2023). FAMS: A Formative Assessment Management System for Generating Individualised Feedback. In 18th European Conference on Technology Enhanced Learning, 2023.

Keywords: formative assessment, iterative feedback, feedback automation, educational technology, higher education pedagogy

Generative AI for Dynamic Micro-Cases in business education

Ana Trevino
Tecnologico de Monterrey, México
actrevin@tec.mx

Engaging students in meaningful learning activities in the AI era presents both challenges and opportunities for higher education. Although case-based learning is widely used to promote authentic and applied learning, traditional cases are often static and limit student interaction. This paper presents a pilot digital learning experience (DLE) that leverages generative AI to transform static cases into interactive, dynamic micro-cases.

The DLE integrates AI tools enabled by Skill Studio and Agent Studio, part of the TECgpt ecosystem developed at Tecnológico de Monterrey. These tools allow instructors to design adaptable and interactive case-based learning activities aligned with specific instructional goals. Configurable AI skills enable instructors to generate cases from scratch and create multiple versions of a case by modifying organizational context, protagonists' roles, or personality traits. To enhance interaction, instructors can configure an AI agent that role-plays as case protagonists, allowing students to engage with the case through dialogue rather than passive reading.

A pilot implementation was conducted with 24 undergraduate students enrolled in the Conscious Business minor at Tecnológico de Monterrey. The Instructor created five micro-cases representing organizations from different countries and sectors. Each case was structured around an employee survey based on Raj Sisodia's Conscious Business Audit. Survey data were generated manually to ensure differentiation across cases while maintaining balance across the four audit dimensions: higher purpose, stakeholder orientation, conscious leadership, and conscious culture.

During the activity, the AI agent prompted students to select their assigned company and the protagonist with whom they wished to interact. The agent responded to students' questions based on the selected context, remaining faithful to the case data while expanding explanations when appropriate.

After completing the activity, students documented the information gathered and responded to a brief survey evaluating their learning experience. Results indicated high levels of engagement and perceived learning impact (mean scores of 9.40 and 9.19 on a 10-point scale). Lower scores and greater variability were observed in clarity, usefulness, and precision of the agent's responses, highlighting challenges related to the effective deployment of generative AI in educational settings.

Overall, this pilot illustrates how AI-driven dynamic micro-cases can enhance student engagement and active learning while also showing limitations related to agent behavior and pedagogical alignment. The study contributes an applied educational case to ongoing discussions about the meaningful integration of the uses of generative AI in business education.

Keywords: dynamic micro-cases, generative AI agents, AI-enabled instructional design, digital learning experience, business education

Cyber Resilience in the Age of AI: Making all Endusers Cybercrime Fighters

Sebastiaan (Basie) von Solms
University of Johannesburg, South Africa
basievs@uj.ac.za

A 2025 estimate is that cybercrime will cost the World $12.2 trillion annually by 2031.
The global cybersecurity awareness training market size was worth around USD 4.30 billion in 2023.
More and more reports indicate that AI will have a further negative impact on cyber resilience. AI is allowing threat actors to create more sophisticated forms of cyberattacks, creating a whole new dimension of cyber risks to all users. The WEF comments that 'AI is supercharging a global cyber fraud crisis'.

While more money is spent on cybersecurity awareness, specifically on cybersecurity awareness, cybercrime keeps increasing! This looks like an anomaly. What is wrong? We argue that one main reason for this anomaly is that most of the cybersecurity awareness training and education miss the target because the target is not well defined or even worse – they often do not know precisely what the target is! The main purpose and target of all cybersecurity awareness (resilience) training should be fighting cybercrime! In the age of AI this target becomes even more important!

Chatham House advises that it is time to re-assess the intersections between cybersecurity and cybercrime. Another view along the same lines advises the breaking down of silos between cybersecurity and cyber fraud detection forcing the convergence of these two concepts.

The main contribution of this presentation is to discuss a model attempting to merge these two silos and to explicitly and unambiguously highlight the real target of all cybersecurity awareness training and education – to fight cybercrime!

It is well-documented that cyber criminals progressively direct their attacks towards the end-user-workforce. In doing that they hope that the specific end-user will not recognize the attack, and compromise corporate sensitive information. AI will help to make such attacks even more difficult to detect. The model presented is based on the concept of creating a workforce of what we call 'enduser-fighters'. These are end-users who had been exposed through relevant training, to the concepts of traditional cybersecurity awareness training and to the impact and modus operandi of cybercrime attacks. The model is visualised as a missile with cybercrime as the target. The payload consists of a skilled group of enduser-fighters which are 'ordinary' end-users but who are trained and skilled to also act in a dual role – as normal end-user workers doing their assigned job and as cybercrime fighters. By default such end-user-fighters are both workers and fighters! The unclear notion of a culture of Cybersecurity Awareness now becomes a much clearer culture of Cybercrime-fighting. This dual role of end-users (workers) and fighters (soldiers) is also based on some ant colonies where every worker is also (by default) a soldier. Both cybersecurity awareness and cybercrime fighting are well-documented areas. The contribution of this model is to combine the two areas into a single model.

Keywords: Cybersecurity Awareness, Cybercrime, Enduser-fighters, Ants

Epistemic Trust and Epistemic Context and the Assessment of the Trustworthiness of AI platforms

Anthony Wensley[1] and Max Evans[2]
[1]University of Toronto, Canada
[2]McGill University, Montreal, Canada
anthony.wensley@utoronto.ca

Increasingly AI platforms are finding their ways into our lives in a multitude of ways. Individuals are consulting AI platforms such as ChatGPT for medical diagnosis, treatment advice and therapy. Businesses are increasingly incorporating agentic AI into their business to support complex, mission-critical processes. We pose the following questions: Is the trust that is increasingly being put into the uses of these AI platforms and Agentic programs appropriate? Are such systems trustworthy? How can we make them more trustworthy?

The established prior research into trustworthiness from the perspective of human-human interaction has identified three principal components that underpin the assessment of trustworthiness – ability, benevolence and integrity. How can we adapt/develop this research to address situations where humans are interacting with AI platforms such as LLMs or model-based systems?

In our presentation we will focus on human – AI interactions which relate to relatively high-risk interactions where the human concerned is seeking advice which, if not appropriately founded can lead to serious negative outcomes. Examples abound in the medical, engineering, defense fields, among others.

We divide our presentation into two parts. In the first part we investigate salient aspects of human-to-human interactions in these contexts. We examine how individuals come to assess the trustworthiness of the provider of suggestion/advice/therapies. In the second part we examine how existing theory concerning trustworthiness can be utilized to explore the assessment of trustworthiness in the context of human-AI interactions. We also add to prior analyses of trustworthiness by introducing and explicating the concepts of 'epistemic context' and 'epistemic trust'.

The concept of 'epistemic context' refers to the concatenation of the beliefs, assumptions, knowledge that represent the overall epistemic stance of individuals and groups. In parallel the term 'epistemic trust' refers to the trust that we place in an individual (or AI) which we are relying on to provide us with reliable knowledge. To the extent that human-human interactions potentially provide richer interaction we examine both the way epistemic trust is established in these contexts and how such trust can be established in a the different and potentially more constrained human-AI contexts.

Keywords: Epistemic trust, Epistemic context, AI platforms, Trustworthiness