ARTIFICIAL INTELLIGENCE

(AI)

in Education

editor

Omid Noroozi

ARTIFICIAL INTELLIGENCE

(AI)

in Education

editor

Omid Noroozi

Artificial Intelligence (AI) in Education

Editors

Omid Noroozi, Wageningen University and Research, Netherlands

ISBN: 978-1-952092-89-3

Date of Publication

December 2025

Publisher

ISTES
Monument, CO, USA

Contact

International Society for Technology, Education and Science (ISTES)
www.istes.org
istesoffice@gmail.com

Citation

Noroozi, O. (Ed.). (2025). *Artificial Intelligence (AI) in Education*. ISTES.

Table of Contents

Chapter 1 - From Grain to Flour: A Gristmill Metaphor for Ethical AI Design

Suzanne Porath ⓘ**, Carolyn Hall** ⓘ

Chapter Highlights

- ➢ The article portrays Artificial Intelligence (AI) as a powerful yet opaque technology that transforms data into impactful decisions while hiding its inner processes.

- ➢ It uses the metaphor of a gristmill, where data is the grain, algorithms are the grinding stones, and human oversight is the miller ensuring quality and safety.

- ➢ The study explores AI applications in agriculture, healthcare, business, and education to show how raw data is refined into meaningful social outcomes.

- ➢ It identifies major ethical challenges, including biased or incomplete data that distort results, algorithmic opacity that weakens accountability, and overreliance on AI outputs that limits human agency.

- ➢ To promote responsible AI, the article introduces six guiding principles: fairness, transparency, accountability, privacy, inclusivity, and sustainability.

- ➢ It emphasizes that automation is not neutral—like a mill, AI amplifies either honesty or corruption depending on the quality of its inputs and oversight.

- ➢ Finally, the article offers an accessible interdisciplinary framework that clarifies ethical responsibility and supports human-centered oversight for equitable AI outcomes.

Introduction

Artificial Intelligence (AI) increasingly shapes decisions in healthcare, finance, education, and other domains. Its capacity to process vast amounts of data generates outputs that influence human life in profound ways, yet its mechanisms often remain opaque and ethically contested. The central concern is not only technical performance, but also the social and ethical conditions under which AI systems are designed, governed, and interpreted.

The metaphor of the gristmill offers a framework for examining these conditions. A gristmill transforms raw grain into flour through calibrated stones, guided by the vigilance of the miller. Similarly, AI systems process raw data through algorithms to produce outputs that must be critically evaluated before they are applied. By tracing this analogy across inputs, mechanisms, outputs, and oversight, the analysis highlights how ethical risks and responsibilities arise at every stage of AI. The gristmill metaphor clarifies how responsible design depends on the integrity of data, the transparency of algorithms, and the discernment of human operators, situating AI within broader debates about fairness, accountability, and human-centered technological development.

This article contributes to ongoing debates in *AI & Society* by offering a metaphorical framework that makes the ethical dimensions of AI systems more visible and comprehensible. While discussions of responsible AI often focus on technical solutions or regulatory frameworks, the gristmill metaphor provides an integrative lens that links data quality, algorithmic design, system outputs, and human oversight into a single cycle of responsibility. By applying this framework across diverse domains - including agriculture, healthcare, business, and education -the analysis demonstrates how ethical risks recur regardless of sector, and how principles of fairness, transparency, accountability, privacy, inclusivity, and sustainability can guide human-centered design. The metaphor advances conceptual clarity, supports interdisciplinary dialogue, and offers a pedagogical tool for explaining why automation is never neutral but always contingent upon the values embedded in its inputs, mechanisms, and oversight.

The Mill in Motion: AI's Transformative Power Across Fields

Like a traditional gristmill, AI takes in unrefined inputs and turns them into a product humans

can immediately use, thereby amplifying their capacity to learn, create, and decide. Just as raw grain becomes data, grinding stones become algorithms, and the resulting flour becomes the decisions, insights, and predictions that shape subsequent action. This transformative function is not limited to a single field, AI is being employed across industries, from agriculture to healthcare to education, where its capacity to process and refine raw data is reshaping daily operations, decision-making, and human potential.

In the food industry, AI systems function like precision-engineered mills; refining raw, unstructured data into insights that optimize production, reduce waste, and support sustainable practices. For example, Agrawal et al.'s, (2025) literature review showed how AI-driven predictive analytics in the food production process enabled precise demand forecasting, thereby minimizing overproduction and associated waste. Real-time monitoring systems, powered by AI, facilitate immediate quality assessments, ensuring product consistency and safety. Computer vision and sensor technologies were utilized for precision assessment, automating inspection processes across production stages. Moreover, AI supported the implementation of circular economy practices by enabling the repurposing of food by-products and improving demand forecasting to prevent overproduction. These applications underscore AI's capacity to transform raw data into valuable outputs, much like a gristmill refines grain into flour. These examples from food production illustrate how AI creates high-quality, usable outputs from raw materials. The healthcare sector mirrors this same cycle; data enters, algorithms process, and life-saving decisions follow.

AI in healthcare demonstrates this transformative cycle. In their general literature review, Al Kuwaiti et al. (2023) highlight AI's role in translating massive health data inputs into actionable clinical decisions, particularly through diagnostic imaging and real-time monitoring. For example, AI-enabled platforms like Ultromics analyze echocardiograms to detect heart disease patterns previously invisible to human eyes, guiding early treatment decisions. AI models have also demonstrated success in predicting diabetes onset and identifying COVID-19 infections from CT scans using deep neural networks. These systems not only detect but classify and forecast outcomes, enabling physicians to intervene earlier and more precisely. While healthcare prioritizes life and diagnosis, the business sector adapts the same grind for profit and performance.

AI in business intelligence mirrors this transformation. According to the literature review by

Eboigbe et al. (2023), the integration of AI "redefined business operations, offering unprecedented insights and fostering more informed decision-making processes" (p. 286). Machine learning and predictive analytics uncover patterns not visible through traditional methods, transforming raw inputs into actionable outcomes. These outcomes enable subsequent actions such as rerouting supply chains to enhance sustainability, detecting and halting fraudulent transactions in banking, or personalizing marketing strategies in real time. Self-service Business Intelligence (BI) tools now empower frontline staff—not just executives—to act on AI-generated insights without technical gatekeepers. In this way, AI doesn't just produce flour—it fuels continuous, strategic, and context-responsive action across entire enterprises. In a similar fashion, education systems now harness AI to interpret student learning behaviors and personalize instruction, turning raw interaction data into pedagogical precision.

AI in education operates as a transformation engine from input to impact. In their systematic literature review, Wang et al. (2024) showed how AI tools like intelligent tutoring systems (ITS) and adaptive hypermedia platforms can diagnose student learning styles, tailor content, and recommend customized learning paths. These systems, using algorithms such as Bayesian networks and neural networks, "analyze and identify students' learning styles within an e-learning system," leading to more effective and targeted instruction (p. 11). For example, ITS track student performance to offer adaptive feedback, tests, or reading recommendations - decisions made in real-time from raw behavioral data. These AI-driven outputs shape classroom actions, lesson adjustments, and instructional strategies, demonstrating how educational AI doesn't simply grind data; it can nourish learning through refined, responsive decisions.

Across these domains - food systems, healthcare, business, and education - AI operates as a gristmill: taking in raw, often chaotic data and producing usable, often transformative output. The patterns are clear: meaningful action depends on the refinement of input, the calibration of algorithmic mechanisms, and the context-sensitive application of outcomes. But just as a gristmill's flour is only as good as the grain and stones that produce it, AI's value hinges on the quality of its data, the design of its algorithms, and the ethical discernment of its users. This is a critical issue: the purity of the grain—the data—at the heart of AI systems.

Sorting the Grain: The Ethics and Consequences of AI's Data Inputs

If the gristmill relies on clean, dry grain to produce nourishing flour, AI depends on high-quality, representative, and ethically sourced data to yield accurate and equitable outcomes. When flawed, biased, or incomplete information is fed into AI systems, even the most advanced algorithms can reinforce harm, perpetuate inequality, or mislead decision-making. Concerns about data provenance, representational fairness, and annotation bias have emerged as central challenges across domains. Attention to the integrity of data inputs, what gets selected, how it's structured, and whose experiences are represented, shapes the reliability and responsibility of every AI-driven insight that follows.

Just as the quality of grain determines whether the flour will nourish or sicken, the integrity of data shapes the safety, usefulness, and fairness of AI outcomes. If flawed, incomplete, or biased data are introduced into a system, even the most advanced algorithms can produce harmful or misleading results. This idea, often referred to as "garbage in, garbage out," means that the quality of the input directly determines the quality of the output. In AI systems, poor input data can yield predictions or decisions that are inaccurate at best and discriminatory at worst.

Biased input data can embed systemic discrimination into algorithmic systems. This is particularly problematic in high-stakes fields like hiring and criminal justice, where decisions are often shaped by patterns drawn from historically inequitable records (Barocas et al., 2019). Facial recognition technologies have shown significantly higher error rates for people with darker skin tones, largely because training datasets were overwhelmingly composed of lighter-skinned faces (Krishnapriya et al., 2020). Similarly, predictive tools in policing and finance often perpetuate historical patterns of inequality embedded in the data (Brayne, 2017; Johnson & Rostain, 2020).

In a series of case studies, O'Neil (2016) demonstrated how opaque, data-driven models, when trained on flawed or exclusionary information, can reinforce inequality and disproportionately impact marginalized communities. Mehrabi et al. (2021) add that even subtle imbalances in data representation or annotation can produce biased predictions, misinform decisions, and erode public trust. If AI functions like a gristmill, then the grain it receives must be clean, representative, and thoughtfully chosen—otherwise, the resulting

flour only reinforces existing harm.

The act of selecting and labeling data is not neutral. It encodes the perspectives, priorities, and blind spots of those who control the process. Just as millers choose which grain is worth grinding, data curators decide what is included, whose experiences are captured, and how those experiences are interpreted. In many medical datasets, for example, patients with darker skin tones are underrepresented, leading to diagnostic tools that perform less accurately across racial groups (Frasier et al., 2025; Spinelli et al., 2024). In education, datasets used to train learning analytics platforms often favor Western, English-speaking student populations, overlooking linguistic and cultural diversity (Cheuk, 2021; Li et al., 2025; Wang et al., 2024). As Gebru et al. (2021) argue, the creation of "datasheets for datasets" (pg. 86) is one way to make these choices more visible, forcing researchers to account for where their data comes from and who is missing.

In Natural Language Processing (NLP), a branch of AI focused on enabling machines to interpret and generate human language, the task of sorting and preparing data is not only technically intricate but also ethically fraught (Bender & Friedman, 2018). Friedman (2018) propose the use of *data statements* as a standardized, transparent method for documenting datasets in NLP. These data statements are intended to mitigate bias and improve the scientific and ethical rigor of NLP systems by requiring developers to disclose detailed information about dataset composition, speaker and annotator demographics, curation rationale, and intended use. By making the characteristics of training and testing data explicit, data statements aim to reduce the risk of emergent and pre-existing bias while promoting more inclusive, accountable, and accurate language technologies. Before any data is processed, the ethical burden rests with those who gather and prepare the grain for grinding.

Grinding the Grain: Algorithms and the Ethics of the Black Box

If data is the grain, then algorithms are the grinding stones of AI; the mechanisms that process, refine, and shape input into output. However, unlike physical millstones, these algorithms often operate as black boxes, producing decisions without transparent reasoning. This opacity is particularly concerning in high-stakes domains such as healthcare, finance, and criminal justice, where understanding the rationale behind AI decisions is crucial. The "black box" problem arises because many machine learning models, especially deep neural

networks, are inherently complex and lack interpretability, making it difficult to trace how specific inputs lead to particular outputs (Adadi & Berrada, 2018).

But unlike the visible stones of a traditional gristmill, many modern algorithms operate as black boxes, concealing how inputs are transformed into decisions. This opacity has prompted urgent calls for explainable artificial intelligence (XAI). Tim Miller (2019) argues that making AI more interpretable requires more than just technical transparency; it demands that algorithms offer explanations humans can actually understand. Drawing on insights from philosophy, psychology, and the social sciences, Miller shows that people seek contrastive explanations ("Why P rather than Q?") (pg. 9), prefer causal over statistical reasoning, and interpret decisions through social context. These human-centered expectations often clash with the way algorithmic models currently explain themselves, if they do at all. To restore trust and accountability, algorithm designers must not only improve the grind but open the mill, allowing observers to see how the stones are set and why they move the way they do.

To address this issue, the field of Explainable Artificial Intelligence (XAI) has emerged, focusing on developing methods to make AI decision-making processes more transparent. Tools like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) have been adopted to interpret complex models by attributing predictions to input features, thereby providing insights into the model's behavior (Wang, 2024). Tjoa and Guan (2020) explore the growing field of explainable artificial intelligence (XAI), with a particular focus on healthcare. They review a wide range of methods designed to help people understand how AI models, especially complex ones like deep neural networks, make decisions. These include visual tools like saliency maps, simplified mathematical models, and natural language explanations. The authors emphasize that in medicine, where lives are at stake, it's not enough for AI to be accurate; it must also be interpretable and trustworthy. They call for greater attention to ethical issues, human-centered design, and clear communication between AI systems and medical professionals.

By implementing such techniques, AI systems can become more transparent and trustworthy, allowing stakeholders to scrutinize and understand the decision-making process. This transparency is akin to installing viewing windows in a mill, enabling observers to witness the grinding process and ensure its integrity. Ultimately, enhancing the explainability of AI algorithms is vital for ethical integrity and public trust in AI-driven decisions.

Flour on the Table: The Ethics of Acting on AI's Output

If data is the grain and algorithms are the grinding stones, then the final product, the flour, is the usable knowledge that AI produces - predictions, classifications, or recommendations. These outputs fuel everything from individualized lesson plans to cancer diagnoses and fraud detection. But this flour is only as useful and ethical as the way it is used. A growing concern across disciplines is not simply whether AI is accurate, but whether humans are over-relying on it without sufficient oversight.

The final output of AI, its flour, can empower human decisions, but when left unchecked, it can also diminish human autonomy and blur the line between assistance and control. Taddeo and Floridi (2018) argue that while AI offers powerful tools for efficiency and innovation, its improper use, particularly when humans unquestioningly defer to its authority, can displace human agency in subtle and profound ways. As AI becomes increasingly embedded in daily life, it "blends into our lives, experiences, and environments and becomes an invisible facilitator that mediates our interactions in a convenient, barely noticeable way (pp. 751-752). This "influencing power," when left unchecked, threatens the "fragile, and yet constitutive, ability to determine our own lives and identities" (p. 752). Delegating decisions to AI without mechanisms for oversight or accountability risks eroding self-determination, especially when systems operate imperceptibly and without meaningful human input. To prevent this, the authors call for robust ethical frameworks, human oversight, and foresight methodologies to ensure AI supports, not supplants, human flourishing.

In high-stakes fields like medicine, the flour that AI systems produce, diagnostic predictions, treatment recommendations, or triage decisions, can have profound consequences for human life. When these outputs are accepted without question, the risks multiply. As Guan et al. (2022) warn, decisions grounded in incomplete or biased data can "result in ethical risks such as privacy breaches, risk to human life, and undermining social justice" (p. 3). These consequences are especially acute in medical contexts, where algorithms lack the tacit knowledge, emotional nuance, and ethical discernment that experienced clinicians bring to complex, ambiguous cases. Without human review, even a well-trained system may miss the context, misinterpret the data, or perpetuate systemic bias. The mill may be efficient, but if no one inspects the flour before it's served, the results may be deeply harmful.

Even the finest flour can become harmful if exposed to the wrong conditions: moisture, contamination, or improper handling. In the same way, the flour of AI - its predictions, classifications, or recommendations - must be critically examined before it is applied, as even well-designed systems can produce outputs that are misleading or ethically problematic when used without review. This is the purpose of Human-in-the-Loop (HITL) design, an approach that embeds human judgment directly into the development and operation of AI systems. Rather than allowing models to function autonomously, HITL ensures that humans are involved at key stages; labeling training data, validating model accuracy, and overseeing live decisions. This oversight is especially critical in high-stakes domains like healthcare and finance, where ambiguous, novel, or ethically sensitive cases demand human discernment. HITL systems are not only more accurate; they foster accountability and trust. As noted in a recent report, "Human-in-the-loop systems don't just reduce the risk of errors; they help build a culture of accountability" (RadarFirst, 2025, para. 17). Just as a miller must inspect and manage the final product to ensure it is safe for consumption, AI systems require deliberate human oversight to ensure that their outputs are not only technically accurate, but also ethically sound and socially responsible.

Empirical evidence underscores the need for caution when relying on AI outputs. Guan et al. (2022) note that AI systems, while efficient, often lack the depth of human reasoning required to navigate complex or ethically sensitive scenarios. They emphasize that when AI is used without critical human oversight, it may fail to incorporate tacit knowledge - such as cultural norms, emotional intelligence, or context-specific understanding - that is essential to sound judgment. These limitations can lead to decisions that are technically precise but socially or ethically misaligned. The reliability of AI-generated outcomes depends not only on the sophistication of the algorithm, but also on thoughtful integration with human review. Without mechanisms for interpretation, challenge, or revision, the outputs of even the most advanced systems risk becoming misleading or harmful when applied uncritically.

Ensuring that AI systems produce outcomes that are not only efficient but ethically sound requires intentional design. This is where Human-in-the-Loop (HITL) approaches become indispensable. Ensuring that AI outcomes are ethically sound requires design approaches that do more than optimize for accuracy or efficiency—they must account for moral complexity. The Human-in-the-Loop (HITL) system, as described by Chen et al. (2023), is not only a technical tool but a philosophical and ethical safeguard. Their framework integrates human

annotation and iterative feedback throughout the lifecycle of AI development, enabling systems to respond to complex, value-laden situations. They emphasize that HITL "takes into full consideration the perspectives of human annotators, incorporating their moral viewpoints into the training data" (p. 5) allowing models to be "more contextually ethical" through repeated loops of refinement. Rather than displacing human judgment, HITL supports and extends it, ensuring that AI's flour, so to speak, is milled not just with precision, but with principled care.

Milling with Care: Toward Ethical, Human-Centered AI

Across sectors, from agriculture to education, healthcare to business, AI functions like a modern-day gristmill, processing raw data into usable insights, predictions, and decisions. But the quality of this output, the flour, depends on much more than computational power. As this article has shown, ethical and effective AI requires clean, representative data (the grain), transparent and accountable algorithms (the grinding stones), and engaged human oversight (the miller). When these components are handled with care, AI systems can serve human flourishing. When neglected, they risk reinforcing harm.

To run the AI gristmills responsibly, we must follow core principles that align automation with ethical intention: *fairness*, to prevent bias and ensure equitable outcomes; *transparency*, to make systems understandable and traceable; *accountability*, to ensure humans, not machines, remain answerable for decisions; *privacy*, to protect individuals from surveillance or misuse; *inclusivity*, to represent diverse perspectives and populations; and *sustainability*, to consider the environmental and social impacts of AI development and deployment.

These principles frame AI not as an autonomous actor, but as a tool within a broader human system. The data it ingests reflects society's values or its blind spots; the algorithms it applies can either amplify justice or replicate inequality; and the outputs it delivers must be interpreted, questioned, and applied with human care. Ethical AI, then, is not just about engineering; it is about intentional stewardship.

Ultimately, the gristmill metaphor is a reminder that automation is never neutral. Left on its own, the system will grind whatever it's given, regardless of consequences. But with ethical principles guiding the selection of the grain, maintenance of the stones, and inspection of the

flour, AI becomes a powerful tool for good - refined not just for efficiency, but for equity, dignity, and the shared good of all.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138-52160. http://dx.doi.org/10.1109/ACCESS.2018.2870052

Agrawal, K., Goktas, P., Holtkemper, M., Beecks, C., & Kumar, N. (2025). AI-driven transformation in food manufacturing: a pathway to sustainable efficiency and quality assurance. *Frontiers in Nutrition*, *12*, 1553942. http://dx.doi.org/10.3389/fnut.2025.1553942

Al Kuwaiti, A., Nazer, K., Al-Reedy, A., Al-Shehri, S., Al-Muhanna, A., Subbarayalu, A. V., Al Muhanna, D., & Al-Muhanna, F. A. (2023). *A review of the role of artificial intelligence in healthcare*. *Journal of Personalized Medicine*, *13*(6), Article 951. https://doi.org/10.3390/jpm13060951

Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, *6*, 587-604. http://dx.doi.org/10.1162/tacl_a_00041

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, *82*(5), 977-1008. http://dx.doi.org/10.1177/0003122417725865

Chen, X., Wang, X., & Qu, Y. (2023). Constructing ethical AI based on the "Human-in-the-Loop" system. *Systems, 11*(11), 548. https://doi.org/10.3390/systems11110548

Cheuk, T. (2021). Can AI be racist? Color-evasiveness in the application of machine learning to science assessments. *Science Education*, *105*(5), 825-836. http://dx.doi.org/10.1002/sce.21671

Eboigbe, E. O., Farayola, O. A., Olatoye, F. O., Nnabugwu, O. C., & Daraojimba, C. (2023). Business intelligence transformation through AI and data analytics. *Engineering Science & Technology Journal*, *4*(5), 285-307. https://doi.org/10.51594/estj.v4i5.616

Frasier, K., Hash, M. G., Werpachowski, N., & Fritts, H. (2025). The blind spots of artificial intelligence in skin cancer diagnosis. *Dermis*, *5*(2),35. http://dx.doi.org/10.35702/Derm.10035

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., &

Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. https://doi.org/10.1145/3458723

Guan, H., Dong, L., & Zhao, A. (2022). Ethical risk factors and mechanisms in artificial intelligence decision making. *Behavioral Sciences*, *12*(9), 343. http://dx.doi.org/10.3390/bs12090343

Johnson, R. A., & Rostain, T. (2020). Tool for surveillance or spotlight on inequality? Big data and the law. *Annual Review of Law and Social Science*, *16*(1), 453-472. http://dx.doi.org/10.1146/annurev-lawsocsci-061020-050543

Krishnapriya, K. S., Albiero, V., Vangara, K., King, M. C., & Bowyer, K. W. (2020). Issues related to face recognition accuracy varying based on race and skin tone. *IEEE Transactions on Technology and Society*, *1*(1), 8-20. http://dx.doi.org/10.1109/TTS.2020.2974996

Li, M., Enkhtur, A., Yamamoto, B. A., Cheng, F., & Chen, L. (2025). Potential societal biases of ChatGPT in higher education: A scoping review. *Open Praxis*, *17*(1), 79-94. http://dx.doi.org/10.55982/openpraxis.17.1.750

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35. http://dx.doi.org/10.1145/3457607

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, *267*, 1-38. http://dx.doi.org/10.1016/j.artint.2018.07.007

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown.

RadarFirst. (2025). *Why a Human-in-the-Loop is Essential for AI-Driven Privacy Compliance*. https://www.radarfirst.com/blog/why-a-human-in-the-loop-is-essential-for-ai-driven-privacy-compliance/

Spinelli, G., Khatun, N., & Colecchia, F. (2024). Technology innovation to reduce health inequality in skin diagnosis and to improve patient outcomes for people of colour: A scoping review and future research agenda. *Frontiers in Artificial Intelligence* 7:1394386. https://doi.org/10.3389/frai.2024.1394386

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, *361*(6404), 751-752. http://dx.doi.org/10.1126/science.aat5991

Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems*, *32*(11),

4793-4813. http://dx.doi.org/10.1109/TNNLS.2020.3027314

Wang, S., F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, *252*, 124167. http://dx.doi.org/10.1016/j.eswa.2024.124167

Wang, Y. (2024). A comparative analysis of model agnostic techniques for explainable artificial intelligence. *Research Reports on Computer Science*, 25-33. https://doi.org/10.37256/rrcs.3220244750

Author Information

Suzanne Porath

https://orcid.org/0000-0003-3158-2958

Kansas State University

USA

Contact e-mail: *sporath@ksu.edu*

Carolyn Hall

https://orcid.org/0009-0004-1138-1555

Kansas State University

USA

Citation

Chapter 2 - AI Literacy of Primary Education Pre-service Teachers: Self-assessment and Performance-based Measurements

Anthi Karatrantou , **George Panetas** , **Christos Panagiotakopoulos**

Chapter Highlights

> ➢ The paper discusses the results of performance-based scale assessing AI Literacy of primary school pre-service teachers / University Students attending the 1st year of studies.

> ➢ The paper discusses the results of self-assessment based scales to assess AI Literacy of primary school pre-service teachers / University Students attending the 1st year of studies.

> ➢ The paper discusses aspects such as Competence in using AI, Knowledge of generative AI technologies, Concerns about AI technologies, Benefits of using AI technologies, Challenges in using AI technologies, Willingness to use AI technologies.

> ➢ The paper discusses the relation between performance-based and self-assessment based assessment of AI literacy of primary school pre-service teachers / University Students attending the 1st year of studies.

> ➢ The paper provides a range of references on AI literacy scales.

Introduction

In recent years, artificial intelligence (AI) has been rapidly penetrating everyday life, but a lack of knowledge about what it is and how it works is widespread across all ages and professions. Due to a lack of knowledge or superficial understanding of AI, misconceptions about it and how it affects our lives can lead to false beliefs or to excessive enthusiasm, resulting in the ignoring of ethical issues such as objectivity and responsibility. Therefore, cultivating artificial intelligence literacy is necessary as it refers to the ability of individuals to use and critically evaluate AI tools and to collaborate effectively with them in multiple contexts (Long & Magerko, 2020). Familiarizing individuals who do not have technical background and educating them on fundamental topics of AI has become an urgent need that requires immediate attention (Long & Magerko, 2020; Ding et al., 2023).

Research has emerged worldwide on how to promote AI literacy among individuals without a technical background to meet this need. Opportunities have been researched and proposed on how to teach AI literacy in early childhood (Su et al., 2023), in K-12 education (Touretzky, 2019; Yue et al., 2022; Antonenko& Abramowitz, 2023; Du, 2024; Karatrantou, 2025c), in higher education and adult education (Laupichler et al., 2022; Biagini, 2024) in both formal and informal education settings (Long et al., 2022), as well as in workforce education (Cetindamar et al., 2022). According to the report of (Miao & Shiohira, 2022), eleven countries have already designed and implemented AI curriculum for K-12 education. The goals of AI education worldwide focus on educating the younger generation to be able to adapt to AI-infused workplaces and inspire them to become not just users but also designers, developers and researchers (Kandlhofer, 2016; Pedro, et al., 2019).

However, the assessment of AI literacy relies strongly on subjective measures, such as qualitative assessment and self-reported surveys, which can be subjective and biased (Rosenman et al., 2011; Laupichler et al., 2022; Maitz et al., 2022; Mertala et al., 2022; Nader et al., 2022; Ding Antonenko & Abramowitz, 2023; Casal-Otero et al., 2023). Recently, many studies have explored how to measure knowledge of AI (Ng, Luo, Chan, & Chu, 2022; Laupichler, Aster, & Raupach, 2023; Wang & Chuang, 2023; Chiu, Ahmand, Ismail, & Sanusi, 2024), most of which suggest using self-reported questionnaires from citizens. These studies actually assessed respondents' perceptions of their AI literacy rather than their actual knowledge of AI, as self-reported answers rarely correspond accurately to

actual measurements.

On the other hand, objective measurements for accurately assessing individuals' AI literacy, like the way assessments evaluate their natural sciences literacy, mathematics literacy, digital literacy, and computational literacy (as demonstrated by the PISA, TIMISS, and ICILS programs for students at all levels of education) are needed (Chiou et al., 2024). Recently, studies proposed objective measures of AI literacy, with the aim of designing learning objectives and educational programs that cultivate responsible citizens of the new AI era, but most of them were designed for higher education students or attendees of specific AI literacy courses (Chiou, et al., 2022; Hornberger, Bewersdorff, & Nerdel, (2023), Weber et al., (2023), Knoth et al., 2024)

The chapter discusses the AI literacy and their perceived AI literacy of 1st year university students studying to be primary school teachers (pre-service primary school teachers) utilizing self-assessment and performance-based measurements. Both scales are used, self-assessment and performance-based, aiming to assess participants' knowledge and skills concerning fundamentals of AI technologies and applications as well as to write down their perceived benefits, challenges and concerns about AI and their willingness to use AI technologies.

AI Literacy and AI Literacy Measures

As the opportunities and challenges arising from advancements in artificial intelligence and its pervasiveness in society are becoming more apparent, understanding artificial intelligence is becoming a crucial issue. The use of high-quality, valid, and reliable AI literacy tools is vital for understanding and promoting AI literacy growth. As the definition of AI literacy is multidimensional (knowledge, skills, attitudes, social/ethical aspects), reliable tools for education and research are needed. Nowadays, there are a variety of scales for measuring AI literacy with different dimensions targeting general population, higher education students, secondary education students, and teachers. Some of them focus on technical knowledge while others focus on attitudes/ethics, while differences in psychometric documentation are identified, resulting in some being reliable and others being less reliable (Davenport & Ronanki, 2018; Cave, 2019; Lintner, 2024). Most of the tools are self-reported questionnaires/scales (self-assessment based scales) actually assessing self-perceived AI

literacy and less are knowledge-based questionnaires/tests (performance-based scales) aiming to achieve objective measures of AI literacy.

AI Literacy

Recent researchers have proposed the term "artificial intelligence (AI) literacy" to describe individuals' understanding and perception of artificial intelligence, without focusing on its technical aspects. AI literacy can be broadly defined as the ability to understand and effectively and responsibly use AI systems. Beyond simply becoming familiar with AI, AI literacy enables individuals to critically evaluate information related to artificial intelligence and make informed decisions (Ding, 2024). AI literacy involves a multifaceted understanding of the basic principles, applications, and ethical implications of AI (Ng et al., 2021a; Ng et al., 2021b). While AI literacy initially only required of individuals with a technical background, with the increasing popularity and development of AI technology, this skill has expanded to individuals without technical knowledge, particularly in the last two years (Long & Magerko, 2020). Individuals without a technical background who are knowledgeable about artificial intelligence do not need to be programmers or engineers. Instead, they need to have the knowledge necessary to make informed decisions about products and services that have been enhanced with AI, be aware of the ethical aspects of AI technologies, understand news and discussions related to AI, and interact effectively with AI systems.

According to Long & Magerko (2020) literacy in the field of artificial intelligence is defined as 'the ability to critically evaluate AI technologies, communicate and collaborate effectively with AI, and use AI as a tool online, at home, and in the workplace' (Long & Magerko, 2020). This literacy is closely related to other forms of literacy. Digital literacy refers to an individual's ability to find, evaluate, create and communicate information using digital tools and technologies (Martin & Grudziecki, 2006). Digital literacy is a prerequisite for artificial intelligence literacy because people need to understand AI through digital tools and applications. Data literacy refers to the ability to use an inquiry-based approach to read, understand, analyze, create and communicate data as information (Ow-Yeong et al., 2023; Wolff et al., 2016). This literacy overlaps with AI literacy because the main idea of AI is "learning from data", which is also a subfield of machine learning. Mathematical literacy refers to an individual's ability to formulate, use and interpret mathematics in various contexts (Jackson et al., 2021), as defined by PISA 2012. It is the foundation of AI and

promotes advanced AI knowledge, such as algorithms for machine learning. Computational literacy is defined as the ability to use and apply coding and other computational tools to develop and explain solutions to problems (Jacob & Warschauer, 2018). It would cultivate individual' skills in developing AI applications. However, as not all individuals should be AI programmers or researchers, it is not required for AI literacy.

Overall, this definition of AI literacy is directly linked to its assessment. In the study by Chiu and colleagues (2022), AI engineering professors proposed three main areas of AI literacy: *knowledge of artificial intelligence, artificial intelligence in process*, and *the impact of artificial intelligence*. In their multidisciplinary analysis of the literature, Long and Magerko (2020) presented a framework for assessing AI literacy with five main axes concerning: *What is artificial intelligence? What can artificial intelligence do? How does artificial intelligence work? How should artificial intelligence be used? How do people perceive artificial intelligence?*

Self-assessment based Scales Measuring AI Literacy

Developing AI literacy scales started recently with the oldest being MAIRS-MS(Karaca, Çalışkan & Demir, 2021) since 2021. MAIRS-MS, SNAIL (Laupichler, et al., 2023), and AILS (Wang, Rau & Yuan, 2022) are the most reused scales researchers. Most of the scales target general population or higher education students and just a few target secondary education students and teachers (Lintner, 2024; Biagini, Cuomo & Ranieri, 2024).

Researchers while creating AI literature scales drew their conceptualizations of AI literacy from different sources and target different populations but they include common competencies comprising AI literacy. Almost all of them *emphasize the technical understanding of AI*, *consider the societal impact of AI as a critical component* and *acknowledge AI ethics as an essential aspect* (Lintner, 2024; Knoth, et al., 2024).

However, the authors of the scales differ in perceiving higher order AI-related skills such as *creation* and *evaluation* of AI, as components of AI literacy. In the original Ng et al.'s conceptualization (Ng, et al., 2021), creation and evaluation of AI are core components of AI literacy. MAILS (Carolus, et al., 2023) based on the Ng et al.'s conceptualization identified creation of AI as a related, but separate construct from AI literacy. AILQ (Ng, et al., 2023),

based on the same conceptualization includes creating AI as a core part of AI literacy. Several other scales consider the ability to critically evaluate AI as a core part of AI literacy (Wang, Rau & Yuan, 2022; Hornberger, Bewersdorff & Nerdel, 2023; Pinski & Benlian, 2023; Zhang, Perry & Lee, 2024; Soto-Sanfiel, Angulo-Brunet. & Lutz, 2024; Lee & Park, 2024). Considering the widespread integration of AI into daily and professional life, a question arises whether the skills to create AI will not have to be included as core competencies of AI literacy in near future, as those competencies might be crucial for functional AI literacy.

Lintner (2024) carried out a systematic review of AI literacy scales describing in detail, evaluating and discussing thirteen (13) self-assessment-based scales consisting of Likert items. He assessed the quality of the scales based on the COSMIN measurement properties (Prinsen, et al., 2018; Mokkink, et al., 2020) as well as on interpretability and feasibility. Most scales demonstrate satisfactory structural validity and internal consistency but, only a few have been tested for content validity, reliability, construct validity, and responsiveness. None of the scales have been tested for cross-cultural validity and none of the scales fully cover all dimensions of AI literacy. None of the scales showed positive evidence for all COSMIN measurement properties and most studies characterized by weak methodological rigour. Additionally, the scales' interpretability and feasibility are unknown as most studies do not report the necessary indicators and do not provide open data (Lintner, 2024).

Research derives that tools are needed that integrate technical, social, ethical, and pedagogical aspects under a unified and reliable assessment framework. Interdisciplinary collaboration is needed to develop more valid, reliable and objective tools (Lintner, 2024; Knoth, et al., 2024).

For the purposes of this study the theoretical frameworks and items of the scales of Ding, 2024, Chiu 2024 and Yau, et al., 2022 were used.

AI Literacy Assessment for Non-technical Individuals by Ding, et al. (2024)

Ding, et al. (2024) developed an AI literacy assessment based on the framework developed by Long and Magerko (2020) consisted of 17 AI competencies organized into five key facets of AI literacy: *understanding AI's nature, recognizing AI's capabilities, grasping AI's*

underlying mechanisms, discerning appropriate AI utilization, and comprehending public perceptions of AI. Pre- and in-service teachers participated at the validation procedure of the Ding, et al. assessment tool. The final version of the tool consisted of 25 items, each of which is designed to assess one or two competencies.

AI Literacy Test by Chiu, et al. (2024)

The study aimed to develop and validate an AI literacy test for school students within the project 'AI4future'. Engineering and education researchers created and selected 25 multiple-choice questions to create the test. School teachers validated the 25 multiple-choice questions while developing an AI curriculum for middle schools. More than 200 students in grades 7 to 9 took the test and offered valuable data. The test aims to enable education researchers and teachers to appropriately evaluate their AI-related interventions.

AI Literacy Test for Junior Secondary Students by Yau, et al. (2022)

The study focuses on the creation and validation of an AI literacy test based on the *Five Big Ideas* (*Perception, Representation and reasoning, Learning, Natural Interaction, Societal impact*) for grade 7 to 9 students. It is connected to the 'AI for the future' project by the University of Hong Kong. The researchers based on the content knowledge in Awareness, Ethics and Impact and Knowledge teaching units addressed for students to create and validate 10 questions.

Performance-based Scales Measuring AI Literacy

As already mentioned above, existing measurements of AI literacy are mainly based on self-report questionnaires, which often overestimate or underestimate actual knowledge and skills of individuals. Objective tools are needed to measure what individuals actually know and can do. Such tools could be knowledge tests with multiple-choice questions on fundamental AI concepts (e.g., machine learning, neural networks, AI applications). Research on objective measures of AI literacy is needed (Chiu, et al., 2024).

Most relative studies evaluate AI learning activities and curriculum designs, as well as learning outcomes during courses and curricula (Ng et al., 2022; Olari, 2023; Kong, Man-Yin

Cheung & Zhang, 2023; Laupichler et al., 2023; Sanusi et al., 2023; Weber, Pinski & Baum, 2023; Ding, Kim & Allday, 2024). There are very few relevant and appropriate objective assessment tools concerning individuals' AI literacy. That may be since measures of perceived ability to use AI and AI literacy may have been confused in previous studies (Laupichler et al., 2023; Ng et al., 2022; Karatrantou, 2025a). Furthermore, the definition of AI literacy may not fit the needs of all individuals (students, university students, adults, workforce citizens). Additionally, as literacy relates to knowledge and skills, its assessment should integrate practical skills assessment and hands-on activities (Lee, et al., 2021; Chiu, et al., 2024).

Lintner (2024) during his systematic review of AI literacy scales found and described in detail, evaluate and discussed only three (3) performance-based scales, the AI literacy test (Hornberger, Bewersdorff & Nerdel, 2023), the AI-CI test (Zhang, Perry & Lee, 2024) and the SAIL4ALL scale (Soto-Sanfiel, Angulo-Brunet & Lutz, 2024). AI literacy test and SAIL4ALL scale are based on Long & Magerko's (Long & Magerko, 2020) conceptualization of AI literacy. The three tools provide evidence for good structural validity and internal consistency while AI-CI provide evidence for satisfactory content validity and AI literacy test for satisfactory construct validity as well.

Future studies should suggest scales including items for practical tests. More investigation is needed to determine what individuals should learn and what knowledge and skills they should acquire, and future studies should include more skills and knowledge for assessment (Yau, et al., 2022; Karatrantou, 2025b).

For the purposes of this study the theoretical frameworks and items of the scales of Bingcheng Wang, et al., 2023 and Chan & Hu, 2023 were used.

User Competence in Using Artificial Intelligence Scale by Bingcheng Wang, et al. (2023)

The researchers identified the core constructs of AI literacy, such as Awareness, Usage, Evaluation and Ethics. The proposed scale, AILS, was based on the conceptual idea of digital literacy suggested by Balfe, Sharples, and Wilson (2018) and Calvani et al. (2008). The scale consists of 12 items concerning Awareness, Usage, Evaluation and Ethics.

Students' Voices on Generative AI by Chan & Hu (2023)

This study explores and discusses university students' perceptions of generative AI technologies, such as ChatGPT, in higher education, with a focus on familiarity with the technologies, willingness to engage, potential benefits and challenges they face, and effective integration. Undergraduate and postgraduate students from various disciplines in Hong Kong participated in the study expressing their use and perceptions of generative AI in teaching and learning. The final version of the instrument consists of a pool of 26 items, with a 5-point Likert scale ranging from "Strongly agree" to "Strongly disagree," as well as 3 open-ended questions concerning additional insights and perspectives from the participants.

Aim and Research Questions

The chapter discusses a study aiming to measure AI literacy of 1st year university students studying to be primary school teachers (pre-service primary school teachers) utilizing self-assessment and performance-based measurements. Both scales used, self-assessment and performance-based, aiming to assess participants' knowledge and skills concerning fundamentals of AI technologies and applications. Many items of self-assessment based scale are focused on generative AI technologies and applications.

More specifically, AI literacy of the 1st year university students studying to be primary school teachers is investigated assessing the Knowledge and their Competence in using AI (in terms of Awareness, Usage, Evaluation, Ethics), their Concerns about AI technologies, the Benefits and Challenges of using AI technologies they perceive, their Use of Generative AI tools and their Willingness to use AI technologies in the future.

The research questions aimed at being answered are:
- What do pre-service teachers know about AI technologies?
- What is the pre-service teachers' competence in using artificial intelligence?
- What Benefits and Challenges do students perceive using AI technologies?
- What are the pre-service teachers' concerns about and willingness to use AI technologies?

Method

Research Design and Data Collection

The research followed a correlational quantitative methodological approach, with data collection and analysis from self-assessment and performance-based questionnaires/scales distributed in two sequential phases. The sampling of the research was convenient as the participants (1st year university students – pre-service primary school teachers) invited to answered the questionnaires and they did it voluntarily and anonymously. The participants in the research were 152 1st year students of the department of Educational Sciences and Social Work of University of Patras in Greece.

Research Tools

A questionnaire (scale) was created and used as the research tools of the study consisting of four (4) parts:

Part A*: Demographic data.*

Part B*: 12 questions on the use of AI applications (5-grade Likert type scale).*

Part C*: A performance-based scale concerning knowledge about AI consisted of 60 Questions/items (questions with True/false answer, matching questions and multiple-choice questions) based on the scales of Ding 2024, Chiu 2024 and Yau, et al., 2022.*

Part D*: A self-assessment based scale concerning dimensions of AI literacy (*Competence in using AI, Knowledge of generative AI technologies, Willingness to use AI technologies, Concerns about AI technologies, Benefits of using AI technologies*), consisted of 40 Questions/items (with answers in a 5-grades Likert scale) based on the scales of Bingcheng Wang, et al., 2023 and Chan & Hu, 2023.*

Table 1. The Structure of the Questionnaire

Part	Questions/ items	Content	Source of questions/items
Part A Demographics	4	Sex, age, ICT skills level, source of knowledge concerning AI	

Part		Questions/ items	Content	Source of questions/items
Part B	Use of Generative AI	12		
Part C	AI literacy (Performance-based)	31	Knowledge	Ding 2024
		24		Chiu 2024
		5		Yau, et al., 2022
Part D	AI literacy (Self-assessment based)	12	Competence in using AI	Bingcheng Wang, et al., 2023
		6	Knowledge of generative AI technologies	Chan & Hu, 2023
		8	Willingness to use AI technologies	Chan & Hu, 2023
		4	Concerns about AI technologies	Chan & Hu, 2023
		5	Benefits of using AI technologies	Chan & Hu, 2023
		5	Challenges in using generative AI technologies	Chan & Hu, 2023

Part D was distributed one week after Parts A, B and C. A code on each questionnaire was used to pair the parts for each participant.

Validity and Reliability

To ensure the validity and reliability of the research, emphasis was given to the formulation of the items of each questionnaire to be clear, understandable, short, non-directive with a focus on the research questions (Creswell, 2011). The back-translation procedure was applied to translate and adapt the items of each scale utilized (Brislin, 1970). To ensure the content validity of the questionnaires, they were thoroughly checked by two experts, one expert in ICT in education, and one expert in AI technologies. To test face validity the questionnaires were piloted with five pre-service teachers (Creswell, 2011).

Cronbach' s alpha coefficient was utilized to check the internal consistency of the self-assessment scales and subscales used (Cronbach, 1951). The Cronbach' s alpha coefficient values calculated for the scales and subscales are shown in Table 2. All values are acceptable $(0.67 \leq \alpha \leq 0.87)$.

Table 2. Cronbach' s Alpha Coefficient Values

Scale and subscales	Number of Questions/ Items	Cronbach's Alpha
Competence in using artificial intelligence	12	0.67
Knowledge of generative AI technologies	6	0.77
Willingness to use AI technologies	8	0.8
Concerns about AI technologies	4	0.76
Benefits of using AI technologies	5	0.68
Challenges of using AI technologies are concerning	6	0.75
Use of generative AI tools	15	0.87
Overall	56	0.83

Data Analysis

For the data analysis, descriptive and inferential statistics tests utilized. Kolmogorov-Smirnov criterion used that showed non-normal data distributions $(0.08 < K\text{-}S < 0.16$; df=152; $p < 0.05)$. Therefore, non-parametric tests were used to detect statistically significant differences in the values of the factors and variables between groups and subgroups of the sample and correlations between the factors and variables. The χ^2 - goodness of fit tests, Mann-Whitney's U test for two independent samples, Kruskal-Wallis for more than two independent samples and Spearman's correlation coefficient (rs) were used. Data was processed using the SPSS 27 package.

The Sample

One hundred fifty-two (152) pre-service primary school teachers, 1st year University students, participated in the study. One hundred thirty-one (131) were women (86.2%) and twenty-one (21) were men (13.8%), with ages from 18 to 22 years old, with mean age of 18.2 years old.

Ninety-six (96) came from urban areas (63.2%) and fifty-six (56) from semi-urban/rural areas (36.8%). According to their answers only seven (7) students had studied articles on the internet related to AI, only eight (8) had attended a seminar on ai, 41 had studied scientific papers on AI, 43 had studied articles in newspapers and magazines related to AI, and 91 had discussed AI during lessons at the university.

Results and Discussion

Detailed results of the descriptive and the referential statistics applied are described assessing the Knowledge about AI fundamentals of the participants and their Competence in using AI (in terms of Awareness, Usage, Evaluation, Ethics), their Concerns about AI technologies, the Benefits and Challenges of using AI technologies they perceive, their Use of Generative AI tools and their Willingness to use AI technologies in the future. Correlations among the variables of the study and differences in the calculated variables' values for groups of the participants are also discussed.

Use of Generative AI Tools

Most of the pre-service teachers answer that they use AI tools for creating simple text content (e.g. writing emails) never or rarely (72.4), creating longer textual content (e.g. professional material, part/parts of an essay/thesis) sometimes (38.2%), collecting and finding information often or very often (54.6), summarizing large amounts of texts (e.g. articles, book chapters) sometimes (30.3) or rarely (27.6), create visual content (e.g. DALL-E, MidJourney) never or rarely (85.6), editing or creating video never or rarely (82.2), language learning (e.g. Duolingo) never or rarely (73.0), improving the drafting, style of text/foreign language text never or rarely (63.8), transcription of audio material (transcription, e.g. WhisperAI) never or rarely (90.8), grammar checking, stylization (e.g. Grammarly) never or rarely (76.0), support self-learning (e.g. explaining concepts/phenomena, asking for examples, etc.) sometimes (32.2) or often (250.), generating feedback/critique on other people's work) never or rarely (80.9), checking plagiarism, AI generated content Checking) never or rarely (63.8), correcting or creating program code) never or rarely (81.0). Most of the participants use AI tools for translation of texts (e.g. deepl.com) (sometimes: 21.1%), often:21.1%, very often:5.9%).

Table 3. Relative Frequencies for the Use of Generative AI Tools (%)

	Never	Rarely	Sometimes	Frequently	Very often
Creating simple text content (e.g. writing emails)	44.1	28.3	16.4	8.6	2.6
Creating longer textual content (e.g. professional material, part/parts of an essay/thesis)	8.6	28.3	38.2	21.1	3.9
Collecting and finding information	2.6	11.2	31.6	42.8	11.8
Summarising large amounts of texts (e.g. articles, chapters)	16.4	27.6	30.3	19.1	6.6
Create visual content (e.g. DALL-E, MidJourney)	64.5	21.1	8.6	8.6	1.3
Editing or creating video	65.8	16.4	7.9	7.9	2.0
Language learning (e.g. Duolingo)	56.6	16.4	11.8	11.2	3.9
Translation of texts (e.g. deepl.com)	25.0	27.0	21.1	21.1	5.9
Improving the drafting, style of text/foreign language text	40.1	23.7	16.4	13.8	5.9
Transcription of audio material (transcription, e.g. WhisperAI)	77.6	13.2	5.3	3.3	0.7
Grammar checking, stylization (e.g. Grammarly)	58.6	18.4	11.8	7.2	3.9
Support self-learning (e.g. explaining concepts/phenomena, asking for examples, etc.)	13.8	19.7	32.2	25.0	9.2
Generating feedback/critique on other people's work	61.2	19.7	13.2	4.6	1.3
Checking plagiarism, AI generated content	39.5	24.3	21.1	12.5	2.6
Checking, correcting or creating program code	64.2	16.8	15.2	3.9	0.0

Rarely: *once every few months,* Sometimes: *several times a month,* Frequently: *several times a week,* Very often: *every day or almost every day*

A total score for the self-assessment based sub-scale measuring the Use of Generative AI

tools was calculated by the mean value of the values of the 15 items consisting of the scale.

Table 4. Measures of Central Tendency, of Dispersion and Asymmetry for the Use of Generative AI Tools Score

min	max	mean	std. deviation	median	range	skewness	kurtosis
1.07	4.27	2.17	0.65	2.00	3.20	0.877	0.306

According to the participants' answers they use generative AI tools rarely. This is low value because most participants use AI tools for collecting and finding information often or very often, summarizing large amounts of texts sometimes, supporting self-learning sometimes or often, translation of texts sometimes, often and often but never or rarely all the other activities.

Performance-based Measurements

Table 5 and 6 present the number of pre-service teachers answered correctly and non-correctly the items of the performance-based scale.

Table 5. Questions Correctly Answered

Questions	Participants with correct answers
(T) AIs rely on algorithms to make decisions.	146 / 96.1%
(T) Data can be error-prone and require interpretation.	144/94.7%
(T) AIs are programmable.	143/94.1%
Matching the techniques that AIs use to the corresponding examples: (a) natural language processing: use a chatbot to respond to customer queries, (b) reinforcement learning: train a robot to navigate a maze, & (c) decision tree: predict whether a customer will buy a product.	141/92.8%
(T) How computers store and organize information about the world is not always fully captured.	141/92.8%
(T) AI can help humans avoid dangerous work (e.g., collecting and packaging of radioactive waste).	141/92.8%

Questions	Participants with correct answers
Which of the following devices can help in the development of speech recognition in AI? A. Acoustic sensor, B. Infrared sensor, C. Mobile phone camera, D. camera, D. Computed Tomography (CT) scanner	139/91.4%
(F) AI always makes fair decisions.	138/90.8%
(F) AI systems are always smarter than humans.	138/90.8%
(F) All AIs are created the same way.	136/89.5%
(T) AIs "see" and "hear" the world through the process of extracting information from sensory signals.	134/88.2%
(T) Self-driving cars are a type of narrow AI.	134/88.2%
(T) Data is always shaped by decisions and assumptions made during the process of data collection, processing, and analysis.	130/85.5%
Which one is not type of data: (α) numbers, (b) images, (c) texts, & (d) actions.	130/85.5%
(F) AI and machine learning are interchangeable terms.	128/84.2%
(T) Face lock feature in phones is a type of AI.	126/82.9%
(T) AI cannot learn without human input.	124/81.6%
You have trained a computer vision model to recognize pictures of cats. It works very well except when shown a hairless cat, which it does not classify at all. What might be the problem? A. Poor image quality, B. The training data did not feature any pictures of hairless cats, C. The computer vision model is flawed, D. The training data did not have enough cat images.	123/80.9%
(F) Self-driving cars only need object detection sensors for them to drive properly.	119/78.3%
(F) All human jobs will be replaced by AI in the future.	119/78.3%
(T) Machine learning is a kind of statistical inference.	116/76.3%
Which of the following actions by an AI developer are not ethical? A. Avoiding providing details of the AI application to cover the limitations of its products and services. B. Cross-checking all training data to ensure it is balanced. C. Data collected is only used for training and testing AI applications. D. Ensure that their products and services will not cause any	112/73.7%

Questions	Participants with correct answers
foreseeable or unintentional harm.	
(F) AI robots can walk along a preprogrammed path, but they cannot avoid obstacles on the path when they sense one.	105/69.1%
Which of the following operations is considered appropriate when building computer vision applications? A. Collecting and storing sensitive data during video surveillance, B. Faking others' identity during facial recognition, C. Developing medical imaging applications without tests and ethical reviews, D. Making safety the priority when building auto-driving cars.	100/65.8%
Which of the following jobs are more likely to be replaced by artificial intelligence? A. Complex clinical procedures, B. Development of military strategy, C. Simple question answers, D. Treating depressed patients.	100/65.8%
(T) AI is not entirely automated & always requires human decision-making.	100/65.8%
(F) It is easy for an AI system to recognize objects in unfamiliar or atypical situations, for example, a partially hidden pencil.	98/64.5%
(F) AI algorithms can figure out all your messy data.	95/62.5%
(T) AI cannot solve problems the way humans can.	89/58.6%
Which of the following is an application of Automatic Speech Recognition (ASR)? A. License plate recognition, B. Text translation, C. Spam classification, D. Meeting transcription	88/57.9%
(F) AI machines cannot keep updating their knowledge by using their own data.	85/55.9%
Which is correct about sentimnt analysis? A. List the topics that a document deal with, B. Compress a document as much as possible without losing meaning, producing another document, C. Assess the emotional content of a document, D. Given a question in natural language, provide an appropriate answer in natural language.	84/55.3%
Which of the following is an application of Computer Vision (CV)? A. Text summarization, B. Intelligent assistant (i.e. Siri, Alexa, etc.), C. Object detection, D. Named entity recognition	81/53.3%

Questions	Participants with correct answers
(F) AI uses same ways to organize & store information for different tasks.	80/52.6%
Put the following machine learning steps in order: (a) collect data, (b) train model, & (c) deploy the model.	79/52%

Most of the participants answered correctly 35 questions-items. The five (5) more often correctly answered questions show that they understand that *AIs are programmable and rely on algorithms to make decisions*, *Data can be error-prone, require interpretation and, how computers store and organize information about the world is not always fully captured., a decision tree application could predict whether a customer will buy a product, AI can help humans avoid dangerous work (e.g., collecting and packaging of radioactive waste)*.

The five (5) less often correctly answered questions show that the participants cannot realize that *AI machines can keep updating their knowledge by using their own data and AI uses same ways to organize & store information for different tasks, sentiment analysis applications assess the emotional content of a document, Object detection is an application of Computer Vision, the machine learning steps in order is collect data, train model, deploy the model.*

Table 6. Questions Non-correctly Answered

Questions	Participants with non-correct answers
(F) Voice-activated digital assistants (Alexa & Siri) are examples of general AI.	141/92.8%
Which of the following functions belong to AI? 1) Use human language, 2) Form Concepts, 3) Self-learning, 4) Simulate human intelligence	130/85.5%
Which of the following functions are considered ethically appropriate when developing computer vision applications? A. Collecting and storing data using video at school gates, B. Using facial recognition to learn more about	118/77.6%

Questions	Participants with non-correct answers
people, C. Keeping safety as the top priority when building self-driving cars, D. Developing medical imaging applications without informing patients.	
What criteria are used to evaluate text-to-speech (TTS) technologies? 1) Intelligibility, 2) Naturalness, 3) Similarity to the characteristics of the target speech.	113/74.3%
Which of the following is the correct explanation of "Transparency" in AI ethical principles? A. AI respects equality, inclusion and diversity, B. The AI application can be understood and explained, C. AI is accountable and honest, D. AI protects and secures sensitive data.	113/74.3%
Which of the following makes use of natural language processing (NLP) technologies? 1.) Clinical text records analysis, 2.) Gesture-controlled robot, 3.) Story generator, 4.) Questions and answers tool	102/67.1%
(F) An automatic washing machine is an example of an AI device.	98/64.5%
Which of the following could be the input source for computer vision? A. Wireless radio, B. Infrared thermograph, C. Detectaphone, D. Blood pressure monitor	97/63.8%
Which of the following mimics the operation of the human brain? A. Reasoning, B. Intelligent recommendation, C. Chatbot, D. Neural network	96/63.2%
Which one of ethical issues is least likely caused by AI if it's used inappropriately: (a) discrimination, (b) lack of accountability, (c) lack of privacy, & (d) lack of compassion	95/62.5%
Which of the following technique is used to analyze the emotion of consumers through the reviews of a product on online shopping app? A. Image Classification, B. Natural Language Processing, C. Automatic Speech Recognition, D. Text to Speech	92/60.5%
(T) Walking down a street as well as a 5-year-old can be very difficult for an AI robot.	86/56.6%
In a football match, an AI camera used to track the ball mistook the	81/53.3%

Questions	Participants with non-correct answers
referee's bald head for a football. Which of the following statements is correct? A. There may not be bald pictures in the training data of the artificial intelligence model, B. This error should be solely the responsibility of the developer of the AI model, C. The reason for this mistake must be that the camera is not high definition enough. D. People should not trust artificial intelligence technology.	
Which of the following does NOT apply AI technology? A. Web browsing, B. Facial recognition, C. Semantic analysis, D. Speech recognition	79/52%
Compared with human reasoning, what are the weaknesses of AI reasoning? A. Inability to handle highly complex tasks, B. Operating as a black box (i.e. hard to explain), C. Having negative emotions, D. Inability to process metadata.	78/51.3%
Some languages, such as Cantonese, are often NOT available in various text-to- speech systems, what is the main reason? A. Languages with insufficient data will not be able to get good training and technical support, B. There is no developer to develop the text-to-speech systems, C. Current text-to-speech system technology and processes cannot be applied to such languages, D. There is no requirement to transfer Cantonese text into speech systems.	77/50.7%

Most of the participants answered non-correctly 16 questions/items. The five (5) more often wrongly answered questions show that they cannot understand that *Voice-activated digital assistants (Alexa & Siri) are not examples of general AI, use human language and simulate human intelligence are functions that belong to AI, keeping safety as the top priority when building self-driving cars is considered ethically appropriate when developing computer vision applications, Intelligibility, Naturalness and Similarity to the characteristics of the target speech are criteria used to evaluate text-to-speech (TTS) technologies, "Transparency" in AI ethical principles means that the AI application can be understood and explained.*

The five (5) less often wrongly answered questions show that about 50% of the participants cannot realize that *Compared with human reasoning, weaknesses of AI reasoning are operating as a black box (i.e. hard to explain) and inability to process metadata, why some rare languages, are often not available in various text-to- speech systems, Why, for example, in a football match, an AI camera used to track the ball mistook the referee's bald head for a football, that walking down a street as well as a 5-year-old can be very difficult for an AI robot and Web browsing does not apply AI technology.*

A total score for the performance-based scale was calculated by the sum of the points offered for each correct answer (1 point/correct answer). The resulting score scale was from 0 points (none correct answer) to 52 points (all the answers are correct). Table 7 presents measures of central tendency, measures of dispersion and measures of asymmetry for the Knowledge total score.

Table 7. Measures of Central Tendency, of Dispersion and Asymmetry for the Knowledge Total Score

min	max	mean	std. deviation	median	range	skewness	kurtosis
20	46	32.34	4.65	33.00	26	-0.387	0.283

According to the values in Table 8, the performance level of the participating students can be considered moderate. Most participants answered correctly 35 questions/items and non-correctly 17 questions/items. There were no statistically significant differences in the scores considering the sex, the origin area, the attendance of AI related seminars, the study of AI related articles on the internet or scientific papers on AI, or articles in newspapers and magazines, the discussion about AI during lessons at the university.

Self-assessment based Measurements

The Self-assessment based measurements assessed the competence of the participants in using AI (in terms of Awareness, Usage, Evaluation, Ethics), their Concerns about AI technologies, their perceived Benefits and Challenges of using AI technologies and their Willingness to use AI technologies in the future.

Competence in using AI

A total score for the self-assessment based scale measuring the competence in using AI was calculated by the mean value of the values of the 12 items consisting of the scale. The total scores for the subscales Awareness (3 items), Usage (3 items), Evaluation (3 items), Ethics (3 items) were calculated in the same way. Table 8 presents measures of central tendency, measures of dispersion and measures of asymmetry for the Knowledge total score and subscales scores.

Table 8. Measures of Central Tendency, of Dispersion and Asymmetry for the Competence in using AI Scores

scale	min	max	mean	std. deviation	median	range	skewness	kurtosis
Awareness	2.00	5.00	3.60	0.55	3.67	3.00	-0.094	0.012
Usage	2.00	5.00	3.63	0.62	3.67	3.00	-0.366	-0.220
Evaluation	2.00	5.00	3.41	0.55	3.33	3.00	-0.090	0.449
Ethics	1.00	5.00	3.67	0.72	3.67	4.00	-0.392	0.394
Total scale	2.42	4.67	3.58	0.42	3.62	2.25	-0.261	-0.011

According to the values in Table 8, the competence in using AI level of the participating students can be considered high. Levels of competence concerning Awareness, Usage, Evaluation and Ethics can be considered moderate to high, moderate to high, moderate and moderate to high respectively. Statistically significant differences between men and women arose concerning their perceived competence in using AI ($U(152) = 876.00$; $Z = -2.710$; $p < 0.05$; $r = -0.22$) as women showed higher scores (mean rank = 80,31) than men (mean rank =52,71). Also, students who had discussed AI during their lessons showed statistically significant ($U(152) = 1939.00$; $Z = -3.184$; $p < 0.05$; r = -0.26) higher scores concerning the ethics dimension of the competence in use AI scale. Furthermore, students who study books and scientific papers about AI showed statistically significant ($U(152) = 1802.00$; $Z = -1.990$; $p < 0.05$; r = -0.16) higher scores concerning the ethics dimension too.

Knowledge of Generative AI Technologies

A total score for the self-assessment based sub-scale measuring the Knowledge of generative

AI technologies was calculated by the mean value of the values of the 6 items consisting of the scale.

Table 9. Measures of Central Tendency, of Dispersion and Asymmetry for the Knowledge of Generative AI Technologies Score

min	max	mean	std. deviation	median	range	skewness	kurtosis
1.67	5.00	3.69	0.68	3.83	3.33	-0.217	-0.387

According to the values in Table 9, the level of Knowledge of generative AI technologies for the participating students can be considered moderate high.

Students who had discussed AI during their lessons showed statistically significant ($U(152) = 1939.00$; $Z = -3.184$; $p < 0.05$; $r = -0.26$) higher scores concerning the Knowledge of generative AI technologies subscale. Furthermore, students who study books and scientific papers about AI showed statistically significant ($U(152) = 2244.00$; $Z = -2.005$; $p < 0.05$; $r = -0.16$) higher scores concerning the ethics dimension too. Students who study books and scientific papers about AI showed statistically significant ($U(152) = 1790.50$; $Z = -2.020$; $p < 0.05$; $r = -0.16$) higher scores concerning the same subscale too.

Concerns about AI Technologies

A total score for the self-assessment based sub-scale measuring the Concerns about AI technologies was calculated by the mean value of the values of the 4 items consisting of the scale.

Table 10. Measures of Central Tendency, of Dispersion and Asymmetry for the Concerns about AI Technologies *Score*

min	max	mean	std. deviation	median	range	skewness	kurtosis
1.00	5.00	3.35	0.87	3.37	4.00	-0.164	-0.543

According to the values in table 10, the Concerns about AI technologies level of the participating students can be considered moderate. No statistically significant differences

arose concerning any demographic data of the participants and the scores of the Concerns about AI technologies subscale.

Benefits and Challenges of using AI Technologies

Total scores for the self-assessment based sub-scales measuring the Benefits and Challenges of using AI technologies were calculated by the mean value of the values of the 5 and 6 items consisting of the scales correspondingly.

Table 11. Measures of Central Tendency, of Dispersion and Asymmetry for the Benefits and Challenges of using AI Technologies Scores

scale	min	max	mean	std. deviation	median	range	skewness	kurtosis
Benefits	1.8	5.00	3.49	0.58	3.40	3.20	-0.139	0.459
Challenges	1.67	4.83	3.34	0.66	3.33	3.17	-0.132	-0.215

According to the values in Table 11, the level of perceived benefits of the participating students can be considered moderate, and the level of perceived challenges can be considered moderate tool. No statistically significant differences arose concerning any demographic data of the participants and the scores of the Benefits and Challenges of using AI technologies subscale.

Willingness to use AI Technologies in the Future

A total score for the self-assessment based sub-scale measuring the Willingness to use AI technologies was calculated by the mean value of the values of the 8 items consisting of the scale.

Table 12. Measures of Central Tendency, of Dispersion and Asymmetry for the Willingness to Use AI Technologies in the Future Score

min	max	mean	std. deviation	median	range	skewness	kurtosis
1.75	4.88	3.72	0.57	3.75	3.13	-0.833	1.375

According to the values in Table 12, the Willingness to use AI technologies in the future level of the participating students can be considered moderate to high.

Statistically significant differences between men and women arose concerning their willingness to use AI ($U(152) = 1000.50$; $Z = -2.009$; $p < 0.05$; $r = -0.16$) with men showing higher scores than women. Furthermore, students who study articles in newspapers and magazines about AI showed statistically significant ($U(152) = 1855.00$; $Z = -2.005$; $p < 0.05$; $r = -0.16$) higher scores concerning the Willingness to use AI technologies in the future subscale.

Correlations

Statistically significant correlations arose among the variables of the study as they supported by the scales and subscales scores with the use of the non-parametric Spearman's correlation coefficient (r_s).

The *knowledge about AI* (scores of performance-based scale) is related positively to the *Competence in using artificial intelligence* (rs $(152)= 0.19$; $p < 0.05$; low effect size), the *Knowledge of generative AI technologies* (rs $(152)= 0.28$; $p < 0.01$; low effect size), *Challenges of using AI technologies* are concerning AI (rs $(152)= 0.20$; $p < 0.05$; low effect size) but the *knowledge* about AI (scores of performance-based scale) is related negatively to the *Use of generative AI applications* (rs $(152) = -0.23$; $p < 0.01$; low effect size).

The *Competence in using artificial intelligence* is related positively to *Challenges of using AI technologies* (rs $(152)= 0.25$; $p < 0.05$; low effect size), the *Knowledge of generative AI technologies* (rs $(152)= 0.24$; $p < 0.01$; low effect size) and the *Willingness to use AI technologies* (rs $(152)= 0.22$; $p < 0.01$; low effect size.

The dimension *Usage* of the Competence in using artificial intelligence is related positively to the dimension *Evaluation* (rs $(152)= 0.46$; $p < 0.01$; moderate effect size), the dimension *Ethics* (rs $(152)= 0.20$; $p < 0.01$; low effect size) and the *Challenges of using AI technologies* (rs $(152)= 0.20$; $p < 0.05$; low effect size). Also, it is related positively to the *Willingness to use AI technologies* (rs $(152)= 0.26$; $p < 0.01$; low effect size). Furthermore, the dimension *Evaluation* is related positively to the *Willingness to use AI technologies* (rs $(152)= 0.31$; $p <$

0.01; moderate effect size), to the *Challenges of using AI technologies* (*rs* (152)= 0.27; *p* < 0.01; low effect size), to the *Knowledge of generative AI technologies* (*rs* (152)= 0.24; *p* < 0.01; low effect size) as well as to the *age* of the student (*rs* (152)= 0.17; *p* < 0.05; low effect size). It also related positively with the dimension *Ethics* (*rs* (152)= 0.25; *p* < 0.01; low effect size). The dimension *Ethics* is related positively to the *Knowledge of generative AI technologies* (*rs* (152)= 0.19; *p* < 0.05; moderate effect size).

The *Knowledge of generative AI technologies* is related positively to the *Challenges of using AI technologies* (*rs* (152)= .32; *p* < .01; moderate effect size), *Willingness to use AI technologies* (*rs* (152)= 0.22; *p* < 0.01; low effect size). On the other hand, *Benefits of using AI technologies* are related to Challenges of using AI technologies TN (*rs* (152)= 0.31; *p* < 0.01; moderate effect size).

Finally, *Willingness to use AI technologies* is realetd positively to *Benefits of using AI technologies* (*rs* (152)= 0.45; *p* < 0.01; moderate effect size) but also with the *Challenges of using AI technologies* (*rs* (152)=0.25; *p* < 0.01; low effect size).

Conclusion

The chapter discusses a study aiming to measure AI literacy of 1st year university students studying to be primary school teachers (pre-service primary school teachers) utilizing self-assessment and performance-based measurements. Both scales used, self-assessment and performance-based, aiming to assess participants' knowledge and skills concerning fundamentals of AI technologies and applications. Many items of self-assessment based scale are focused on generative AI technologies and applications.

The research questions of the study concerning what do pre-service teachers know about AI technologies, what is their competence in using artificial intelligence, what are their concerns about and their willingness to use AI technologies as well as what Benefits and Challenges do they perceive using AI technologies. One hundred fifty-two (152) pre-service primary school teachers, 1st year University students, participated in the study.

According to the participants' answers they use generative AI tools not often. Most participants use generative AI tools often or very often for collecting and finding information,

summarizing large amounts of texts sometimes, supporting self-learning, translation of texts, but never or rarely for creating simple text content (e.g. writing emails), creating longer textual content (e.g. professional material, part/parts of an essay/thesis), creating visual content (e.g. DALL-E, MidJourney), editing or creating video, language learning (e.g. Duolingo), improving the drafting, styling of text/foreign language text, transcription of audio material (transcription, e.g. WhisperAI), grammar checking, stylization (e.g. Grammarly), generating feedback/critique on other people's work, checking plagiarism, AI generated content, checking, correcting or creating program code.

Most of the participants answered correctly 35 questions-items of the per4formance-based questionnaire. The more often correctly answered questions show that they understand that AIs are programmable and rely on algorithms to make decisions, Data can be error-prone, require interpretation and, how computers store and organize information about the world is not always fully captured., a decision tree application could predict whether a customer will buy a product, AI can help humans avoid dangerous work (e.g., collecting and packaging of radioactive waste). The less often correctly answered questions show that the participants cannot realize that AI machines can keep updating their knowledge by using their own data and AI uses same ways to organize & store information for different tasks, sentiment analysis applications assess the emotional content of a document, Object detection is an application of Computer Vision, the machine learning steps in order is collect data, train model, deploy the model.

Most of the participants answered non-correctly 16 questions/items of the same questionnaire. The more often wrongly answered questions show that they cannot understand that Voice-activated digital assistants (Alexa & Siri) are not examples of general AI, use human language and simulate human intelligence are functions that belong to AI, keeping safety as the top priority when building self-driving cars is considered ethically appropriate when developing computer vision applications, Intelligibility, Naturalness and Similarity to the characteristics of the target speech are criteria used to evaluate text-to-speech (TTS) technologies, "Transparency" in AI ethical principles means that the AI application can be understood and explained. The less often wrongly answered questions show that about 50% of the participants cannot realize that Compared with human reasoning, weaknesses of AI reasoning are operating as a black box (i.e. hard to explain) and inability to process metadata, why some rare languages, are often not available in various text-to- speech systems, Why, for

example, in a football match, an AI camera used to track the ball mistook the referee's bald head for a football, that walking down a street as well as a 5-year-old can be very difficult for an AI robot and Web browsing does not apply AI technology.

The performance level of the participating students can be considered moderate. There were no statistically significant differences in the scores considering the sex, the origin area, the attendance of AI related seminars, the study of AI related articles on the internet or scientific papers on AI, or articles in newspapers and magazines, the discussion about AI during lessons at the university.

The competence in using AI levels of the participating students can be considered high. Levels of competence concerning Awareness, Usage, Evaluation and Ethics can be considered moderate to high, moderate to high, moderate and modarate to high respectively. Women showed higher scores for the perceived competence in using AI scale than men. Furthermore, students who had discussed AI during their lessons showed higher scores concerning the ethics dimension of the competence in use AI scale and students who study books and scientific papers about AI showed higher scores concerning the ethics dimension too.

The level of Knowledge of generative AI technologies for the participating students can be considered moderate to high. Students who had discussed AI during their lessons showed higher scores concerning the Knowledge of generative AI technologies subscale. The Concerns about AI technologies level of the participating students can be considered moderate. No statistically significant differences arose concerning any demographic data of the participants and the scores of the Concerns about AI technologies subscale. The level of perceived benefits of the participating students can be considered moderate, and the level of perceived challenges can be considered moderate too. No statistically significant differences arose concerning any demographic data of the participants and the scores of the Benefits and Challenges of using AI technologies subscale

The Willingness to use AI technologies at the future level of the participating students can be considered moderate to high. Differences between men and women arose concerning their willingness to use AI with men showing higher scores than women. Furthermore, students who study articles in newspapers and magazines about AI showed higher scores concerning

the Willingness to use AI technologies in the future subscale.

Statistically significant correlations arose among the variables of the study as they supported by the scales and subscales scores. The knowledge about AI (scores of performance-based scale) is related positively but not strongly to the Competence in using artificial intelligence, the Knowledge of generative AI technologies, Challenges of using AI technologies are concerning AI and negatively but not strongly to the Use of generative AI applications. The Competence in using artificial intelligence is related positively but not strongly to Challenges of using AI technologies, the Knowledge of generative AI technologies and the Willingness to use AI technologies. The dimension Ethics of the Competence in using artificial intelligence is related positively to the Knowledge of generative AI technologies. Finally, as expected, Willingness to use AI technologies is realetd positively to Knowledge of generative AI technologies, to Usage and Evaluation dimensions of the Competence in using artificial intelligence, perceived Benefits of using AI technologies but also with the Challenges of using AI technologies.

There are some limitations that affect the generalization of the results and conclusions potentiality as the sample size is not very large, the participants are students of one only department of one only university at one country. But they can contribute to the discussion concerning cultivating artificial intelligence literacy of individuals to use and critically evaluate AI tools and to collaborate effectively with them in multiple contexts. More extensive research employing more specific referential statistics could give more valid conclusions and is needed.

Recommendations

As the opportunities and challenges arising from the pervasiveness of AI in society at all levels, AI Literacy is becoming a crucial issue. The use of high-quality, valid, and reliable AI literacy tools is vital for understanding and promoting AI literacy growth. More extensive research with the same tools and research questions on the same basis as the present research could offer additional data and thoughts to the field under consideration. Participants from all study years of primary school teachers' preparation or from all the departments of the university with universities participating from all Europe countries and more would offer valuable data and insights. Secondary school students would offer valuable insight too. The

AI literacy of in-service teachers, at all levels of education, is extremely crucial as they have to work with their students on AI related activities.

Acknowledgements

References

Antonenko, P., & Abramowitz, B. (2023). In-service teachers' (mis)conceptions of artificial intelligence in K-12 science education. *Journal of Research on Technology in Education, 55*(1), 64-78. https://doi.org/10.1080/15391523.2022.2119450

Balfe, N., S. Sharples, and J. R. Wilson. (2018). "Understanding Is Key: An Analysis of Factors Pertaining to Trust in a Real-World Automation System." *Human Factors* 60 (4): 477–495. doi:10.1177/0018720818761256.

Biagini, G., Cuomo, S., & Ranieri, M. (2024). Developing and Validating a Multidimensional AI Literacy Questionnaire: Operationalizing AI Literacy for Higher Education. *Proceedings of the First International Workshop on High-performance Artificial Intelligence Systems in Education*.

Bingcheng Wang, Pei-Luen Patrick Rau & Tianyi Yuan (2023) Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale, *Behaviour & Information Technology, 42*:9, 1324-1337, DOI: 10.1080/0144929X.2022.2072768

Calvani, A., A. Fini, and M. Ranieri. (2009). "Assessing Digital Competence in Secondary Education. Issues, Models and Instruments." Issues in Information and Media Literacy: *Education, Practice and Pedagogy*, 153–172

Carolus, A., Koch, M., Straka, S., Latoschık, M. E., Wıenrıch, C. (2023). MAILS—meta AI literacy scale: development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies.

Comput. Hum. Behav. 1, 100014.

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A systematic literature review. *International Journal of STEM Education, 10*(1), 29. https://doi.org/10.1186/s40594-023-00418-7

Cave, S., Coughlan, K., & Dihal, K. (2019). "Scary robots" examining public responses to AI. *In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 331-337). ACM. https://doi.org/10.1145/3306618.3314232

Cetindamar, D., Kitto, K., Wu, M., Zhang, Y., Abedin, B., & Knight, S. (2022). Explicating AI literacy of employees at digital workplaces. *IEEE Transactions on Engineering Management, 71*, 810-823. https://doi.org/10.1109/TEM.2021.3138503

Chan, C.K.Y., Hu, W. Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *Int J Educ Technol High Educ* **20**, 43 (2023). https://doi.org/10.1186/s41239-023-00411-8

Chiu, T. K. F., Ahmand, Z., Ismail, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers & Education Open, 6,* 100171.

Chiu, T. K. F., Meng, H., Chai, C. S., King, I., Wong, S., & Yeung, Y. (2022). Creation and evaluation of a pre-tertiary Artificial Intelligence (AI) curriculum. *IEEE Transactions on Education, 65*(1), 30–39. http://dx.doi.org/10.1109/TE.2021.3085878.

Davenport, T. H., and R. Ronanki. (2018). "Artificial Intelligence for the Real World." *Harvard Business Review 96* (1): 108–116

Ding, L., Kim, S., & Allday, R. A. (2024). Development of an AI literacy assessment for non-technical individuals: What do teachers know? Contemporary Educational Technology, 16(3), ep512. https://doi.org/10.30935/cedtech/14619

Ding, L., Li, T., Jiang, S., & Gapud, A. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *International Journal of Educational Technology in Higher Education, 20*(1), 63. https://doi.org/10.1186/s41239-023-00434-1

Du, H., Sun, Y., Jiang, H. *et al.* Exploring the effects of AI literacy in teacher learning: an empirical study. *Humanit Soc Sci Commun* **11**, 559 (2024). https://doi.org/10.1057/s41599-024-03101-6

Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about artificial intelligence? Development and validation of an AI literacy test. *Computers in Education: Artificial Intelligence, 5*, Article 100165.

Jackson, C., Mohr-Schroeder, M. J., Bush, S. B., Maiorca, C., Roberts, T., Yost, C., &

Fowler, A. (2021). Equity-oriented conceptual framework for K-12 STEM literacy. *International Journal of STEM Education, 8*, 38. https://doi.org/10.1186/s40594-021-00294-z

Jacob, S. R., & Warschauer, M. (2018). Computational thinking and literacy. *Journal of Computer Science Integration, 1*(1). https://doi.org/10.26716/jcsi.2018.01.1.1

Kandlhofer, M., G. Steinbauer, S. Hirschmugl-Gaisch, and P. Huber. (2016). "Artificial Intelligence and Computer Science in Education: From Kindergarten to University." *2016 IEEE Frontiers in Education Conference (FIE)*, Erie, PA, USA, 2016, pp. 1-9, doi: 10.1109/FIE.2016.7757570.

Karaca, O., Çalışkan, S. A. & Demir, K. (2021). Medical artificial intelligence readiness scale for medical students (MAIRS-MS)–development, validity and reliability study. BMC Med. Educ. 21, 1–9.

Karatrantou, A., Panetas, G., Panagiotakopoulos, C. (2025a). Primary school Teachers' Artificial Intelligence Literacy and Anxiety. *Proceedings of 18th annual International Conference of Education, Research and Innovation*. Seville (Spain). 10- 12/11/2025.

Karatrantou, A., Panetas, G., Panagiotakopoulos, C. (2025b). ChatGPT, MagicSchoolAI and Wiserwork supporting Primary School teachers to design educational scenarios: A Case Study . *Proceedings of 18th annual International Conference of Education, Research and Innovation*. Seville (Spain). 10- 12/11/2025.

Karatrantou, A., Panetas, G., Panagiotakopoulos, C. (2025c). Pre-service teachers' acceptance of Machine Learning (ML) tools for primary education students. *Proceedings of International Conference on Educational Technology and Online Learning*, Cunda, Ayvalık, Balıkesir, Türkiye, 26-29/08/2025.

Knoth, N., Decker, M., Laupichler, M. C., Pinski, M., Buchholtz, N., Bata, K., & Schultz, B. (2024). Developing a holistic AI literacy assessment matrix–Bridging generic, domain-specific, and ethical competencies. *Computers and Education Open, 6*, Article 100177. https://doi.org/10.1016/j.caeo.2024.100177

Kong, S.-C., Man-Yin Cheung, W., & Zhang, G. (2023). Evaluating an artificial intelligence literacy program for developing university students' conceptual understanding, literacy, empowerment, and ethical awareness. *Educational Technology & Society, 26*(1), 16-30. https://doi.org/10.30191/ETS.202301_26(1).0002

Laupichler, M. C., Aster, A., Haverkamp, N., Raupach, T. (2023) Development of the "Scale for the assessment of non-experts' AI literacy"—ān exploratory factor analysis. *Comput. Hum. Behav. Rep. 12*, 100338.

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. In *Computers and education: Artificial intelligence*. Elsevier. https://doi.org/10.1016/j.caeai.2022.100101

Lee, I., Ali, S., Zhang, H., Dipaola, D., & Breazeal, C. (2021). Developing middle school students' AI literacy. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 191-197). ACM. https://doi.org/10.1145/3408877.3432513

Lee, S. & Park, G. (2024). Development and validation of ChatGPT literacy scale. *Curr. Psychol.* https://doi.org/10.1007/s12144-024-05723-0.

Lintner, T. (2024). A systematic review of AI literacy scales. *npj Sci. Learn. 9*, 50 (2024). https://doi.org/10.1038/s41539-024-00264-4

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *In Proceedings of the Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376727

Long, D., Teachey, A., & Magerko, B. (2022). Family learning talk in AI literacy learning activities. In *Proceedings of the Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3502091

Maitz, K., Fessl, A., Pammer-Schindler, V., Kaiser, R., & Lindstaedt, S. (2022). What do construction workers know about artificial intelligence? An exploratory case study in an Austrian SME. *In Proceedings of the ACM International Conference* (pp. 389-393). https://doi.org/10.1145/3543758.3547545

Martin, A., & Grudziecki, J. (2006). DigEuLit: Concepts and tools for digital literacy development. *Innovation in Teaching and Learning in Information and Computer Sciences, 5*(4), 249–267. https://doi.org/10.11120/ital.2006.05040249

Mertala, P., Fagerlund, J., & Calderon, O. (2022). Finnish 5th and 6th grade students' pre-instructional conceptions of artificial intelligence (AI) and their implications for AI literacy education. *Computers and Education: Artificial Intelligence, 3*. https://doi.org/10.1016/j.caeai.2022.100095

Miao, F., Shiohira, K. (2022). *K-12 AI curricula. A mapping of government-endorsed AI curricula*. UNESCO. Document code: ED-2022/FLI-ICT/K-12 REV.

Mokkink, L. B. et al. (2020). COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med. Res. Methodol. 20*, 1–13.

Nader, K., Toprac, P., Scott, S., & Baker, S. (2022). Public understanding of artificial intelligence through entertainment media. *AI and Society, 39,* 713-726. https://doi.org/10.1007/s00146-022-01427-w

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F, Chu, S. K. W. Design and validation of the AI literacy questionnaire: the affective, behavioural, cognitive and ethical approach. *Br. J. Educ. Technol. 54*, 1–23 (2023).

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021a). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology, 58*(1), 504-509.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021b). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence,* 2. https://doi.org/10.1016/j.caeai.2021.100041

Olari, V. (2023). Introducing artificial intelligence literacy in schools: A review of competence areas, pedagogical approaches, contexts and formats. In T. Keane, C. Lewin, T. Brinda, & R. Bottino (Eds.), *Towards a collaborative society through creative learning*. Springer. https://doi.org/10.1007/978-3-031-43393-1_21

Ow-Yeong, Y. K., Yeter, I. H., & Ali, F. (2023). Learning data science in elementary school mathematics: A comparative curriculum analysis. *International Journal of STEM Education, 10*, 8. https://doi.org/10.1186/s40594-023-00397-9

Pedro, E., Subosa, M., Rivas, A., Valverde, P. (2019). Artificial Intelligence in education: Challenges and opportunities for sustainable development. *Working papers on education policy, 7* [14] Document code: ED-2019/WS/8,

Pinski, M. & Benlian, A. (2023). AI Literacy-towards measuring human competency in artificial intelligence. In *Hawaii International Conference on System Sciences*. https://hdl.handle.net/10125/102649.

Prinsen, C. A. et al. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual. Life Res. 27*, 1147–1157.

Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioral and Healthcare Research, 2*(4), 320. https://doi.org/10.1504/IJBHR.2011.043414

Soto-Sanfiel, M. T., Angulo-Brunet, A. & Lutz, C. (2024). The scale of artificial intelligence literacy for all (SAIL4ALL): a tool for assessing knowledge on artificial intelligence in all adult populations and settings. Preprint at arXiv https://osf.io/bvyku/.

Thomas K.F. Chiu, Yifan Chen, King Woon Yau, Ching-sing Chai, Helen Meng, Irwin

King, Savio Wong, Yeung Yam. (2024). Developing and validating measures for AI literacy tests: From self-reported to objective measures, *Computers and Education: Artificial Intelligence, 7,* 100282, ISSN 2666-920X, https://doi.org/10.1016/j.caeai.2024.100282.

Touretzky, D., Gardner-Mccune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence* (pp. 9795-9799).

Wang, B., Rau, P. L. P. & Yuan, T. (2022). Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behav. Inf. Technol. 42*, 1324–1337

Wang, Y. Y., & Chuang, Y. W. (2023). Artificial intelligence self-efficacy: Scale development and validation. Education and Information Technologies. https://doi.org/10.1007/s10639-023-12015-w. *Advanced Online Publication.*

Weber, P., Pinski, M., & Baum, L. (2023). Toward an objective measurement of AI literacy. In PACIS 2023 proceedings (Vol. 60). https://aisel.aisnet.org/pacis2023/60.

Wolff, A., Gooch, D., Cavero Montaner, J.J, Rashid, U., Kortuem, G., (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics, 12*(3),9-26.

Wong, G. K. W., & Huan, J. (2020). Broadening artificial intelligence education in K-12: Where to start? *ACM Inroads, 11*(1), 20-29.

Yang, W. (2022). Artificial intelligence education for young children: why, what, and how in curriculum design and implementation. *Comput. Educ. 3*, 100061.

Yang, Y., Zhang, Y., Sun, D. *et al.* Navigating the landscape of AI literacy education: insights from a decade of research (2014–2024). *Humanit Soc Sci Commun* **12**, 374 (2025). https://doi.org/10.1057/s41599-025-04583-8

Yau, K., Chai, C., Chiu, T., Meng, H., King, I., Wong, S., Saxena, C., Yam, Y. (2022). Developing an AI literacy test for junior secondary students: The first stage. *2022 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, Hung Hom, Hong Kong, 59-64.

Yue, M., Jong, M. S. Y., & Dai, Y. (2022). Pedagogical design of K-12 artificial intelligence education: A systematic review. *Sustainability, 14*(23), 15620. https://doi.org/10.3390/su142315620

Zhang, H., Perry, A. & Lee, I. (2024). Developing and validating the artificial intelligence literacy concept inventory: an instrument to assess artificial intelligence literacy

among middle school students. *Int. J. Artif. Intell. Educ.*

Creswell, J. (2011). *Educational Research Planning: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Pearson Education. ISBN: 9780131367395

Brislin, R. W. (1970). Back-Translation for Cross-Cultural Research. Journal of Cross-Cultural Psychology, 1, 185-216. https://doi.org/10.1177/135910457000100301

Cronbach, L. J. (1951). "Coefficient alpha and the internal structure of tests". *Psychometrika. 16* (3). Springer Science and Business Media LLC: 297–334. doi:10.1007/bf02310555. hdl:10983/2196

Author Information

Anthi Karatrantou

https://orcid.org/0000-0002-5347-6251

University of Patras

26504 Rio

Greece

Contact e-mail: *akarat@upatras.gr*

George Panetas

https://orcid.org/0009-0008-7914-7182

University of Patras

26504 Rio

Greece

Christos Panagiotakopoulos

https://orcid.org/0000-0003-3162-9218

University of Patras

26504 Rio

Greece

Citation

Chapter 3 - Surveying Attitudes towards AI at Western: Promises, Perils and Possibilities

Raymond Archee , **Levi Osuagwu** , **Myra Gurney** , **Walid Bakry** , **Mauricio Novoa Muñoz** , **Edward Anyasodor**

Chapter Highlights

- This chapter investigated attitudes towards artificial intelligence (AI) and surveillance among academics, students, and professional staff at a large Australian university using a mixed-methods survey approach.
- The attitude scale results show both positive and negative attitudes towards AI with direct experience with AI often leading to higher optimism about the technology and lesser familiarity leading to higher pessimism.
- Open-ended responses suggest a shared understanding of AI as involving computer software, data, tasks, learning, information, machines, and algorithms, with sentiment analysis revealing nuanced differences in perceptions between university roles.
- These findings reveal unrecognized knowledge about how individuals and their roles engage with emerging AI technologies and digital surveillance within a total university context.

Introduction

In recent years, there has been increasing use of digital technologies in a range of industries such as education, finance, health, manufacturing, agriculture and commercial transportation. The evolution of social media, especially, has seen AI-generated agents increasingly used to monitor, moderate and determine what we see and hear, as well as what we buy and sell, on these platforms. Software companies such as Microsoft and Google have been constantly using aggregated user data to produce metrics that resemble surveillance techniques creating a possible new era of scientific management.

The latest new technology to gain public attention is artificial intelligence (AI). The idea of a computer imitating a human arguably started with Alan Turing's (1950) paper that imagined a benchmark by which a machine could be questioned and judged to exhibit AI, or not – the so-called Turing Test. Since that time, the AI field has slowly, but assiduously, gained global attention via films such as *2001: A Space Odyssey,* mainstream and social media, and popular culture.

While ChatGPT became world news with its public release in late 2022, AI comprises many other kinds of applications. These applications utilise machine learning, natural language processing, and computer-based perception enabling programs to simulate human intelligence and perform tasks traditionally requiring human intervention. In the context of higher education, a range of existing AI applications such as *Dialogflow* (from Google), *Querium* and *Classcraft* have the potential to enhance teaching and learning. AI's potential benefits extend to the streamlining of administrative processes and improving student support services. The dark side of AI includes compromises to personal privacy and copyright issues, digital surveillance, bias and erroneous decision-making. The advent of generative AI (GenAI) has allowed students to bypass the traditional norms and processes of scholarship, potentially eroding the integrity and reputation of academic qualifications and institutions.

The initial implications of the widespread adoption of GenAI triggered diverse reactions internationally. The Italian government, for example, originally banned its citizens from using ChatGPT, citing privacy concerns due to the "massive collection and processing of personal data in order to train the algorithms on which the platform relies" (Browne, 2023, Apr 4). The United Kingdom government announced plans to regulate but not ban the use of

AI while the Australian government did not originally take any position. While parent company, *OpenAI* has admitted that its early public release was a huge beta test of the technology, the company faced fines of up to 20 million euros if it did not conform to the European Union's strict privacy policies (Meyer, 2023, Apr 1). In 2025, most governments acknowledge AI tools are inevitable, and potentially beneficial to their future economies but are still struggling with ethical, security and regulatory problems yet to be solved (Google Gemini, 9 Aug, 2025).

Literature Review

While the introduction of OpenAI's ChatGPT in November 2022 caused a major disruption to traditional norms of teaching and learning and research in universities, the applications of AI in education has been practised for several decades, rapidly evolving and reshaping the overall educational landscape. More broadly, the implications for the use of AI and GenAI is being widely discussed in terms of its impact on economies, and higher education (Rudolph, Tan & Tan, 2023; Wang, et al., 2024). Major uses and abuses of AI and GenAI in higher education encompass four main clusters of research: 1. intelligent tutoring systems and personalised learning, 2. automation of administrative processes including grading, 3. ethical considerations, and 4. adaptive assessment (including would-be cheating).

Intelligent tutoring systems can use AI algorithms to provide personalised instruction, adapting to individual student needs and learning styles. These systems offer immediate feedback, track student progress, and tailor instructional content. Studies have shown that the use of AI algorithms for tutoring can improve student engagement, knowledge acquisition, and learning outcomes (Alam, 2023). There are reports that adaptive learning platforms such as *Knewton* and *Smart Sparrow*, use AI algorithms to dynamically adjust course content based on individual learner performance, promoting personalised learning pathways (Johnson et al., 2015). AI algorithms can also analyse student interactions with learning management systems and discussion forums, identifying learning gaps, and suggesting relevant resources (Cantabella, et al., 2019). Student motivation, engagement, and knowledge retention are all supposedly measurably enhanced, using AI technology.

AI technologies in higher education can possess broader, more administrative functions. They can simplify enrolment and admission processes by automating the analysis of an applicant's

qualification and previous results, thus removing bias in the decision-making process (Gaftandzhieva et al., 2023). The software can also assist in predicting student enrolment patterns, enabling institutions to optimise resource allocation, and plan for future student intakes (see Sihare, 2024). Intelligent academic advising systems can also bestow guidance to students, helping them understand course selection, degree planning, and possible career pathways (Hamadneh, et al., 2022). Such systems can analyse student records, including academic performance and career interests, and offer more desirable alternatives. By considering student preferences and faculty availability, course scheduling can be augmented and improved (Huang, et al., 2021).

Some ambitious institutions are currently employing AI chatbots and virtual assistants to perform tasks, usually via their websites. These intelligent agents can provide 24/7 support by replying to frequently asked questions, helping students with administrative tasks, and recommending academic options (Labadze et al., 2023). GenAI-based chatbots employ natural language to understand and respond to student queries, which increases the availability and responsiveness of support services. By offering individualised recommendations and assistance, chatbots can potentially offer academic and professional advising services. Software tools can provide specific guidance on course choices, internships, and employment prospects, and can examine student data, such as academic achievement and professional interests (Gill, et al., 2024). AI-driven tools can enable students to make highly informed decisions regarding their educational and career pathways.

AI applications can also improve student engagement and retention by identifying at-risk students and implementing targeted interventions. Predictive analytics models can analyse student records, such as academic performance, attendance, and engagement patterns, to identify early warning signs of disengagement, absenteeism or potential dropouts (Almalawi, et al., 2024). Institutions can then provide personal support, such as additional tutoring or mentoring, to support student success and retention.

However, there are a range of policy issues and concerns to be considered before institutions can confidently and ethically employ AI technologies. For example, the integration of AI in higher education raises concerns about algorithmic bias and potential discrimination based on race, gender, and social class (Chen, et al., 2023). AI systems can inadvertently perpetuate biases present in the data used for training, resulting in unequal treatment of individuals

(Noble, 2018). Research concludes that institutions must ensure transparency and equity in AI algorithms to avoid possible errors, and to uphold fairness in decision-making processes (Floridi, et al., 2018). The use of AI technologies in higher education would involve the collection and analysis of large amounts of student data, meaning that institutions need to prioritise data privacy and protection to maintain student trust and comply with relevant regulations. Safeguards such as anonymity, encryption, and secure data storage should be implemented to protect sensitive student information (Daniel, 2020).

AI-based assessment systems can offer adaptive evaluation methods that adjust the difficulty and type of questions based on individual learner responses. Such adaptive assessments provide immediate feedback and generate detailed analytics on student performance. Research suggests that adaptive assessments improve the accuracy of measuring student proficiency, allowing for tailored interventions and support (Gligorea et al., 2023). AI-enabled assessment tools hold the potential to enhance the effectiveness and efficiency of evaluation processes in all industries, including higher education.

One pressing AI assessment problem experienced in higher education has been the student use of ChatGPT to create their assignments without acknowledging this usage. One study found half of the participants admitted to being tempted to resort to GenAI tools if the need arose (Ventayen, 2023). Formal policies regarding acceptable AI use at most higher education institutions have been slow to be formulated but such student behaviour appears unfair and cannot be easily identified or regulated. For some institutions, the academic program, *Turnitin* has become the only method of revealing student AI usage, but *Turnitin* itself disavows 100 per cent reliability of its detection processes. Our own experiments with Turnitin show that the AI percentage drops if one uses other software to paraphrase ChatGPT output or manually revise the output text. After the second or third paraphrasing iteration, the Turnitin score drops to zero.

Surveys of attitudes to AI and ChatGPT are relatively uncommon in the academic literature and have predominantly been performed using samples of students, not academics, and/or professional staff. While several studies found that students recognize the potential benefits of AI in education (Baca, & Zhushi, 2025; Pataranutaporn, et al., 2021; van der Vorst & Jelicic, 2019), other studies see a plethora of risks caused by the disruptive use of AI and ChatGPT (Dwivedi, et al., 2023) or reluctance to rely on AI exam marking (Freeman, 2025).

Much of the recent survey research has been conducted among medical and healthcare students, given the levels of AI already being employed in these industries. Whole university attitudes are under-represented in the literature to date.

We regard AI as potentially helpful, but GenAI, in particular, as a double-edged sword insofar as the range of possible GenAI technologies possess the ability to provide opportunities for innovative teaching and improved student learning, at the same time as encouraging students to cheat their way through their degrees. We were unsure if our ideas and concerns were shared by other stakeholders at Western Sydney University. Thus, we surveyed staff and students to understand the existing knowledge, attitudes, and risks of students, academics and professional staff about AI in general and ChatGPT in particular. Our guiding research question was: *What are the current attitudes of all stakeholders to AI at Western?*

Method
Instrument

Given the dearth of findings on wide-ranging, institutional attitudes to AI in the research literature in early 2023, we could not find any satisfactory survey instruments upon which to base our own questionnaire. Thus, we asked ChatGPT 3.5 on four separate occasions (in May 2023) to produce four series of questions. Resultant output comprised some overlapping questions but also contained totally different items. The Early-Mid Career Research (EMCR) group was asked to vote for the best questions, and we asked executive staff of the university to add their own items if some issues were missing from our draft survey.

The final questionnaire consisted of four (4) demographic questions (gender, age, job role, and school), 39 Likert scales related to attitudes, and six (6) open-ended questions. Professional staff were asked to nominate their department or division, not school. The questionnaire was then placed on the Qualtrics platform and designed for ease of use for both computer and mobile phone. Demographic items were radio buttons. Likert scales were asked in two ways: 1. Five-point scales -Strongly Disagree to Strongly Agree; 2. as a slider 8-point scale with a zero. All scales went from less to more or Strongly Disagree to Strongly Agree in terms of directionality. Open-ended questions were text boxes.

Ethics and Recruitment

The project obtained Human Ethics approval (H15601) through the WSU subproject FLEXPoR process on 31 July, 2023. A Western webpage was created and advertised to all staff and students via email and Western forums. Posters (A3, A4 and A6) were also created and placed at university libraries and grounds at Parramatta South, Bankstown and Campbelltown campuses. Direct emails from the EMCR group to all staff of individual schools were distributed on numerous occasions. Student recruitment, however, was the real problem given the low participation rate of about 22 per cent of the sample, or 0.003 per cent of an estimated 48,000 enrolled students.

Demographics and Sampling Reliability

From August 1, 2023, to March 25, 2024, a total of 838 online surveys had been submitted to the Qualtrics platform. However, 126 of these were incomplete and had to be removed from the study. Answering more than one Likert scale question was the criterion for inclusion. Many students and a few staff opened the survey but failed to continue answering questions. Thus, the valid sample was 712 participants after removal of incomplete surveys.

Gender balance was relatively unequal with 39.7 per cent males, and 54.8 per cent females. Students, on average, were the youngest participants, academics and professional staff were the oldest. About 46 per cent of participants were aged between 35-54 years old. Students accounted for 30.8 per cent of the valid sample, academics for 36.5 per cent and professional staff for 31.5 per cent.

All schools within the university were represented in the sample. In rank order, Humanities and Communication Arts (12.5%), Health Sciences (6%), Business (6%) and Computing-Data-Maths (6%) were the most frequent school contributors making up over 30.5 per cent of the sample. The healthy rate of participation by students makes comparing students to staff attitudes a reliable process. Additionally, some research has suggested that 50-75 student responses may be sufficient for an unbiased sample irrespective of the size of the student population (Fosnacht, Sarraf, Howe & Peck, 2017).

A traditional 50 per cent valid return rate used for paper-based surveys is applicable when

researchers know how many invitations were distributed. This norm relies on the assumption that invitees make a conscious decision to either participate or not participate in the process. WSU students were not given a questionnaire to fill out in class, but instead, they were invited via email announcements and posters. However, this does not automatically mean that students made any conscious decisions. Apart from technical mishaps, we know that many students simply ignore institutional emails and routinely delete messages without reading them. We also believe that staff behave in similar ways. The relatively low student participation rate may, in fact, represent a reliable sample of a small but unknown population of students who had the requisite experience to demonstrate considered and defensible attitudes towards AI technologies.

Analytical Software

We used a variety of software to analyse the data. IBM SPSS 29 was used to analyse scale data using both descriptive and inferential procedures. Table results were imported to MS Excel for creation of graphs. The website, Voyant-Tools.org was used for content analysis of open-ended answers. We examined some of the open-ended answers by employing an advanced version of the computational linguistics application called VADER (Valence Aware Dictionary and Sentiment Reasoner) (see Roehrick, 2020) to determine the sentiment of each participant's answers and also calculated an average score for the aggregated three roles. VADER scores each word and sentence on a dictionary from -1 to +1. We used an extended web-based version of VADER that also produced scores for whole paragraphs, i.e. answers to open-ended questions. If we calculated a mean for all the paragraphs for each role, we could then compare sentiment between all students, academics and professional staff in the sample.

Results
Descriptives

The aggregated results of Likert scales are shown in three tables because 8-point slider scales (with a zero point) should be distinguished from 5-point, Strongly Disagree-Strongly Agree scales. The ChatGPT questions are also distinguished from the other items of a more generic meaning of AI. Slider scales were an innovation for the Qualtrics platform and were designed to capture mobile phone user engagement but proved to be slightly problematic in several

cases due to the scale being incorrectly displayed on the participant's mobile phone. These items had to be given missing data status for these participants.

Table 1. Slider Scale (zero to 7), Means, Medians and Modal Scores, in question order, with Selected Scores highlighted

Scale Item	Median	Mode	Mean	SD	N
Q7.1 Efficiency & productivity	5	7	5.09	1.58	640
Q7.2 Healthcare benefits	4	5	4.39	1.75	590
Q7.3 Decision-making	4	4	4.11	1.69	570
Q7.4 Automation of tedious tasks	6	7	5.70	1.59	644
Q7.5 Advancing research	5	5	4.45	1.72	601
Q7.6 Privacy-safety benefits	4	3	3.96	1.76	542
Q8.1 Job unemployment	5	7	4.81	1.82	639
Q8.2 Privacy-security risks	6	7	5.43	1.67	646
Q8.3 Ethical-bias risks	6	7	5.44	1.69	644
Q8.4 Surveillance risks	6	7	5.21	1.80	635

The slider scales examined opinions about possible benefits and risks of utilising AI technologies at WSU. The zero option was never used yielding a de-facto 7-point scale. From Table 1, the main costs of AI were perceived to be high in terms of Ethical-Bias Risks (M=5.44), Privacy/Security Risks (M=5.43), and Surveillance Risks (M=5.21). In terms of benefits, the Automation of Tedious Tasks was highly valued (M=5.70) as well as Efficiency/Productivity benefits (M=5.09). It should be noted that all mean, median and modal scores are above the mid-point of 3.5 for this set of questions meaning that on average, all users agreed with the assertion of risk or benefit, as opposed to disagreeing. These two sets of questions show that participants largely believe that AI possesses a range of advantageous benefits but even higher risks to society, given the modal and median scores.

Table 2. Traditional 5 pt. Likert Scale Items, Means, Medians and Modal Scores, in question order with Selected Scores highlighted

Survey Scale Item	Median	Mode	Mean	SD	N
Q5 FAMILIARITY	3	3	3.04	0.93	703
Q9 HOW COMFORTABLE	3	4	3.27	1.14	670

Survey Scale Item	Median	Mode	Mean	SD	N
Q10 TRUST AI	2	1	2.13	1.08	671
Q11 POTENTIAL IMPLICATIONS	4	4	3.96	1.06	671
Q12 AI IMPACT	4	4	3.46	0.98	670
Q14.1 AI will improve lives	4	4	3.73	0.95	648
Q14.2 AI will add to economic growth	4	4	3.58	0.92	647
Q14.3 AI can make better decisions	3	4	3.04	1.08	648
Q14.4 AI can enhance creativity	4	4	3.40	1.11	646
Q14.5 AI will worsen inequality	3	3	3.47	1.02	647
Q14.6 AI will lead to unforeseen risks	4	5	4.27	0.80	649
Q14.7 AI will need regulation	5	5	4.37	0.81	647
Q14.8 AI will lead to copyright problems	5	5	4.27	0.89	647
Q14.9 I am excited about AI	4	4	3.52	1.17	648
Q15.1 Govt bodies involvement	4	5	3.94	1.08	647
Q15.2 AI organisations involvement	4	4	3.54	1.16	645
Q15.3 Ethicists involvement	4	5	4.13	1.05	647
Q15.4 WSU involvement	4	3	3.56	1.12	642

Table 2 shows the aggregate means, medians, modes and standard deviations of all the 5-point scale items (minus ChatGPT scales). Answers to these items range from negative to positive opinions, unlike Table 1 scores. TRUST in AI (M=2.13), represents the lowest mean, falls below the scale mid-point of 3 and thus should viewed negatively for the majority of participants. Participants agreed with all the other item statements especially: the need for regulation (M=4.37), increased copyright issues (M=4.27), and unforeseen risks (M=4.27), and the need for ethicist involvement (M=4.13). Conspicuously, most benefits of AI are given a moderate score, such as economic growth (M=3.58), AI impact (M=3.46) and. Familiarity (M=3.04).

Table 3. Traditional 5-pt. Likert Scale Items, ChatGPT-specific, Means, Medians and Modal Scores, in question order with Selected Scores highlighted

Survey Scale Item	Median	Mode	Mean	SD	N
Q16.1 ChatGPT can enhance research	4.00	4	3.58	1.14	514
Q16.2 ChatGPT is valuable for new ideas	3.00	3	3.21	1.24	511

Survey Scale Item	Median	Mode	Mean	SD	N
Q16.3 ChatGPT provides accurate & reliable info	2.00	2	2.48	1.08	514
Q16.4 ChatGPT helps me with new ideas	4.00	4	3.49	1.11	513
Q16.5 ChatGPT can answer discipline specific qns	3.00	3	2.98	1.15	504
Q16.6 I prefer ChatGPT over traditional methods	2.00	1	2.26	1.14	504
Q16.7 ChatGPT has limitations	5.00	5	4.50	0.85	505
Q16.8 ChatGPT has the potential to reduce time	4.00	4	3.44	1.08	504
Q16.9 ChatGPT has reliable & accurate responses	2.00	2	2.35	1.07	508
Q16.10 ChatGPT can effectively summarise	4.00	4	3.48	1.09	507
Q16.11 Overall ChatGPT is valuable tool for my work	4.00	4	3.49	1.12	511

Almost all participants strongly agreed that ChatGPT has limitations (M=4.50). Reliability and accuracy (M=2.35) is believed to be the most problematic feature of ChatGPT (M=2.35). However, ChatGPT's overall reputation, Q16.11 as a valuable tool is only moderately high (M=3.49).

Anecdotally, in terms of opinions about ChatGPT overall, students and staff appear equivocal towards the tool - they either love it or hate it. This could be because of experiential problems with existing AI tools or this may be because of the spectre of unending automation or the threat of existential problems. The media-fuelled association of AI with the commercial interests of Microsoft, Google, and Meta does not make AI an altruistic gift to humanity. Mega corporations spending millions of dollars on AI leads to a proportionate degree of suspicion. The arrival of ChatGPT has created major academic integrity issues which devalue the intrinsic appraisal of higher education degrees - a student chatting with a sophisticated AI chatbot can possibly obtain a degree, now.

Profiling Academic, Professional and Student Attitudes

After examining the survey as a whole and aggregating all participants, the next stage in the analysis was to differentiate academic staff, professional staff and students from each other to identify attitudinal differences.

It was hypothesised that students and professional staff should possess their own AI opinions given their respective educational and professional identities. Academics should be different again and perhaps exhibit a more critical range of attitudes related to both teaching and research responsibilities. A one-way ANOVA was conducted on the scale items as dependent variables, with Role as the main predictor variable. The robust Welch procedure in SPSS 29 was selected to ensure homogeneity of variance would not need to be assumed.

Table 4. Significant Welch ANOVA tests with Role as predictor at 0.05 alpha level

Significant Items listed	Statistic	Df1	Df2	Sig
Healthcare and medical diagnosis	6.06	2	368.09	.003
Privacy-security risks	5.30	2	406.52	.005
Ethical-bias risks	14.97	2	400.85	<.001
Surveillance opportunities	6.99	2	399.67	.001
POTENTIAL IMPLICATIONS	16.37	2	426.66	<.001
AI IMPACT	6.69	2	438.17	.001
AI will worsen inequality	22.16	2	414.25	<.001
AI will lead to unforeseen risks	16.19	2	399.30	<.001
AI will need regulation	5.55	2	409.65	.004
AI will lead to copyright problems	8.17	2	407.31	<.001
Govt bodies involvement	29.63	2	399.67	<.001
Ethicists involvement	16.50	2	397.52	<.001
WSU involvement	22.60	2	405.77	<.001
ChatGPT is valuable for new ideas	12.34	2	325.71	<.001
ChatGPT provides accurate & reliable info	12.84	2	324.48	<.001
ChatGPT helps me with new ideas	13.42	2	334.09	<.001
I prefer ChatGPT over traditional methods	15.89	2	314.46	<.001
ChatGPT has limitations	6.93	2	309.33	.001
ChatGPT has the potential to reduce time	3.03	2	322.87	.049
ChatGPT has reliable and accurate responses	17.26	2	324.49	<.001
ChatGPT can effectively summarise	3.28	2	324.51	.039

The demographic, *Role* of the WSU participants differentiated nearly half of the attitudinal

scales of the survey. Thirteen out of 27 general attitudes, and 8 out of 11 of the ChatGPT-specific items were significantly affected by Role. Sixteen of the 21 significant items were highly affected at the .001 or less, significance level. Clearly, academics, professional staff and students have markedly different opinions about AI probably due to their role-based needs, duties and responsibilities at the university.

Apart from complex post-hoc tests a simple way to examine items was to cross-tabulate student, academic and professional staff mean scores across the Likert scale points. To make the comparisons fair, we used percentages of scale scores within each institutional role, not raw frequency counts that the SPSS software produces by default. As can be seen from Figure 1, the main difference occurs with scale point 7 where about 45% of academics express significantly more concern (sig <.001) than either professional staff (38%) or students (26%) for the possible ethical and bias issues found in AI systems. Similar reactions to ChatGPT can be also seen in Figure 2
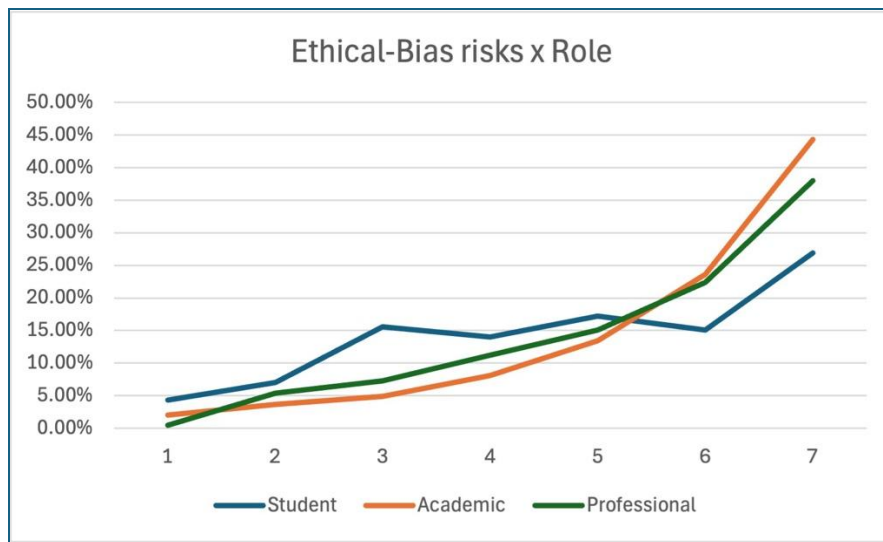
Figure 1. Line Graph of Crosstab of "Ethical-bias risk" x Role, using Percentage within Role. (The x-axis represents 7 used slider scale points, zero was never selected.)

In Figure 2, the obvious main difference occurs at scale point 1. Strongly Disagree. Again academics (37.5%) are much higher (sig <.001) than students (19%) and professional staff (19.5%) in Strongly Disagreeing with the assertion of AI reliability and accuracy. This pattern of response where academics are most critical, and students are least critical of the potential of AI, with professional staff positioning themselves between the two roles, was

seen throughout most of the significant items. The reason for these differences is most likely due to the inherent nature of university roles.

.
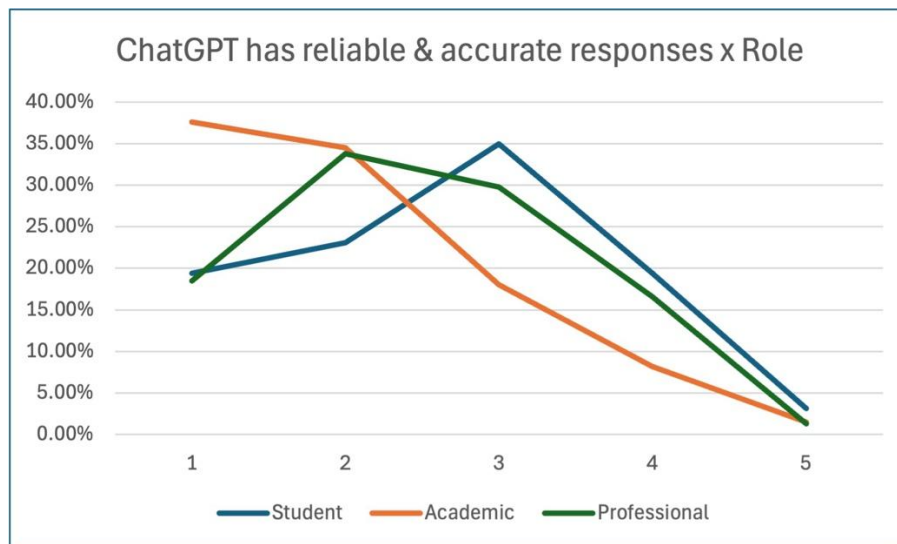
Figure 2. Line Graph of Crosstab of "ChatGPT has Reliable and Accurate Responses" x Role, using Percentages within Role. (The x-axis represents 5 Likert scale points.)

Academics are a specialised group of participants and usually subscribe to the highest standards of critical thought, and insightful understanding of society. Academic research is regarded as the most rigorous form of research and its proponents are respected as being objective, ethical, less self-serving and less prone to mistakes. Academics are also critical users of technology and less susceptible to popular opinion or media sensationalism. A good example of this critical but somewhat cautious approach to technology is the fact that it took many Australian universities five (5) years to develop their own websites at the end of the last century. Many private enterprises possessed custom websites by 1995 in Australia.

Students are much more pragmatic members of the university in that they use technology and presumably, AI for important tasks such as assignments, and also for their concurrent employment. Many students are currently employed in jobs that ask them to utilise AI to perform their daily jobs. Anecdotally, many jobs have already changed because AI has allowed several job roles to be amalgamated into a single position. Organisations are able to leverage AI to reduce their staff overhead costs and increase profits. While student employees may protest their extra workloads, such students are certainly appreciative they can successfully navigate this changed employment landscape using AI tools. Overall, students

apply the same compromises to using AI within the university.

The term "Professional staff" refers to a range of positions and skills that support the running of the university's teaching and research functions. These staff were difficult to categorise because they included a large range of roles including administrative positions, computer support, human relations, finance, PR, training, learning support, and librarians. Professional staff are currently not subject to the HR cutbacks that AI tools have permitted the private sector to endorse. Working for a university is still a prestigious job, with many staff possessing qualifications above their pay grade. Many professional staff who work for universities prefer to stay in their jobs because they appreciate the conditions and the culture of academia. They are not academics, but many advocate academic customs of freedom of thought, diversity of opinion, honesty, and intellectual rigour. Professional staff attitudes tend to not be as critical as those of academics but are still suspicious of the grandiose promises of AI technology.

Differences between WSU Schools on the AI Survey Scale Items

It was possible to test whether the 16 categories of schools, institutes and professional divisions significantly differed in their attitudes to each of the survey items. We conducted an ANOVA on the scale data using schools, divisions and institutes as predictors, employing robust Welch tests, and Games-Howell post-hoc comparisons. Inspection of the resultant SPSS output showed non-systematic, random differences between WSU schools that cannot easily be explained. However, one major difference stands out – the School of Computer Data and Mathematical Science and to a lesser extent the School of Engineering and Built Environment, were significantly higher (p<.001) to all other schools in terms of their familiarity with AI. This fact influences their attitudes and opinions about many of the other survey items because these two schools appear to currently have the most experience with AI technologies.

Familiarity (or Experience) with AI – Demographic or Attitude?

While participants of two WSU schools appeared to have an advantage over the rest of the university in terms of familiarity with AI, this variable overall was unremarkably normally distributed as can be seen in Figure 3.
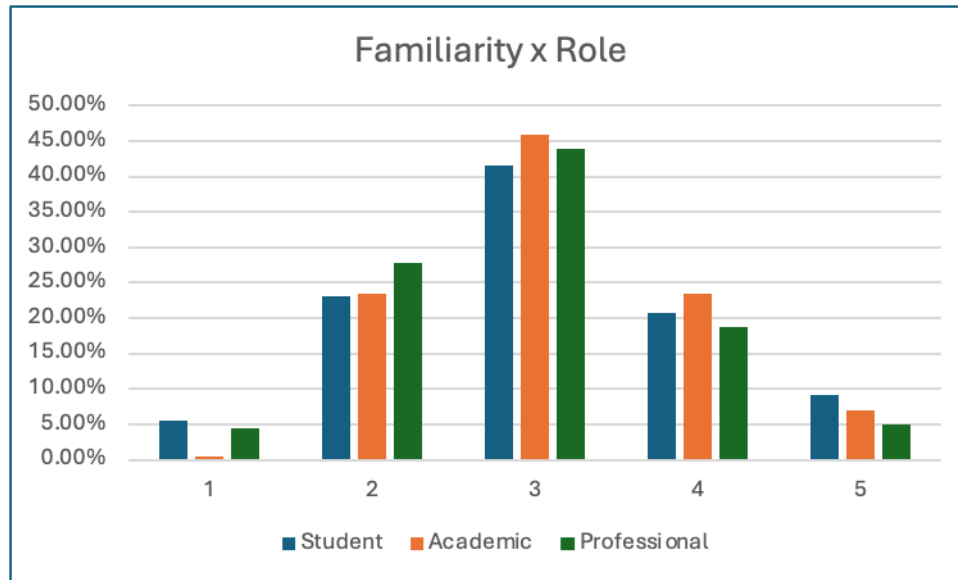
Figure 3. Histogram of Crosstab for Familiarity x Role, using Percentages within Role. (The x-axis represents the 5 Likert scale points, Not familiar to Extremely familiar.)

Arguably having moderate Familiarity (scale point 3) can be interpreted as having hands-on experience or understanding the term from mainstream or social media. The question is ambiguous and may or may not be useful for the analysis. Notably, academics have a small head-start with over 45% scoring 3, followed by professional staff (44%), and then students (41%). Different scale points show slightly different ratios of percentages for each role. One problem of scale point 3 is that in some Likert scale surveys this mid-point item means No Opinion or Not Applicable.

We hypothesised that Familiarity would be a valuable precursor to possessing informed attitudes about AI, and thus we conducted a one-way ANOVA with Familiarity as the main predictor variable, and the other scale items as dependent variables.

Table 5. Table of Sig. Welch ANOVA Tests with Familiarity as Predictor at 0.05 Alpha Level

Significant items listed	Statistic	df1	df2	Sig.
Efficiency & productivity	11.93	4	101.69	<.001
Healthcare and medical diagnosis	6.02	4	84.37	<.001
Decision-making	15.41	4	85.05	<.001
Advancing research	7.58	4	91.42	<.001

Significant items listed	Statistic	df1	df2	Sig.
Job unemployment & displacement	2.52	4	99.20	0.046
Privacy-security risks	2.53	4	98.62	0.045
HOW COMFORTABLE	17.85	4	105.78	<.001
TRUST AI	15.40	4	123.15	<.001
AI will improve lives	8.99	4	102.20	<.001
AI will add to economic growth	4.81	4	104.01	0.001
AI can make better decisions	3.20	4	102.459	0.016
AI can enhance creativity	4.53	4	103.96	0.002
AI will need regulation	2.81	4	99.41	0.029
I am excited about AI	10.94	4	101.13	<.001
ChatGPT has limitations	2.64	4	40.84	0.047

From Table 5 it can be seen that *Familiarity* does differentiate participants' attitudes on 15 of the 39 scale items. Nine of these 15 items were highly significant at the .001 sig level or less. Histograms are the easiest charts to understand patterns for each 7-point scale.
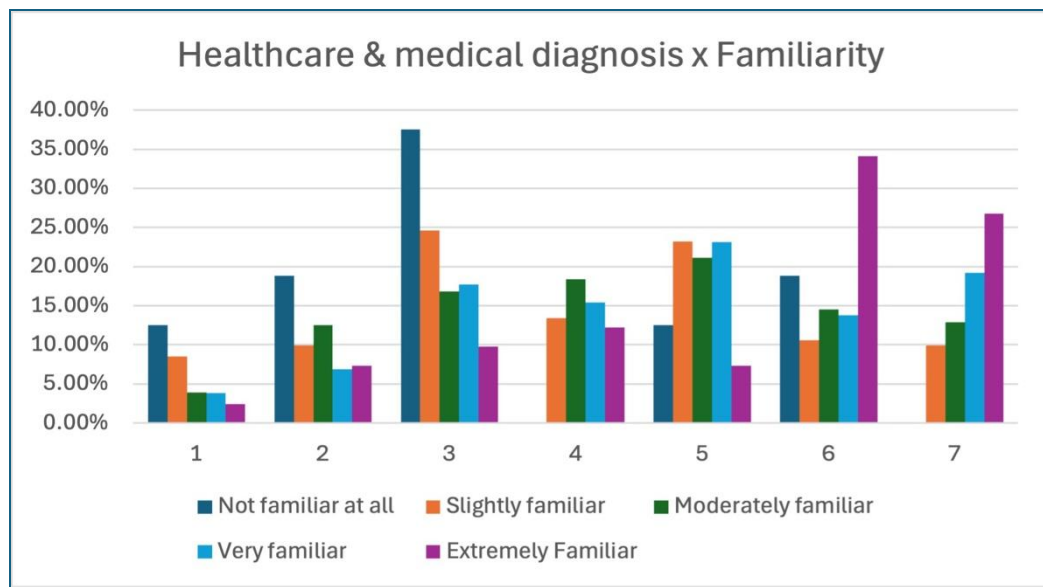
Figure 4. Histogram of Crosstab Q7.1 Healthcare & Medical Diagnosis x Familiarity Scale, $p<.001$

The main differences seen in Figure 4 is the association of high percentages of zero *Familiarity* (dark blue bars) for scale items 1, 2 and 3. (68% combined) and high percentages of extremely Familiar (purple bars) for scale items 6 and 7 (60%+ combined). In other words,

the lower the experience the participant has with AI, the lower the assessed potential benefits of healthcare and medical diagnosis using AI. Conversely, the more experience with AI then the higher the participant assessment of healthcare and medical diagnosis with AI. The following graph typifies the effect of *Familiarity* on the 5-point scale, "Trust in AI to make important life decisions" and many of the other attitudes of the survey.
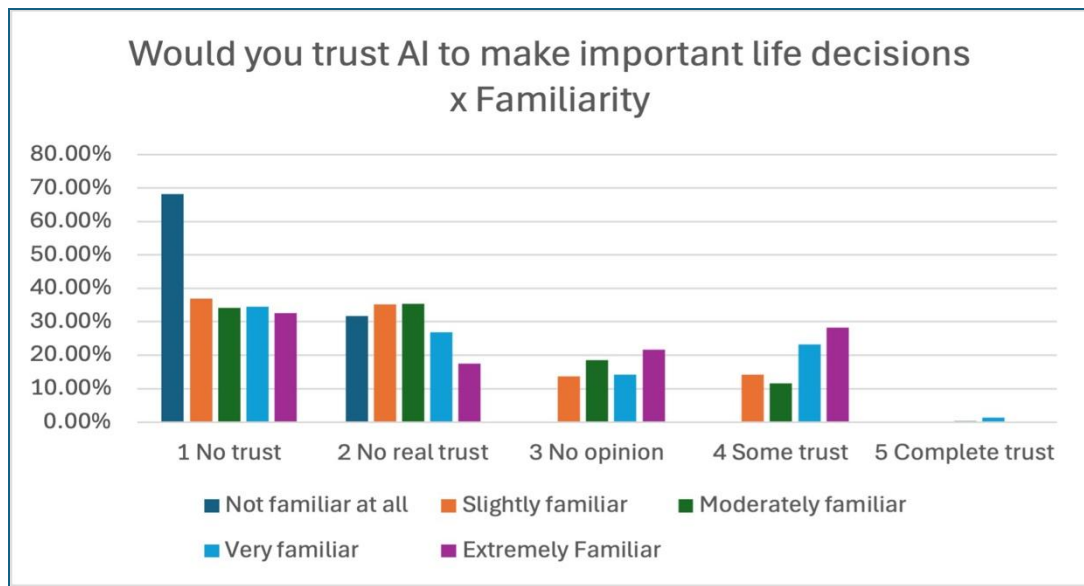
Figure 5. Histogram of Q10 Trust in AI x Familiarity with AI scale, p<.001

Figure 5 depicts a compelling finding. The dark blue line, scale point 1 (No trust) for *Q10 Trust in AI to make important decisions that could impact your life* was associated with *Not Familiar with AI at all* by 69% of this group. The remaining choice is for this group is scale point 2 (No real trust, 31%). These two levels of trust (100%) in AI are the only choices that this group offers. This finding strongly shows that lack of Familiarity with AI is closely associated with serious mistrust of AI technologies. The purple bar (Extremely Familiar) participants also have serious trust misgivings with only 29% saying they have any sort of trust in AI (scale point 4).

Close inspection of the histograms for the other significant survey items shows a very similar pattern of results – less experience with AI leads to a fear of the unknown and thus greater pessimism, whereas greater experience with AI usually leads to greater optimism (but not always). It is apparent that AI tools are currently held in awe by the media and viewed as mysterious inventions, and this is reinforced by media statements that assert that scientists themselves do not quite understand how AI works or how it will evolve (see Bailey & *The*

Conversation US, 2023). University staff and students are wary of such comments and slow to praise AI tools in 2023.

This finding that experience with AI results in higher optimism about the technology but casts some doubt whether any random sample of participants can reliably express their attitudes to new technologies such as AI. Are many of the survey scales simply measuring familiarity/experience with AI tools, and not real opinions?

Open-ended Questions

There were six open-ended questions:

- Q6 Please briefly define AI in your own words.
- Q13 Have you personally experienced any positive or negative interactions with AI?
- Q17 What measures should be employed to mitigate potential risks and ensure human rights?
- Q18 How would you respond if you received a Microsoft warning that you should not send email outside of business hours?
- Q19 How would you respond if you received an employer warning that your screen time was down by 25% this week?
- Q20 Do you have any additional comments or insights you would like to share?

Q6 Please briefly define AI in your own words

Table 6 was created by uploading open-ended answers for students, academics and professional staff to the online content analysis site, Voyant-tools.org. The research website allows for uploads of text files and .csv files (from Excel) and provides numerous screen-based panels that analyse text automatically. Using the Terms panel and removing small "stop" words allowed for a direct comparison between WSU roles and the 12 most popular concepts for Q6. Apart from the ranked order of words, the concepts used are very similar with the top three (3) concepts being exactly the same, given "information" is a synonym of "data". It can be concluded that students, academics and professional staff have similar understandings of the term, "artificial intelligence". Use of the jargon word "algorithms" was more highly used by academics, but this slight difference was not substantive.

Table 6. Voyant-Tools Content Analysis of Q6 Briefly define AI, showing top 12 concepts and frequencies of 3 roles

Rank	Students (N=84)	Academics (N=206)	Professionals (N=168)
1	human 30	human 84	human 44
2	computer 24	computer 64	computer 26
3	data 24	information 56	data 25
4	tasks 23	learning 53	tasks 23
5	information 23	machine 41	information 21
6	learning 19	tasks 30	learning 20
7	machines 17	technology 26	machine 20
8	perform 13	algorithms 25	technology 19
9	technology 13	software 24	language 16
10	language 12	questions 22	text 15
11	text 10	language 22	algorithms 14
12	algorithms 8	text 20	models 14

Q13 Have you personally experienced any positive or negative interactions with AI?

For Q13, 129 students, 181 academics and 136 professional staff produced valid responses to this open-ended question. We deleted simplistic answers such as "No". For the vast majority of student participants, ChatGPT was the main AI technology discussed. Text analysis of the answers showed that students' main positive experience with AI was related to employment where ChatGPT helped them to perform their external jobs more efficiently. The most common negative complaint was being falsely awarded a Turnitin plagiarism score when in fact students stated they did not cheat, i.e., false positives from WSU software.

From the University's point of view, false negatives from Turnitin software are the more worrying problem because this is tantamount to cheating that is never discovered, never known about and never discussed in surveys such as this one. If Turnitin cannot reveal paraphrases of generative AI output, then plagiarism will become increasingly commonplace affecting high percentages of both struggling students and high-achieving students. The invisible use of ChatGPT is probably already a part of the educational process of a great many students.

Table 7. Voyant-Tools Content Analysis of Q13 Have you personally experienced positive or negative interactions with AI technologies, showing top 12 concepts and frequencies

Rank	Students (N=129)	Academics (N=181)	Professionals (N=136)
1	use, using 49	students 91	positive 31
2	work 30	information 24	ChatGPT 27
3	positive 24	negative 23	work 23
4	ideas 22	writing 21	chat 18
5	help, helped 25	learning 21	writing 16
6	time 16	positive 20	students 15
7	like 10	assessments 18	useful 15
8	research 10	research16	negative 14
9	students 9	questions 12	information 12
10	writing 9	teaching 9	things 10
11	ChatGPT 8	academic 8	task 9
12	information 8	time 6	data 9

From Table 7, academics (N=181) seem to be nearly equally "positive" (20 instances) and "negative" (23 instances). But many academics complained about the difficulty of marking assignments with student welfare being a frequent concern (91 instances). One exemplary academic quote was:

> *Marking has become more difficult and setting assessments that avoid cheating has become really difficult.*

Professional staff used the concept, "positive" more than twice the number of times as the concept, "negative" indicating their enthusiasm for AI. A positive quote from a professional staff member was:

> *Generative AI is an absolute game changer for the computer work we do – especially programming in Python, R or any other language.*

The VADER total compound means for WSU roles, as seen in Table 8 appear as low (but positive) scores because positive experience scores are cancelled by negative experience scores for each participant, thus producing scores approaching zero in the aggregated sample. However, the order of enthusiasm for AI tools can be ranked as 1. professional staff first, then 2. students, with 3. academics least impressed with their previous experiences.

Table 8. VADER Mean Scores, plus Positive and Negative Exemplars for 3 roles for Q13

Role	VADER score	Participant Exemplars
Student	**M=0.182**	
negative	-0.552	*(AI) can be misleading and repetitive*
positive	0.625	*(AI) provides me with great ideas for my assessments*
Academic	**M=0.079**	
negative	-0.415	*An alarmingly large number of students are using it (AI) to cheat*
positive	0.608	*Improved teaching materials*
Professional	**M=0.215**	
negative	-0.433	*Cheating by students, no innovation by users*
positive	0.413	*Very useful third party/critical eye for curriculum development, lesson planning, combing through large data sets and documents*

Table 8 also reveals the range of emotional responses for students, academics and professional staff by listing some of their most negative and most positive answers according to individual extreme VADER scores. The positive mentions of "great ideas for assessments", "improved teaching materials", and "better curriculum development, and enhanced sorting of data sets" represent the most positive personal sentiments towards AI at Western. However, related negative sentiments from other participants are also displayed for comparison purposes.

Q17 What measures should be employed to mitigate potential risks and ensure human rights?

From Table 9 it can be seen that the answers were very similar, with many of the same words and concepts being used across the three roles. Most participants who answered this question believed that regulation was definitely needed to protect risks of AI and ensure human rights, but few had any confident ideas about how this could be achieved. The same debate can be seen in the mainstream media as we witness increasing numbers of "wish lists" but few new laws being created so far (see Levy, 2023, 26 May).

Table 9. Voyant-Tools Content Analysis of Q17 What measures should be employed to mitigate risk and ensure human rights, showing top 12 concepts of 3 roles

Rank	Students (N=114)	Academics (N=134)	Professionals (N=122)
1	data 23	regulation 36	human 28
2	work 21	human 27	data 26
3	ensure 18	need 25	information 23
4	human 12	data 24	people 20
5	information 10	people 22	rights 18
6	oversight 7	information 21	oversight 17
7	people 7	development 20	needs 16
8	public 6	government 17	tool 16
9	research 6	oversight 15	access 14
10	academic 5	rights 15	research 14
11	government 5	research 14	risks 12
12	education 4	decisions 14	education 10

Q18 How would you respond if you received a Microsoft Outlook warning that you should not send email outside of business hours?

In 2023, this question arose from some of the group questioning if all staff and students had seen the current Microsoft warnings about out-of-office email. The warning encapsulates tension surrounding desirable worker flexibility versus expected workplace obligations. The Microsoft warning was not really an AI tool but an automated message from Microsoft asking users to consider the consequences of sending emails at a time that is usually deemed to be non-work time (often before 9 am, and after 5 pm). We believe that this topic is important except for "Big Brother" Microsoft appearing to comment on our already accepted work practices. This may be over-stepping the mark in an associated but somewhat petty way that ChatGPT has become a massive disruptive tool.

Previous research had concluded that 79 per cent of full-time workers had not been paid for working outside business hours, which in fact, amounted to overtime. Findings calculated that some employers had 'stolen' more than 280 hours from each employee, equating this with seven standard working weeks per year (MacDonald, 2023). As of August 2025, a new

Australian Federal government bill has given all Australian employees the right to not work (including responding to emails and phone calls) outside of normal working hours (Ziffer, Aug 26, 2025).

However, working within an Australian university is different to working in a government office. Flexibility is one of the blessings of working in a university. Ninety-five students, 180 academics and 159 professional staff answered this question in quite similar ways. A majority of participants had not seen the warning, but many said that they would be outraged, many said that they would ignore the warning if they saw it. Some of the more considered comments are the following:

- *I am not sure that I understand the question? is it a warning that I should consider work life balance? Is it part of business policy? Save the email as a draft and send it first thing in the morning. Discuss it with my supervisor the next day. (Student)*
- *I would consider it inappropriate. I work flexible hours and have a by-line in my emails indicating this is the case, and that I do not expect a response outside anyone else's business hours. I don't need a computer to remind me to be professional and considerate. (Academic)*
- *This would be an improvement on quality of life in maintaining a work-life balance. France has made it illegal to send emails outside work hours; why not Australia? (Professional)*

Q19 How would you respond if you received an employer warning that your screen-time was down by 25% this week?

This question arose from reading Microsoft Viva Insights emails and information that is aggregated from weekly email data. Microsoft explains:

> *The insights for individuals that this app presents are completely personal and private. Personal insights in the app are for your eyes only; neither your manager nor the system administration can see your insights. (https://support.microsoft.com)*

The "insights" are analytics that Microsoft can distil from the data that all Outlook users unknowingly contribute. We think that the insights sound like time and motion studies from the early 20th century when Frederick Taylor's scientific management theory (Taylor, 1911)

revolutionised organisational communication at the time. Microsoft is evidently asking individuals to consider becoming managers of their own labour.

Taylor's four principles were: 1. Replace habit with efficiency research; 2. Do not assign workers to random jobs but match them in terms of capability and motivation; 3. Monitor workers' performance and train them to work at maximum efficiency; 4. Separate managerial functions (planning and training) from task functions (physical labour) (Taylor, 1911). Taylorism has been criticised for its rigidity, its refusal to encourage teamwork, its creation of artificial specialisations, and its presumption that mental work needs to be separated from physical work. Taylorism has essentially been rejected by modern organisational practice, but it is alive and well in Microsoft's "insights" and still found in many organisations today.

Ninety-two students, 180 academics and 158 professional staff answered this question with similar vehemence and outrage. The words, "surveillance", and "privacy abuse" were very commonplace and the phrase, "I think I would find another job" was used many times. Surprisingly, not many Outlook users had seen these Viva emails. Only three (3) participants explicitly mentioned the Viva digest emails. However, 74 per cent (N=303) of participants replied with a negative response from a modest, "ignore the email" to use of one of several expletive swearwords, while 17 per cent (N=71) of participants took the email at face value and responded positively or defensively. A small percentage, 7 per cent, (N=36) were perplexed and needed clarification or stated that the email was not relevant to them.

Q20 Do you have any additional comments or insights you would like to share regarding the intersection of AI, the university or higher education? e.g. risks, biases, accuracy?

For this last question, 88 students, 126 academics and 98 professional staff produced valid responses. We deleted single-word answers such as, "No" and were surprised at the range of opinions and depth of many of the answers. We do not believe we have fully captured this question's results because of space limitations of this chapter. From Table 10 the number one theme is "students" for all three roles at WSU. However, close inspection of the comments reveals that student use of "student" is about personal issues – being helped or hindered by AI (and other) technologies. Academic use of "student" shows concern for student welfare, and educational development. The professional staff discuss students as a generic group of clients and present societal issues and solutions.

Table 10. Voyant-Tools Content Analysis of Q20 Do you have any other additional comments you would like to share…, showing top 12 concepts and frequencies of 3 roles

Rank	Students (N=88)	Academics (N=126)	Professionals (N=98)
1	students, student 41	students, student 70	students, student 59
2	university, universities 27	need, needs 49	need, needs 37
3	work 18	work 29	tool, tools 31
4	people 16	university 23	research 21
5	academic 12	technology 20	staff 21
6	tool 12	human 18	university 19
7	research 10	education 17	work 19
8	ChatGPT 7	research 17	people 17
9	education 7	information 16	ChatGPT 16
10	human 7	learning 15	information 14
11	world 7	potential15	learning14
12	future 6	academic 14	skills 13

Students mention the university at rank #2, not rank 4 or 6, as for staff. AI can be seen as affecting the university for students more so than staff. A good example of this trend are the comments:

- *The <u>university</u> should embrace the use of AI to advance learning. (Student)*
- *I don't think the <u>university</u> is enabling students to critically use AI. I don't have a problem with its use per se. But I do think that students need to remain critical. (Student)*

Some notable warnings for the university include:

- *False positives of Turnitin. When it happens, it is devastating and the process of resolving this is traumatic. (Student)*
- *The fact that we as students are being encouraged to use AI technologies to prove that we are not using AI to write essays is becoming something of a joke. What's more, it places pressure on the students who have never and would never use such technology to prove they are being honest. (Student)*

ChatGPT was ranked highly by students and by professional staff, but not by academics who

took a more wide-ranging perspective. This one comment summarises many of the academic comments:

- *After just one semester of ChatGPT in the hands of students, and enhanced AI functionality for software like Grammarly, I've seen a dramatic decrease in genuine engagement with course materials, a huge upswing in papers failing for AI use that has sent them in completely the wrong direction for the task, and a general lack of understanding on behalf of students of what AI is, how it works, and what the ethical and academic integrity boundaries are for its use in academic study. (Academic)*

This comment shows the genuine benefits and downsides to student use of ChatGPT but fails to record the large workload spike that detecting the use of ChatGPT via Turnitin algorithms, has created for academics. This is especially true for staff whose students currently submit essays and reports, rather than sit invigilated exams. Professional staff tended to make more societal comments, as opposed to the individualised students' and teaching-related academics' comments. Here is one:

- *Education systems will need to evolve with changing technology or be left behind. The nature of society is changing rapidly ... Universities should form an AI department of specialists who can advise all staff and students on all regulatory changes; all AI organisations and their emerging tech, ethical applications and implications for all aspects of higher education life and work; permitted/recommended and prohibited/not recommended AI tools; incorporating AI within assessments, course design and collaboration; develop AI-related policies; offer solutions to disaffected individuals and groups, etc. (Professional staff)*

The above comments represent a small portion of the many additional thoughtful and well-considered contributions by students, academics and professional staff. This last question could not be easily summarised due to lack of space and could be the basis of a new extended qualitative article on university attitudes to AI.

Discussion and Conclusions

This survey of our university was one of the first to provide insight into the attitudes, risks, and opportunities associated with the use of AI within an entire higher educational institution.

The survey revealed a diverse range of existing attitudes towards AI at Western. While some participants expressed enthusiasm about the potential of AI to enhance learning experiences and streamline administrative processes, others expressed concerns about generative AI (mainly ChatGPT), ethical implications, and the potential for AI to exacerbate existing inequalities in education. The Microsoft warnings (not strictly AI tools) about office hours and screen times were highly disliked by most participants. We believe that this range of perceptions indicates the need for comprehensive debate and transparent policies around the ethical and equitable use of AI and computers at Western. We also question why Microsoft should be allowed to contact staff about its email analytics.

The survey has revealed risks such as the lack of staff experience, basic inequities between school experiences with AI, and deeply shared concerns about data privacy, copyright, AI reliability, trust, safety and surveillance. On the other hand, efficient current workplace use of AI, opportunities for personalised learning, predictive analytics for student success, and more efficient learning, teaching and administrative processes were also recognised. It is evident that while AI presents exciting possibilities, careful consideration of its implementation is essential to mitigate potential risks and biases.

The survey findings have significant implications for policy and practice at Western. Lack of familiarity with AI software and tools, other than generative AI tools indicates that the university, first and foremost, needs to educate its staff and students on the range and variety of available AI software for learning, teaching and administration. The university should then prioritise the development of clear guidelines for the ethical use of AI, invest in staff development programs to ensure they are equipped to leverage AI tools effectively, and establish robust data governance frameworks to safeguard student privacy and protection. Additionally, collaborative efforts between academia, industry, and regulatory bodies are fundamental to addressing the broader societal implications of AI in higher education.

We need to be mindful of the current reliance on stop-gap tools such as Turnitin that: 1. markedly increase academic workloads, 2. produce false accusations of innocent students, 3. create complacency that detection of AI cheating is being appropriately managed, and 4. allow high levels of cheating to pass unnoticed. We know that Turnitin detection can be easily defeated because we have accomplished this feat ourselves using free internet services. We see several areas for future research. For instance, longitudinal studies tracking the

impact of AI integration on student learning outcomes and staff practices could provide valuable results. In addition, research focusing on the development of inclusive AI technologies that address diverse student needs and promote equity in higher education is warranted. Higher education does not have an exemplary record of dealing with disability, gender discrimination, racism, sexual harassment or ESL issues (Wolbring & Lillywhite, 2021; Bensimon & Malcom, 2023).

It was a revelation to us that experience or familiarity with AI acted as an extraneous variable significantly affecting many of the other attitudes in the survey. If we had previously known about this finding, then we could have more clearly defined the concept for participants or created other items to control for variation in participant understanding. Participant experience with any form of new technology probably confounds attitudes towards that technology. Our ignorance of the importance of experience is a limitation of the entire survey.

However, the survey findings have allowed us to begin to understand the complex nature of the landscape of AI in higher education. We think it is imperative to critically approach the adoption of AI at Western, ensuring that it aligns with agreed-upon ethical principles and contributes to positive educational outcomes for all students, and all staff. The problem for all of us is that the ethics of ChatGPT (and its clones) in particular, and AI in general, are still being developed by universities, think tanks, governments and experts around the world.

Acknowledgement

References

Alam, A. (2023). Harnessing the power of AI to create intelligent tutoring systems for enhanced classroom experience and improved learning outcomes. In *Intelligent communication technologies and virtual mobile networks* (pp. 571-591). Springer Nature Singapore.

Almalawi, A., Soh, B., Li, A., & Samra, H. (2024). Predictive models for educational purposes: A systematic review. *Big Data and Cognitive Computing*, *8*(12), 187.

Alston, E. (2023). What are AI hallucinations and how do you prevent them? Online document: https://zapier.com/blog/ai-hallucinations/

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 267-270.

Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education, 24*, 387-426.

Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning, 20*(2), 13-29.

Baca, G., & Zhushi, G. (2025). Assessing attitudes and impact of AI integration in higher education. *Higher Education, Skills and Work-Based Learning*, *15*(2), 369-383.

Bailey, M & The Conversation US (2023, October 3). How can we trust AI if we don't know how it works, *Scientific American*. https://scientificamerican/article/how-can-we-trust-ai-if-we-dont-know-how-it-works/

Bensimon, E. M., & Malcom, L. (Eds.). (2023). *Confronting equity issues on campus: Implementing the equity scorecard in theory and practice*. Taylor & Francis.

Browne, R. (2023, April 4). Italy became the first Western country to ban ChatGPT. Here's what other countries are doing. https://www.cnbc.com/2023/04/04/italy-has-banned-chatgpt-heres-what-other-countries-are-doing.html

Campbell, J., DeBlois, P., & Oblinger, D. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review, 42*(4), 40-57.

Cantabella, M., Martínez-España, R., Ayuso, B., Yáñez, J. A., & Muñoz, A. (2019). Analysis of student behavior in learning management systems through a big data framework. *Future Generation Computer Systems, 90*, 262-272. https://doi.org/10.1016/j.future.2018.08.003

Chen, R. J., Wang, J. J., Williamson, D. F., Chen, T. Y., Lipkova, J., Lu, M. Y., ... & Mahmood, F. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature Biomedical Engineering, 7*(6), 719-742. https://doi.org/10.1038/s41551-023-01056-8

Daniel, B. (2020). Artificial intelligence and the 'good society': The US, EU, and UK

approach. *Information, Communication & Society, 23*(2), 279-297. DOI:10.1007/s11948-017-9901-7

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642\

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., & Chazerand, P. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

Fosnacht, K., Sarraf, S., Howe, E., & Peck, L. K. (2017). How important are high response rates for college surveys? *Review of Higher Education, 40*(2), 245-265. Retrieved from Pro-Quest, 18 January 2024.

Freeman, J. (2025). Student generative ai survey 2025. *Higher Education Policy Institute: London, UK*.

Gaftandzhieva, S., Hussain, S., Hilcenko, S., Doneva, R., & Boykova, K. (2023). Data-driven decision making in higher education institutions: State-of-play. *International Journal of Advanced Computer Science and Applications*, *14*(6).

Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., ... & Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems*, *4*, 19-23.

Gligorea, I., Cioca, M., Oancea, R., Gorski, A. T., Gorski, H., & Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. *Education Sciences*, *13*(12), 1216.

Google Gemini within Chrome ver 138.0.7204.184 (2025, 9 Aug). Prompt: "Have governments in 2025 accepted the potential economic benefits of AI but have not solved ethical, security and regulatory issues?"

Hamadneh, N. N., Atawneh, S., Khan, W. A., Almejalli, K. A., & Alhomoud, A. (2022). Using artificial intelligence to predict students' academic performance in blended learning. *Sustainability, 14*(18), 11642. MDPI AG. http://dx.doi.org/10.3390/su141811642

Huang, M., Huang, H., Chen, I., Chen, K., & Wang, A. (2021). Artificial Intelligence Aided Course Scheduling System. In *Journal of Physics: Conference Series*, *1792*(1), p.

012063). IOP Publishing.

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2015). *NMC/CoSN Horizon Report: 2015 Higher Education Edition*. The New Media Consortium.

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International journal of Educational Technology in Higher education*, *20*(1), 56.

Levy, S. (2023, May 26). Everyone wants to regulate AI. No one can agree how. *Wired.* https://www.wired.com/story/plaintext-everyone-wants-to-regulate-ai/

MacDonald, F. (2023). Short changed: Unsatisfactory working hours and unpaid overtime. *The Australia Institute.* https://futurework.org.au/wp-content/uploads/sites/2/2023/11/Short-Changed-GHOTD-2023.pdf

Meyer, D. (2023, March 31). Italy bans ChatGPT until OpenAI makes the hit A.I. respect Europe's privacy laws. *Fortune.* https://fortune.com/2023/03/31/italy-bans-chatgpt-gdpr-violations-privacy-ai/

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism.* NYU Press.

Page, L. C., & Gehlbach, H. (2017). How an artificially intelligent virtual assistant helps students navigate the road to college. *Aera Open, 3*(4), DOI: 10.1177/2332858417749220

Pataranutaporn, P., Danry, V., Leong, J., Punpongsanon, P., Novy, D., Maes, P., & Sra, M. (2021). AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence, 3*(12), 1013-1022. https://doi.org/10.1038/s42256-021-00417-9

Roehrick, K. (2020). Valence aware dictionary and sEntiment reasoner (VADER). https://cran.r-project.org/web/packages/vader/index.html

Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. Journal of Applied Learning and Teaching, 6(1). DOI: 10.37074/jalt.2023.6.1.23

Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1). DOI: 10.37074/jalt.2023.6.1.23

Sihare, S. R. (2024). Student dropout analysis in higher education and retention by artificial intelligence and machine learning. *SN Computer Science*, *5*(2), 202.

Taylor, F.W. (1911). *The Principles of Scientific Management*, Harper & Brothers.

Turing, A.M. (1950). Computing machinery and intelligence. *Mind, 49*, 236, 433–460, https://doi.org/10.1093/mind/LIX.236.433

van der Vorst, T., & Jelicic, N. (2019). Artificial intelligence in education: Can AI bring the full potential of personalized learning to education? 30th European Conference of the International Telecommunications Society (ITS): "Towards a Connected and Automated Society", Helsinki, Finland, 16th-19th June 2019. https://www.econstor.eu/bitstream/10419/205222/1/van-der-Vorst-Jelicic.pdf

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197-221. DOI: 10.1080/00461520.2011.611369

Ventayen, R. J. M. (2023). ChatGPT by OpenAI: Students' viewpoint on cheating using artificial intelligence-based application. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4361548

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications, 252*, 124167.

Wolbring, G., & Lillywhite, A. (2021). Equity/equality, diversity, and inclusion (EDI) in universities: the case of disabled people. *Societies, 11*(2), 49. https://doi.org/10.3390/soc11020049

Zeide, E. (2019). Artificial intelligence in higher education: Applications, promise and perils, and ethical questions. *Educause Review, 54*(3). 11 pages. SSRN: https://ssrn.com/abstract=4320049

Ziffer, D. (2025). The right to disconnect has started, giving Australians legal cover to stop responding to emails and calls after hours. https://www.abc.net.au/news/2024-08-26/what-right-to-disconnect-from-work-means-for-you/104234308

Author Information

Raymond Archee

ⓘ https://orcid.org/0000-0001-5301-9248

Western Sydney University

School of Humanities & Communication Arts

Locked Bag 1797, Penrith, NSW, 2751

Australia

Levi Osuagwu

ⓘ https://orcid.org/0000-0002-1727-6914

Western Sydney University

School of Medicine

Locked Bag 1797, Penrith, NSW, 2751

Australia

Myra Gurney

ⓘ https://orcid.org/0000-0002-8806-1726

Western Sydney University

School of Humanities & Communication Arts

Locked Bag 1797, Penrith, NSW, 2751

Australia

Walid Bakry

ⓘ https://orcid.org/0000-0002-6049-9662

Western Sydney University

School of Business

Locked Bag 1797, Penrith, NSW, 2751

Australia

Contact e-mail:

W.Bakry@westernsydney.edu.au

Mauricio Novoa Muñoz

ⓘ https://orcid.org/0000-0002-4014-010X

Western Sydney University

School of Engineering, Design & Built

Environment

Locked Bag 1797, Penrith, NSW, 2751

Australia

Edward Anyasodor

ⓘ https://orcid.org/0000-0003-1928-7657

Charles Sturt University

Rural Health Research Institute

Australia

Citation

Chapter 4 - Artificial Intelligence Proposal to Eradicate Violence between Children in Social Media Interactions

Luis Mario Reyes Pérez Silva

Chapter Highlights

➢ The present study takes place in Mexico City, where two 6th grade students were discussing about soccer teams on WhatsApp. The final argument of the discussion was a death threat, the date was December 22 of 2023.

➢ This study describes the teacher's pedagogical approach at Centro Educativo Didascalos to solve and reconcile a death threat between students. In this proposal we encourage the use of artificial intelligence (AI) to mediate conversations and interactions among children in digital platforms, such as social media apps or video games. Adult intervention is contemplated on this proposal.

➢ Violence and harassment eradication is one of the Autoridad Educativa Federal priorities in Mexico, and maybe the most important one. The Organization for Economic Cooperation and Development (OECD) affirms that Mexico is at the first place on the international ranking on bullying in basic education. In addition, the National Human Rights Commission (NHRC) indicates that seven out of ten children have been victims of this problem in Mexico.

➢ Artificial Intelligence systems could help us to forewarn actions of violence between students that are related to behavior and mental health. AI technology could give children the opportunity to learn which is the best way to say something, to communicate through empathy and respect; always defending their digital rights. Decision making is the key phrase to accomplish that.

➢ Children's mental health data could be misused for private interests.

Introduction

Social media users, between seven to eleven years old, are increasing in Mexico. Social media is a common entertainment for our students in Centro Educativo Didáscalos, a primary school located in Mexico City (Iztapalapa). According to Encuesta Nacional de Contenidos Audiovisuales (ENCCA); 66% of social media users in Mexico, who are children between seven and eleven years old, use WhatsApp to send messages (Villanueva, 2023, par. 1). WhatsApp is an instant message application, property of Meta. A great number of students in our school spend hours of their lives chatting and watching videos on mobiles and tablets without adult surveillance, especially kids from ten to twelve years old.

The past information is important because children communicate and access into the (cyber)space where adults, conscious or unconscious, are absent. As the principal of Centro Educativo Didascalos, I'm very concerned about this social phenomena; as my coworkers are. The channels and traditional methods used by children to communicate have changed to apps, video games and social media sites. "Every half a second a child goes onto internet for the first time" (Miller, 2023, par. 1), the past statistic talk about the speed of access, in combination with the facility to enter in some places of the world, to a "place" where they are going to "navigate" a brand new world on their own. In which other situations do we, as adults, take the same risk with their well-being? Why do we accept to take such risks? Are we aware that children's rights are under direct threat each time they use the web (Miller, 2023, par. 7)?

"Like other children around the world, a significant portion of their lives is now spent in the digital environment." (Miller, 2023, par. 2). The COVID-19 pandemic enhanced this fact. As many schools around the world, we implemented online classes after the Mexican authorities alerted us about the pandemic hazards in April of 2020. We could not visualize, in terms of social media communication problems, what would come after the pandemic safety measures were suspended for schools two years later, 2022. For two years, our students experienced school and personal lives through tablets, mobiles and computers; this way of life brought a series of extreme necessities that we were not prepared to attend.

After the suspension, we decided to redesign our digital education plan, which now is based on the use of mobiles and tablets in the classroom. We no longer work with computers. These

redesign opened a great spectrum of pedagogical opportunities; however, it also opened a bigger necessity, like the urgence to educate on digital citizenship, cybersecurity; and other topics like social media uses and how to prevent harm to others on the internet.

At the present time, some students use social media as a forum to insult each other, misinform about homework and school tasks. Since 2022, the number of problems coming from the use of social media applications is increasing at our school, especially from WhatsApp. The worst case so far is a death threat amid two students. At that time they were 11 years old, attending 6th grade. The threat began for a misunderstanding about soccer team fondness.

The objective of this essay is to propose the use of artificial intelligence to prevent any type of violence and harassment among students in social media, specifically, for instant messages applications such as WhatsApp. This proposal is based on a real case, which occurred in December of 2023. The proposal addresses a pedagogical approach to improve digital education in the community.

Context

Centro Educativo Didascalos has an average population of 130 students. The school is not public, and is divided in two sections, kindergarten and primary school. The population that attend to Didascalos are middle class. The great majority of our community works at informal commerce like street stands or self employment commerce, with long workdays and no social security. The principal caregivers of our students are family members such as grandmothers and grandfathers; or even older brothers and cousins. Most of our students do not count on adult surveillance during the afternoon. Parents spend most of their time in work and city mobilization. Some of them spent one to two hours arriving at their workplaces.

After school reopening in 2022, as I have said, teachers decided to leverage the digital abilities that the students learned from the online classes (2020-2022). We encouraged the community to bring their tablets and mobiles to school. We started to integrate personal devices to the classroom. Parents responded in a positive way. Since the pandemic, they recognized, more than ever, that children must learn digital abilities (see Figure 1).

Figure 1. Modeling on 3D, Fifth Grade

This project was designed to experience teenage pregnancy throughout a series of solving cases problems, better known as Problem Based Learning (PBL). The 3D modeling had the objective to build an immersive experience, presenting the results they obtained in the cases. As I have explained, almost three years have passed since that pedagogical decision. As the principal, I can say and prove that it is one of the best decisions I have ever made. Nowadays, our students use their devices to design infographics, produce podcasts, record shortcuts, and create 3D models by themselves. For our community the use of mobiles and tablets in the classroom signified "the vehicle of personal computing, internet access, and social media" (Giannini, 2023, p. 1).

We couldn't predict all the benefits that the use of mobiles and tablets would bring to our school around digital skills, as we couldn't predict the communication and social problems that these artifacts; without adult and teachers surveillance, would bring to the classroom, such as cyberbullying or abuse. I think about the past, and probably, those problems still happened before the pandemic measures without our knowledge. Despite what I may think; we, as teachers, must act in accordance with our pedagogical decision of bringing mobiles and tablets to school.

School violence can happen as a result of the misuse of social media sites or apps. That's exactly what was/is happening after classes in our school. School violence is defined as any type of violence (physical, psychological or verbal) that occurred inside the school

(Secretaría de Educación Pública, 2016a, par. 2) and outside the school through digital devices such as video games, social media or apps. Parents, teachers, principals and administrative staff can be involved in it.

With the experience of digital violence in school, we have learned that school also happens in the digital world, no matter what time of the day is. School is where our students coexist. If we approach the digital territory as a social territory or sphere, we will identify the digital world as a pedagogical need. (Passeron, 2023, p. 126). Once we started to analyze and solve the online social problems that occurred during WhatsApp conversations between students, our first advice to parents was to block the contact who committed any kind of threat or insult. Sometimes we even encourage them to restrict the internet access of their sons; when we, in the school, ask the children to use it for homework or any other scholar activity. These actions were contradictory measures in any possible way. We know that now, experience taught us about it. We were violating the Free Violence and Harassment protocol in the school, a document made by us. One of the principal objectives in this document is to preserve and maintain, through dialogue, positive relationships between students.

Thinking about our actions as teachers, when it comes to conflicts on web and video games, we should facilitate to our students the space to have a conversation about what had happened, to share their feelings and find a healthy/convenient resolution (Secretaría de Educación Pública, 2023b, p. 6). How do we teach digital abilities when we do not offer a security environment to express themselves? Do we really understand what is happening? Do we have the digital citizenship skills to solve it?

> *This is so important because while children are adapting to the digital world, their caregivers and the laws, policies and services meant to protect and support them are often not. The knee-jerk restriction of children's access to digital technology is not going to help keep them safe online. They just won't ask for help, and that is significantly more dangerous. Particularly given the very real risks for cyberbullying, abuse, trafficking, exploitation and even online radicalization. (Miller, 2023, par. 6)*

Cyberbullying consists in the use of technology as video games, instant messages platforms or social media platforms to intimidate, insult or any other display of violence (Secretaría de Educación Pública, 2023c, p. 7). If we think about that one out of three internet users is a kid

(UNICEF, 2017, par. 1), we may say that cyberbullying and abuse in cyberspace is a global problem. Let's be more accurate, cyberbullying could be a constant danger in schools.

Children started to keep in secret the online social problems between them. Parents and teachers were in total ignorance of what was happening in WhatsApp or video games. When they have a problem on WhatsApp, as their teachers and their parents ordered them, they just blocked the contact, no matter if they were best friends or just classmates. They took, we took, the easiest method to "solve" the conflict, by ignoring it. They just kept coming to school as nothing had happened, when so much had.

Artificial Intelligence Proposal to Eradicate Violence between Children in Social Media Interactions

Violence and harassment eradication is one of the Autoridad Educativa Federal priorities in Mexico, and maybe the most important one. The Organization for Economic Cooperation and Development (OECD) affirms that Mexico is at the first place on the international ranking on bullying in basic education. In addition, the National Human Rights Commission (NHRC) indicates that seven out of ten children have been victims of this problem in Mexico (Secretaría de Educación Pública, 2016a, par. 2).

From many years to 2024, teachers and principals have thought (and learned) a series of social and digital abilities to inhibit violence in schools. This series of activities and projects are established in Plan Analítico. This paper is designed according to students' needs, such as academic, emotional and physical requirements. Plan analítico is an annual plan. Teachers evaluate (monthly) the progress and decide any kind of actualization/editing to it. Plan analítico is divided into three actions of ranges: community, school and classroom. I mentioned it because our Plan Analítico counted with a Free Violence and Harassment protocol; however, we did not specify the actions to take (protocol) when children have a conflict in WhatsApp or any other issue from digital devices such as video games or social media. A huge mistake that we compensated with a profound protocol redesign.

As teachers, when we must apply academic and emotional-safety measures to mediate cases such as cyberbullying or abuse on the web, we are always concerned about the consequences of it on the school environment. That is to say, the untoward in the learning environment. It's

common to perceive a rupture among the children in the classroom when abuse or cyberbullying happens, where little groups are made to support victims and aggressors.

If we, as teachers, omitted the responsibility to rebuild the state of peace in the classroom, we are giving our consent to act in a violent way, to normalize the violent acts. Violence in schools derives from an environment that accepts and legitimate violent conducts, that is to say; the culture shared by the community, that is rooted in the normalization of aggression, accepts violence as a form of socialization. (Secretaría de Educación Pública, 2016, par. 4).

In the next pages (Table 1) I will describe a series of doings derived by the death threat I mentioned at the introduction of this paper. The objective of it is to clarify how it happened and the actions we took as teachers. I decided to mention it because all the happenings that are mentioned in the chapter are action opportunities for artificial intelligence in education. The reader will notice the shortened texts. I decided to write it this way to keep secrecy. We are talking about personal information. Frames without information represent unknown information

Table 1. Death threat happenings

N°	DATE	HAPPENING	PARENTS ACTIONS	PRINCIPAL ACTIONS	TEACHER ACTIONS	STUDENTS ACTIONS
1		-Some students organized a WhatsApp group. It may have been used to schedule online video games or share information about classes, homework, exams, etc.	-I assumed that parents allowed their sons to be part of the group.	-Complete ignorance about the chat group.	-Complete ignorance about the chat group.	-The chat participants started to communicate.
2	December 22, 2023 (holidays) 13:30 p.m.	-One student committed a death threat to another student via WhatsApp.	-The mother of the student who suffered the threat sent a voice note to teacher Fernanda Rodríguez, informing about it. The mother asked for an urgent meeting with the teacher and the principal.	-Teacher Fernanda Rodríguez informed to principals, Luis Mario Reyes, Felipe Reyes and María del Carmen Silva. They agreed to start the free violence and harassment protocol. The first action will be	-Teacher Fernanda and principals agree to have a meeting with the mother of the threatened student. I will address him as Student B. Once we hear her version of what happened, we will have a meeting with the	-The student who suffered the death threat left the WhatsApp group. -I was informed by this on January 9, 2024.

N°	DATE	HAPPENING	PARENTS ACTIONS	PRINCIPAL ACTIONS	TEACHER ACTIONS	STUDENTS ACTIONS
				taken once the winter break ends. As a result of a death threat, teachers and parents must overhaul backpacks at the entrance of the school, this is the "Mochila Segura" protocol. In English, "Safe backpack" protocol. The priority is to find guns and knives as any other explosive artifact. No student is exempt from the protocol, including kindergarten students and teachers. This protocol occurred once teachers informed the community the hour and the date that it will take place.	student who did the threat, student A. -The principal and the teacher are concerned about the violent behavior of student A, it's not the first time he is involved in violent actions (soccer game injury to another classmate).	
3	December 23, 2023		-The mother of student B, who suffered the threat, tried to talk with the mother of student A. The conversation was very short to discuss the situation.	-Principals keep informed about the incident.	-Teacher Fernanda called the principals and informed them about the attempt. The mother did not talk with the other adult.	
4	January 8, 2024 (first day of school) 8:00 a.m.	-Teacher Fernanda Rodríguez and the school principals, Felipe Reyes and Luis Mario Reyes are waiting for the mother who asked for the meeting. She didn't present.	-The mother sent a message to teacher Fernanda Rodríguez. She will not attend the meeting because she needs to go to work. -Teachers and parents participated in the "Mochila Segura" protocol. The result was 0 guns, 0 knives. There is no trace of explosive artifacts in the school.	-I insisted on the meeting and announced to teacher Fernanda Rodríguez to be alert. She must inform me of any strange behavior or misconduct amid classmates. -Contact games are forbidden for the rest of the weekend. -Team work is not an option. Students will take classes, at least for one week, from their	-Teacher Fernanda notices a rupture in the classroom. The students ignore each other. The learning environment is not propitious to study. The companionship rupture is evident.	

N°	DATE	HAPPENING	PARENTS ACTIONS	PRINCIPAL ACTIONS	TEACHER ACTIONS	STUDENTS ACTIONS
				desks.		
5	January 12, 2024 8:00 a.m.			-Teacher Fernanda and I met with the mother of student B. She explained to us that she used to have a friendship with student's A mother. She says that she talked to her days after the threat, December 23, 2023. Mother of student A told her that she would talk to her son about the incident.	-Teacher Fernand and I discussed the meeting. We agreed that Students' B mother looked calm, as if She did want to finish the discussion about the threat.	
6	January 18, 2024			-Teacher Fernanda and I met with the mother of student A. She looks upset. She argued that her son feels uncomfortable with their classmates. He doesn't want to come to school anymore.	-Teacher Fernanda expresses her concern about the constantly violent behavior by student A. -She told the mother that she must come to school to evaluate the scholar performance and social behavior of student A. The mother was not satisfied with the decision.	
7	January 26, 2024	-Teachers month council. We discussed the case and made agreements to prevent and eradicate violent behavior at school.	-None of the parents shared information with us. The Parents and students that were involved decided to break the relationship.	-In agreement with all the schoolteachers, we decided to rebuild and evaluate our Free Violence and Harassment protocol. We are convinced that parents must be involved and in an acquaintance about social media uses. We need their help to impede violent behavior on the web.	-Teacher Fernanda and I supposed that the threat would not take place, as any other kind of it. However, we must keep vigilant. -Teamwork is reestablished, contact games too.	

N°	DATE	HAPPENING	PARENTS ACTIONS	PRINCIPAL ACTIONS	TEACHER ACTIONS	STUDENTS ACTIONS
8	March 13, 2024	-As part of the Free Violence and Harassment protocol updates, the Mexican cybernetic police visited the school to have a chat with the students about cybersecurity.	-The parents involved in the treat did not come. They excused themselves saying that was impossible because of their work.	-I wrote a report about the presentation given by Sub officer Colín. The most important points were. -Legal implications for minors and adults by the misuse of social media and chat applications, such as cyberbullying, grooming, etc.	-Teacher Fernanda and I shared our frustration and impressions about the problem. We ask ourselves if we really took the right decisions to solve it from the beginning. -We are aware that it could happen again.	

Artificial Intelligence Basics, Algorithms and Social Media

Artificial Intelligence is a very extensive topic. Robotics, natural language processing, automated planning and scheduling, optimization, machine learning (ML) and many other components defined what we know as Artificial Intelligence (AI). In a very simple idea, artificial intelligence (AI) is "automation based on associations" (Cardona et al., 2023, p. 11). That is to say, "systems which have the capacity to process data and information in a way that resembles intelligent behavior, and typically includes aspects of reasoning, learning, perception, prediction, planning or control" (UNESCO, 2021a, p. 10). One of the aspects in which AI can help teachers, among a lot of other aspects in life, resides in the improvement of mental health for students, as well as their social well-being. The problem between classmates mentioned on this paper resides on that necessity.

As the death threat happenings chart mentioned, I must say that the past conflict, as any other, is not necessarily a conduct to violence. The threat did not take action; however, it was highly aggressive. Of course it could have happened. The American Academy of Child & Adolescent Psychiatry (AACAP) says that this kind of threat could lead us to just a try to get attention from our children; however, we must pay attention to it for the potential danger to themselves and the community they live with (2019). The first action to prevent any deathfull danger in the school is "Mochila segura" protocol (see Figure 2). One of the updates of the Free Violence and Harassment protocol refers that parents are obligated to come to the school, as an observer of the "Mochila segura" protocol, when they express any kind of lethal

danger to their sons.

Figure 2. "Mochila Segura" Protocol

That's why parents came to the school and talked to us, to the teachers, because children cannot solve their problems by themselves. Sometimes they get so scared that they do not ask for adult intervention and finally find a solution, because that information it's already viral on instant messages groups in WhatsApp or in any social media. This is how the conflict grows, causing misinformation. When this occurs, teachers and parents are no longer "at the control" of the problem, the chaotic logic of internet is, that's the reason why extreme misinformation could lead us to not recognize the responsible of cyberbullying (Secretaría de Educación Pública, 2023c, p. 7).

A great part of a conflict is conformed, no matter if it was born in reality or on the internet, by contrary values/interests, unattended necessities/desires between subjects (Secretaría Educación Pública 2, 2023, p. 69). In other words, a conflict emerges when our thoughts are not compatible with others, the conflict will exist while the need to interact stays human. Parents and students should know that.

As the conflict I presented in Table 1, AI systems could help us to forewarn this kind of school problems that are related to behavior. I start this recommendation saying that any kind of social media should consider AI technology to give children the opportunity to learn which is the best way to say something, to communicate through empathy and respect; always

defending their digital rights. A very useful knowledge to avoid conflict either on the internet or in their real life. As a teacher, I know that the first step to find a solution starts with the responsibility of what we say and do to others. Through this technology children could make the right decision at the right time. Even saving time to resolve it. As the death threat happenings chart illustrates, it took us months to set actions.

Being objective, we did not agree on a solution within parents and students. We just applied extreme measures to forestall violence as the prohibition of contact games or teamwork in the classroom. Those decisions were not benefit our students, nevertheless, a death threat could not be taken as a simple inconvenience.

What would happen if one of my students saw an alert/window display before what he/she was pretending to send, like a death threat? What would happen if their parents could know about the situation in real time? What would happen if, from the beginning, the social media, powered by AI tech, identifies the users as under ages in potentially danger? The next question is how we can answer –and accomplish– the questions from above.

> *AI systems and tools identify patterns and choose actions to achieve a given goal. These pattern recognition capabilities and automated recommendations will be used in ways that impact the educational process, including student learning and teacher instructional decision making. (Cardona et al., 2023, p. 12)*

Decision making is the key phrase. As a teacher, we know how critical it is to improve this ability to keep and maintain a healthy educational environment in school. The automated recommendation can help them to express their frustration and anger. Inspired them to a desirable behavior, inviting children to learn about the violent situation they are passing through. Even propound, to older children with more experience in digital relationships, a wrong answer to verify their learning process (Savolainen, 2023). The learning process could be as profound as we can think.

Does social media stimulate ethical principles and democratic values such as respect, freedom, justice or empathy among users? Digital game-based learning (DGBL) (Figure 3) could be used as a good pedagogical strategy in social media. Through its use children could strengthen their critical thinking based on real situations, solving (autonomously) dangers

raised by real life problems. AI contribution to videogames includes emotional-gestures recognition in real time, voice appreciation, adaptation to environment measures, among many others. That information helps developers to relate and measure creativity and persistence on users (Jara et al., 2020, p. 9).

We want them to take decisions, to be responsible for it, and to mitigate the risks while they are using the internet by themselves. We supposed that any parent and teacher want to keep their children away from unsuitable materials on the internet. Do not forget that children are the most vulnerable group of internet users (UNICEF, 2017, par. 9). Teachers, students and parents must know that the consequences of their actions on the internet are real, as in the physical world.

Figure 3. Interland, a Video Game to learn Digital Rights

Interland is divided in four lands, it has the objective to teach the basic concepts of digital citizenship to children, such as privacy on the internet, social media sharing content, offensive language and many others. According to AI goals in almost every use of it, we encourage educational AI to help humans but also to anticipate/predict risks amongst students, to mitigate potential harms and maximize the benefits of it; especially as the one we are studying at, one danger that occurred away from the classroom. It's important to say that the student who did the threat presented violent behavior before, in digital environments as well. In past years we received intrafamilial violence reports from the student's grandmother. In that sense, mental health professionals agree that predicting behavior it's a very difficult

task, however, they believe that a kid that presented violent and aggressive conduct in the past could relapse in the same conduct in the future (AACAP, 2019, par. 5).

With AI the school could be present, at any time of the day, to those who need it. Moreover, empowering them by socio-emotional skills, thinking skills and communication expertise.

> *AI systems used in learning should be subject to strict requirements when it comes to the monitoring, assessment of abilities, or prediction of the learners' behaviors. AI should support the learning process without reducing cognitive abilities and without extracting sensitive information, in compliance with relevant personal data protection standards. (UNESCO, 2021a, p. 34)*

Let's not forget children's digital rights. I'm not naive and I know that my proposal needs algorithm power to collect any kind of data. The learning analytics process by algorithms is based on a system that analyzes any kind of data; identifying patrons, variables and tendencies (Jara et al., 2020, p. 14). An algorithm can detect human interests with an 84% of accuracy; approximately, in a period of 120 minutes (Calvo, n.d., par. 6). Based on the case we are analyzing; can we assume that any algorithm would identify violent behavior from the user that committed the threat in a lapse of two hours? According to the fact, it is possible. Regarding the information mentioned, we can affirm that any social media user would visualize slanted content after two hours of using a social media platform (Calvo, n.d., par. 6). The speed of the algorithm to assess human abilities or predict learners behavior it's unimaginable.

Just to mention some algorithm actions/scans according to this proposal, one serious risk is the appropriation of mental health information, biometrics as well (UNICEF, 2017a). What does that mean? That sensible information, as psychological profiles, could be sold to private interests, such as employers, colleges or any other institution that wanted information from one subject or a specific group of people. As a result of that risk, we must debate about children's information stemming from their conversations and profile data. Emotional reactions, localization (GPS), connection record, time and hour of the conversation, search history, conversation history, stickers sent, memes sent, pictures as any other content would be subject to AI analysis.

We cannot take for granted the non-commercial purposes of data collecting. The collected data could be the object of misuse and criminal exploitation. Mark Zukerberg, WhatsApp´s owner, was persuaded by the U.S. government to claim a public apology about the abuses and harms (suicides, self-harms, etc.) children suffered in social media (Redacción, 2024, par. 1). Child abuse and violent behavior on internet is a common interest, we should act in accordance with our criticism. Social media products and publications must be a common interest among schools, governments and families (UNESCO, 2021a, p. 35).

I even aim to use social media services without an AI companion, to "take a critical perspective to any outputs" (UNESCO, 2023b, p. 13). Security digital habits around artificial intelligence starts with the habit to verify their answers and outputs, in other words, not trust entirely on AI judgment (Savolainen, 2023, p. 14). Parents and children should be aware of that. AI researchers have found that algorithms could have a certain level of unpredictability, generating doubts and misunderstandings about their outputs (Jara et al., 2020, p. 17).

My desire is to support social media use, ethically and effectively between students. Having options is one of the best ways to empower students, they become masters of their own behavior. In fact, I invite them to discuss AI technology, pertaining to the influence and the consumer engagement that social media enterprises apply for the user/client amusement (Giannini, 2023, p. 5). Teaching about social media uses is the best way to avoid and prevent any kind of misuse.

I accepted that the risks are unreachable, and maybe; just maybe, the benefits of this proposal are not enough to take them. The truth is that this proposal model will respond with accurate and relevant outputs as a result of the data recollection (UNESCO, 2023b, p. 13) by means of interactions between children in social media.

Is it time to ask if AI in social media would foretell school shootings? Would AI predict it? Theory says it could be possible, however, we must prevent any abuse of it, "humans must determine the types and degree of responsibility we will grant to technology within educational processes, which is not a new dilemma." (Cardona et al., 2023, p. 13). The data recollection related to children's interactions should be discussed from legal terms to scholar points of view. The ethical dilemma that I'm building in this proposal is prone to criticism. I'm aware of that.

The next chart will describe the possible artificial intelligence actions to prevent violent behavior amid students. Table 2 is divided in six numbers; the proposal does not limit the AI scopes. The conflict stages are based on Paco Cascón Soriano, *Educar en y para el conflicto* (2001).

Table 2. Artificial Intelligence Actions to Prevent Violent Behavior in Social Media or Instant Message Applications

Nº	AI fields	Definition	AI actions	Conflict stages
1	Machine learning	"AI systems are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments." (UNESCO, 2021a, p. 10)	Analyze context: culture, language, media trending, search history, etc.	Necessities
	Problem happens			
2	Learning analytics	Continuous and automatic algorithm data used to improve performance (UNESCO, 2023b, p. 8).	Once the conflict is loosed, find common interests and needs to avoid the crisis Evaluate possible solutions. Parents are aware of the problem.	Problem
3	Generative AI	"Imitate human capabilities to produce outputs such as texts, images, videos" (UNESCO, 2023b, p. 2)	Propose a mediated dialogue, based on inferences and inductions, along with parents consent.	Crisis
4	Negotiate a solution	AI systems are designed to operate with varying degrees of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations. (UNESCO, 2021a, p. 10)	Find an autonomous solution according to a learning adaptive system; that is to say, personalized learning.	

Nº	AI fields	Definition	AI actions	Conflict stages
5	Report	The document would inform parents (and teachers) about conflict approaches: -How do they perceive the conflict? -Which topic was the center of the conflict? -Do they use racist, homophobic or anti-immigrant language? -Are there other classmates involved? In which way?		Discuss a solution with adult mediation.
6	Optimization and "deep learning"	Known as algorithm "training", the "deep learning" is a mathematical process where computers "learn" how to resolve a conflict (Jara et al., 2020, p. 4)	Evaluate, following up on the conflict Restart Nº1	

Note: This chart is just a basic approach to understand the capacity of AI in education. It does not have the intention to limit the uses of technology.

As the chart exposes, those six steps/stages could function as the base of the AI system to prevent violence between underage on social media apps. About the chart, from stage number 3 to number 4, How the AI-user interaction would function? "The scope of pattern recognition and automated recommendations will expand" (Cardona et al., 2023, p. 12), that is to say, the AI approach will be focused on warnings, to prevent harm, about what the minor is doing and saying on social media. As I have said, I´m not talking about strict control, not even algorithm conduct manipulation (Calvo, n.d., par. 5). We want users to access a series of mechanisms and notifications to live together on the internet, to be digital citizens without reducing cognitive abilities.

I want to make it clear that this proposal goes beyond the question-and-answer or answer-and-question AI approach (Thomas, 2023, p. 3). The question-and-answer method could be just a first stage system. One fundamental idea of this proposal is the transformation and overcoming of the AI customer model to a student-centered approach (UNESCO, 2023b, p. 13) based on generative AI. This overcome will be implemented by relevant learning methods such as problem solving in digital (and classroom) environments based on decision-making.

The past strategy is already on testing. Khamingo is a pedagogical assistant powered by

generative artificial intelligence, launched in 2023. In Khamingo every interaction made by an underage (eighteen-year-old minors) is registered, parents and teachers have access to the chat/conversation history. To elevate student security, there is a second AI system that supervises conversations. If the student is having a suspicious conversation related to sex, harm, abuse, etc.; the AI suspends the service, alerting parents and teachers (Savolainen, 2023, p. 14).

Continuing with the algorithm actions of this proposal, parents will have access to the conversation history (the option is already in function in WhatsApp) of what and how their children communicate in social media, stage number 5. This conversation history would specify in How he or she perceived the conflict? Which topic was the center of the conflict? Did they use racist, homophobic or anti-immigrant language? Which options, proposed by the AI, did he or she take to avoid the conflict? He or she had the disposition to arrange an agreement.

About the case we are studying, parents would have the data of the incident, as accurately as it can be. According to this proposal, the conversation history would highlight potentially dangerous take-action messages, such as access to weapons, planning to bring a weapon to school, family violent behavior, thinking of harming others, witnessing abuse or harm to others, cruelty to animals, etc. According to AACAP (2019), the past actions are risk behaviors, ones that could increase violence in any context. As I will explain in the next paragraphs, the conversation history will register the actions to prevent harm and violent communication between children correlated to the decisions they made at the time the social media platform (AI) warned them about it.

Once parents have the conversation history, in the company of their children, they could analyze and think over their social media actions/behavior as they do in real life in places like schools and homes. They would study the learning process and why not, the assessment made by the AI based on communication skills on cyberspace. Parents and children will have the liberty to share the conversation history with teachers to discuss how the problem occurred and how they can solve it, of course, with a pedagogical approach.

Hypothetically, if only I could have had the report I´m proposing, it could have been a very useful tool for me as the principal, because I could have taken better decisions, such as

conflict-resolved strategies or emotional discharge activities with them. I certainly could have prevented the death threat. That information (data) could have helped teachers and me "to generate appropriate educational materials such as lesson plans, quizzes and interactive activities that closely align with an effective pedagogical approach and specific curricular objectives" (UNESCO, 2023b, p. 13). Thanks to the learning adaptive system (personalized learning), the report could add specific pedagogical strategies for each student, considering students personality and interests; strengths, social and academic competencies, even moods (Jara et al., 2020, p. 7-8).

Security in AI is not to keep them behind an "eye" that sees everything. Security also involves systems that invite them to interact with the outputs/answers proposed by the AI (Savolainen, 2023, p. 14). That is why I propose a real time accompaniment, generating new ideas and solutions to real-world challenges (UNESCO, 2023b, p. 8), by making a stress situation a pedagogical situation.

> *Educational applications will be able to converse with students and teacher, co-pilot how activities unfold in classrooms, and take actions that impact students and teacher more broadly. There will be both opportunities to do things much better than we do today and risks that must be anticipated and addressed (Cardona et al., 2023, p.12)*

The decision-making record will play a major role to prevent, even anticipated, school harassment. "Parents can help their children to take control of how others see them online and create an online reputation that will highlight their strengths and passions to serve them in the future" (Prakash, p. 4910, 2019). Parents will accompany their children in cyberspace, a "place" that "was not designed with children's safety in mind" (Miller, 2023, par. 7). This is how children will be empowered to build their own digital identity, discussing what content, idea or emotion led him/her to act in the way he/she did. Let's not forget that digital content could influence them in how, and for what reason, apply their values (Passeron, 2023, p. 148). They will decide, no matter the manipulees conspiracies around AI human control, how they want to communicate with others, how to be seen. In a time when human behavior is measured as a product (Calvo, n.d., par. 1), the ability and the opportunity to make non-influenced decisions on social media seems to be a human right.

As I have informed, the data could be presented to the teachers to study the case, where the victim(s) and the responsible(s) of the violent act will be attended. Not to determine a guilty but heal the social tissue and find NO REPETITION agreements and strategies. Parents, students and teachers will participate in the agreement discussion, giving feedback and practicing democracy values as the dialogue with authorities is (Secretaría de educación pública, 2023c, p. 16). I consider the past action as another fundamental part of this proposal, dialogue to recognize rights and responsibilities; in other words, sharing the entire responsibility for what happened. This action will help them to augment their face-to-face communication abilities (Chiu, 2023, p. 3) as they could not practice it in social media with AI consultancy through self-regulated learning activities.

It's important to say that educational AI systems allow personalized learning experiences for students (Jara et al., 2020, p. 6). Based on machine learning (stage Nº 1), alongside learning analytics systems (stage Nº 2), AI technology is capable of offering a learning adaptive system to users, and teachers too. Starting from the profile data (interactions, answers in conversation, etc.), AI systems could provide accurate and efficient academic content to students, considering variables like difficulty, sequence, learning progress and learning rhythm (Jara et al., 2020, p. 7).

Based on my experience as a teacher, when I asked my students how the internet works, a great number of them answered me by describing a scroll on a social media feed as TikTok or YouTube. They do not know that the internet is subject to economic and political interests, "there is nothing neutral on the web" (UNESCO, 2020, p. 5), I said to them each time we analyze internet topics, that's the main reason why we need a democratic oncoming to solve social media and video games conflicts among children. The internet mechanisms are not propitious to democratic discussions.

Going back to the Free Violence and Harassment protocol updates and redesign related to the death threat happenings, the Mexican cybernetic police visited the school on March thirteen to have a chat with students and teachers about cybersecurity. Sub officer Colín visit was (Figure 4) very significant for the students. Teachers and students looked interested in how police officers act when they receive legal complaints related to grooming or cyberbullying. She was very emphatic in the legal consequences derived from internet crimes like the one I'm narrating. I must admit that at some point of the conference the atmosphere was tense. I

suppose that students remembered the problem among student A and student B.

The parents of the children who were involved in the death threat did not assist the conference, they excused themselves arguing a work overload. Moreover, student A did not come to school. I was very disappointed, even angry. We are talking about a potential conflict of extreme violence. In January parents asked us to double our efforts to eradicate it. Why don't they share and assume their responsibility with the school? Do they presume that the death threat was just a misinformation problem? Just a joke? If they do, the conflict has taken another course, pointing at a huge cybersecurity ignorance by our parent community.

Figure 4. Sub Officer Colín Conference, Cyberbullying

 Kids formulated great questions, such as; How can we trust in the cyber police if they have access to personal data? Why do hackers always attack the governments? etc. Sub officer Colín told them that Mexican law is based on one principle: you are guilty until you prove otherwise. Even though I'm conscious about the Mexican law procedures, I started to think that the case I'm studying could have ended in a legal sentence for the children that were involved. She taught us about the internet legal framework. This part was the most challenging for students. They had difficulties understanding that the internet is regulated by the state, or at least there is an intended to. Students looked incredulous about it. They asked Sub Officer Colín, on several occasions, how it's possible for the cyber police to trace them on the web. She explained to them that the internet is built like an endless timeline that police can deconstruct, as a result, everyone left "traces" on it. The teachers that were with us at that

moment, I supposed, felt the same as me. We perceived that children envisage cyberspace as a "place" without law support and institutional limitation. The teachers identified urgent academic needs: digital citizenship (Figure 5) and ethical behavior on the web.
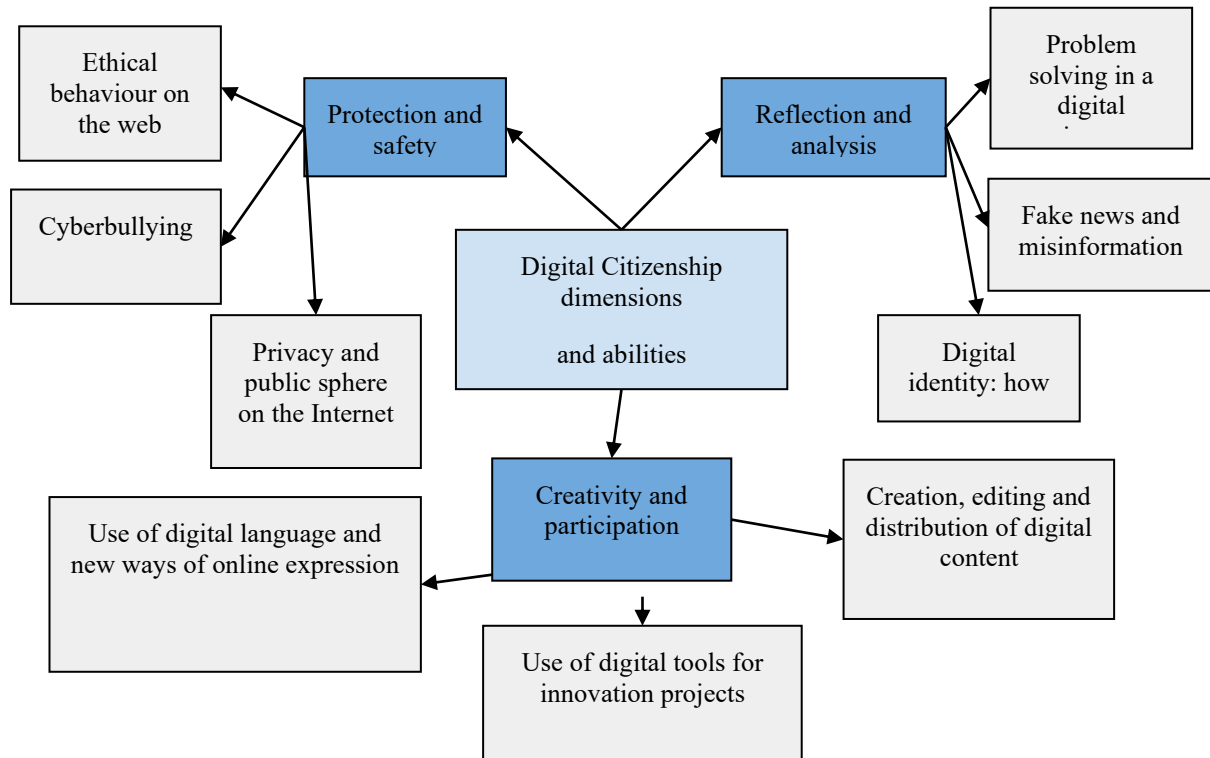

Figure 5. Digital Citizenship as a Public Policy in Education in Latin America

The teaching group committed to work in creativity and perception as an urgent ability to improve in our students. This decision was established in the improvements of the Free Violence and Harassment protocol. As the chart exemplifies, one of the most important characteristics for a digital citizen is the ability to analyze and evaluate content to make informed decisions. If we educate children based on the digital citizenship dimensions and abilities, algorithm actions will not influence them as we think it does. Algorithms do not make decisions, persons do.

Consider that technology "exhibits and privileges certain worldviews and reflects ways of thinking and knowing. New generative AI models and utilities are no exception" (Giannini, 2023, p. 3). We must debate if the internet legal framework considers the intrinsic economic and political interests on the web as forces that influence humans. Those forces work as they do in reality, however, we must regard the amazing data analysis, as accurate as it can be, by AI technology. Otherwise, we live in an obsolete and contradictory cyberspace legal

framework.

Sub Officer Colín also told them that "Information stays on the internet, no matter your efforts to erase it". I felt powerless at that time, because it meant that the death threat committed by student A could transcend over his/her digital life, it can be used as a legal antecedent of violent behavior. The sub officer sent by the Mexico city cyber police expressed that legal rulings apply to children and their caregivers; their parents. If both of them are dealing with any kind of cyberspace procedure they would appear to the judge. Any threat, abuse or "joke" committed on video games, social media or any other web app is subject to a legal denouncement. In Mexico, children can appear to a judge starting at the age of twelve.

The scope of this proposal, to apply AI technology in the benefit of social media interaction or video games among children; could serve to forestall unnecessary legal procedures, even law abuses; starting with removing "incriminatory content against children and implement the right to erasure" (Prakash, p. 4910, 2019). That is to say, children should have the right to rebuild their identity on social media and video games along with artificial intelligence, training on new ways of online expression and; especially and more importantly, in the company of adults and teachers. For this proposal, children will know that they could be free of the algorithm influence, they will understand how algorithms work and how persuasive they could be.

Just imagine how transcendent it is to be the responsible of a death threat on the web with eleven years old, to be judged by classmates, and even teachers, as a "violent" kid. A child cannot be classified as a criminal just because he or she lacks adult assistance on the web. Even though our students should have the opportunity to mold their own identity in cyberspace without algorithm influence, apparently, they are not aware of that digital right.

Thinking about my decision to bring a police officer to the school, I accept that the strategy to invite her could have been intimidating for some children. The police officer did not bring a gun, the emotional impact could have been stronger otherwise. I realized that I had called a legal authority, based on the urgency to keep them safe. Did I accomplish my goal to prevent violent behavior by observing a real police officer? I don't think so. My intentions were not intimidatory, however, I really wanted to make a clear statement, parents and children must

be responsible for their acts, to avert unhealthy conduct on the internet stemming from the online disinhibition effect (Farrar, 2019, par. 1) (Figure 6), as the fantasy of being "invisible" on the web. The disinhibition effect it´s a conduct that normalizes violence based on the feeling of being anonymous. Anonymity on the internet is not a privilege to act as you wouldn´t act in the real world (Passeron, 2023, p. 131), anonymity on the internet should be a right in the name of freedom, as journalism is.
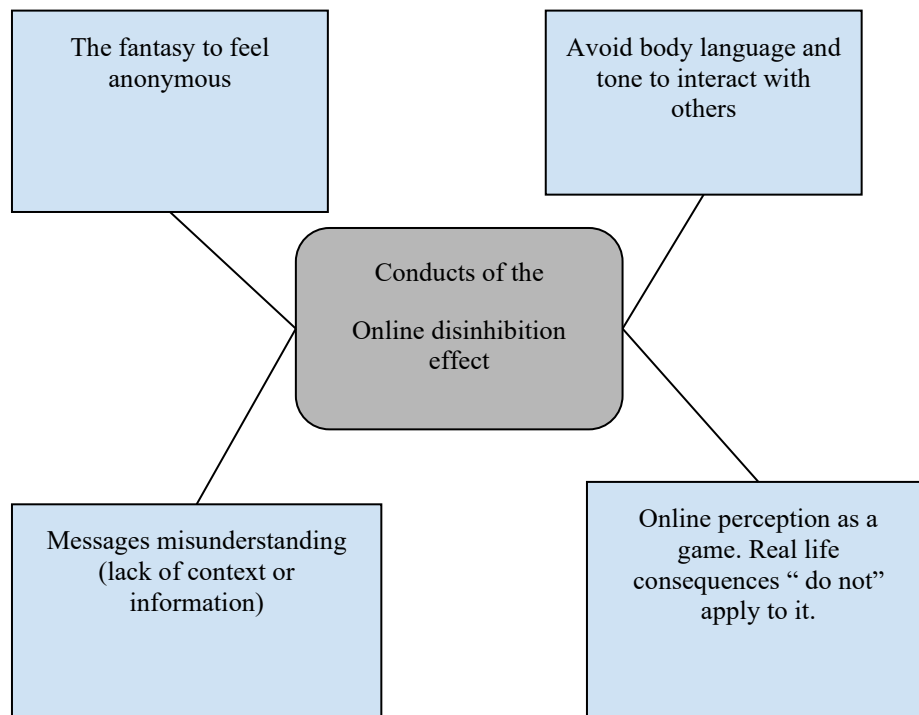

Figure 6. Conducts of the Online Disinhibition Effect

I consider the online perception as a game as the most dangerous conduct. Before this paper, I must accept that I did not know about the online disinhibition effect in people. In consequence of what I observed in officer Colin´s conference, where children assumed the Internet as a "place" without law, I started to investigate how and why this phenomenon was happening.

Children should know when and how to share photos and personal details on cyberspace, even the use of stickers and memes. Some people can interpret certain types of messages as offensive, misunderstanding the significance of it. As teachers and parents, we are lacking digital literacy strategies to teach how to use digital language in the proper way. As Sub official Colín warned us, the misappropriation of memes, stickers or photographs could lead

us to legal procedures. I accept that our efforts as a school are not enough, maybe we should redirect our digital education school program; changing the Master of Software and programming for habits of communication to construct healthy relationships in social media. We must think that internet infrastructure/configuration promotes communication problems such as cyberbullying (Secretaría de Educación Pública, 2023c, p. 18).

> *Our children are born and growing up in a Dystopian world where computers and internet play an indispensable role in the choice of food they eat, the friends they make, the clothes they should wear, the games they should play and the future partners they may choose. (Prakash, p. 4097, 2019)*

As I have already said. I'm not saying that teachers and parents should watch every chat conversation and regulate every social interaction in WhatsApp or TikTok. Not even to assign this enormous task to a predictable AI chatbot that tries to teach students how to be polite on the internet. "AI chatbots function, therefore, like all-knowing oracles" (Giannini, 2023, p. 3). There's nothing worse than an AI Chatbot that emulates a "smarty" human talk. No matter the technology advances, Chatbots limitations are visible to any user. They lack "abilities" to efficiently detect intentions, meanings and intrinsic implications in the human language (Hill et al. 2015). What I'm proposing is the consent and the responsibility of the decisions they made on internet, the propel digital literacy based on the protection and safety of themselves and others, even though reflection and analysis of their actions and words (UNESCO, 2020, p. 7); appealing to a series of abilities and skills arise from the dialogue within families-classmates-teachers. The AI will be just a method to know themselves and how they act on the internet, the automated recommendations and warnings suggested by the AI will be just a channel to clarify their communication problems and affection needs, adapting to them in a positive learning experience (Thomas, 2023, p. 3). The incorporation of AI technology as a conversation mediator It's already proven, stimulating collaborative work and problems resolution skills (Jara et al., 2020, p. 9), however, it's crucial to balance individual assignments with collaborative activities.

I think that this proposition would function not only for extreme cases like the one I explained, BUT the AI could ALSO work in different scenarios and different modes, such as schools groups, study groups, or any other academic approach on WhatsApp and video games.

The use of AI in this proposal will be successful if only students, parents and teachers work together, by solving the conflict with adult mediation. If we take the conflict as an opportunity to learn and work on critical thinking abilities, we have just started to build the solution (Secretaría de Educación Pública, 2023b, p. 69). We must understand that children learning is under, most of the time, on teachers' supervision, "This implies that teachers play a key role in supporting learning and teaching using technology in classrooms" (Chiu, 2023, p. 2).

Teachers will play a very important role in this proposal because; as we do it in the classroom, we have to observe, in company of kids and parents, the body language, vocal tones and perceived their attitude to solve a conflict. The myth that narrates teachers will be substituted by AI is just that, a fiction. If machine intelligence cannot be equated with human values, we, as humanity; must maintain the responsibility to judge and measure our actions (Giannini, 2023, p. 3). The moral judgments must stay under human control, otherwise we are abandoning the principles of education.

In our scholar community we are used to arbitrating and conciliating any kind of communication breakdown/misunderstanding from social media and video games, approximately once a week. Students come to school with these bothering that impact on their academic performance. The nature of media technology at present times causes the immediate spread of any harassment or violent behavior between persons, exceeding the scholar atmosphere; at this point we lose any attempt to control it (Secretaría de Educación Pública, 2023c, p. 7). I mentioned it because we must understand how common it is for us and how difficult and slow it is for teachers to arbitrate and solve a conflict that was originated in social media and video games. Most of the time, violence on the internet stops throughout time; just after days, weeks or maybe months of being viral, to finally be ignored by the public, substituted for another matter (Passeron, 2023, p. 155). Meaning that the violence victim was forgotten, but not helped in terms of mental health or physical health.

The consequences for not acting with respect, tolerance and empathy in the digital world should resound in reality, as sub official Colín told us; however, we are committing a terrible mistake as humanity if we continue to use the internet as a "place" without legal guarantees; even worst, not taking actions to qualify our students in terms of digital citizenship dimension and abilities.

Limitations and Responsibility

Is it possible to identify values and interests, through AI technology, to predict a conflict? Once again, the theory says yes; nevertheless, the risk to achieve that goal could be dangerous. At the time I wrote this paper, another student committed a death threat to another student via WhatsApp. This case happened in fifth grade (March 2024).

When that happened, parents come to me in the search of guilty and justice for their sons. I inform them that we are going to work as far as our role as teachers demands, based on the official and federal guide to administrate schools functions in Mexico, *Guía Operativa* (2023); in addition to the Free Violence and Harassment protocol designed (and updated) by us. However parents and social media enterprises must assume the responsibility. They should know, based on officer Colín information, when a death threat happens they should go to the public ministry to report a crime. Even Though, they must know that willpower and treat abilities are essential elements to accomplish a conflict solution (Secretaría de Educación Pública, 2023b, p. 74). Are they prepared to work with us?

The use of social media platforms requires new ways and methods of conflict solution, as the one I'm proposing. Education policy-makers and institutions should have a clear understanding about why, how, and what children are learning in this new phase of the digital era (UNESCO, 2023b). A global accepted normative could be a very difficult goal to reach, nevertheless, technology works on global uses in any culture.

> *This must comply with international law, including the United Nations Charter and Member States' human rights obligations, and should be in line with internationally agreed social, political, environmental, educational, scientific and economic sustainability objectives, such as the United Nations Sustainable Development Goals (SDGs). (UNESCO, 2021a, p. 18)*

How to end this school problem no matter if it happened outdoors, in the time they are supposed to be in the care of an adult? "Parental monitoring may be the best mechanism, however, it may not be possible without the backing of a suitable legal process" (Prakash, p. 4097, 2019). Just to mention one fact to understand the seriousness of the case we are studying, thirteen years old is the minimum age, for the American continent, to participate in

social media such as WhatsApp. In Europe, the minimum age is sixteen. (WhatsApp, 2024, par. 2). When I asked parents why they allowed to download WhatsApp on their children's devices in the first place, they answered that it's an easy way to communicate with them while they are at work. I responded to them that I understand the situation, we can not change the context we lived in. However, we can manage how to coexist on the web without adult monitoring, to be safe. The use of AI to improve security, at any level, is already proved (Thomas, 2023, p. 3), Why not apply it to schools?

I´m aware of the security methods implemented by WhatsApp to under ages. One of them is Parental control, this mechanism is based on restrictions, like blockings. Blocking is a mechanism of security, a technical response; not a dialogue strategy. Children need strategies to repair relationships. Furthermore, WhatsApp security measures to minors consider, by legal procedures, that the user is a teenager (thirteen years old), not a child. Strategies based on limitations avert children to take decisions, to be responsible for their acts, it also encourages them to be just consumers and no proactive internet users (Jara et al., 2020, p.8).

I consider that, as a first approach, regulatory limitations to mediated user behavior on social media is just the first step to the use of AI. Parent control is an obvious choice to mediate children's communication on web, and the most important one; however, and overage of this strategy could incite to a vicious circle of dependence in a virtual world where they need to make their own decisions, by growing and maturing on their own rhythm (UNICEF, 2017b). This sort of politics encourages digital citizenship ignorance; limiting the development of their digital citizenship abilities and dimensions (Secretaría de Educación Pública, 2023c, p. 18). We, as society, should overpass the prohibition method to guarantee the security of our children on the internet.

As teachers we demand that digital products, such as social media or video games, should reflect children's needs (UNICEF, 2017, par. 4), even more, to be design-manufactured with a human-centered vision (UNESCO, 2023b) (Figure 7), we can accomplish that by laws and international protocols that construct mechanisms for the benefit of internet communication, to exercise social abilities on the web (UNICEF, 2017b). I insist, I'm not proposing a punishment regime to children on the internet. My intentions are far from those. I persist in this point because they put themselves in danger by not learning from the experience, the legal consequences are obvious. That is why parents should know about their son's activity on

the web. It's time to act for the benefit of our future, making educational decisions through evidence (Chiu, 2023, p. 3).
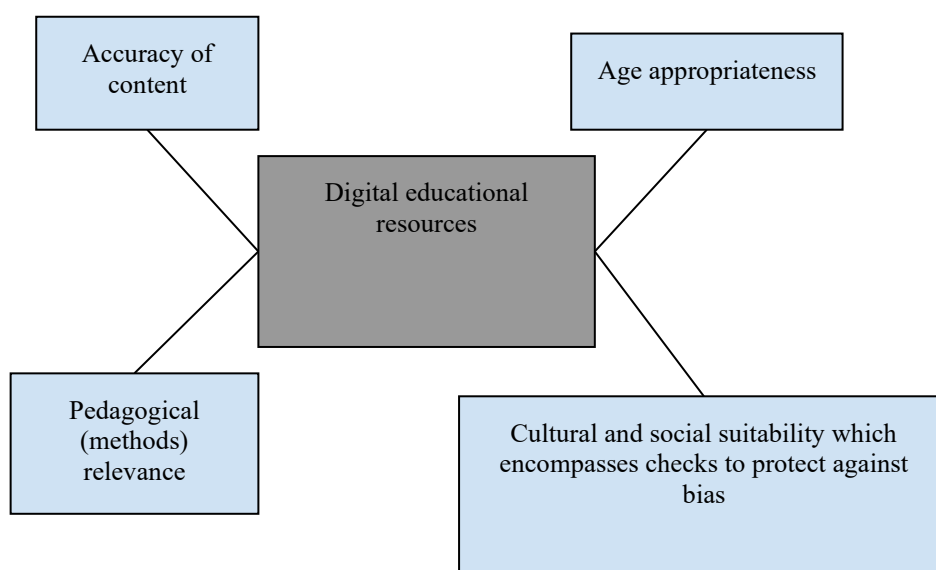

Figure 7. Main Criteria for Educational Artificial Intelligence Resources, proposed by Stefannia Giannini (2023)

In my understanding, one of the most important digital educational resources, around AI in education, is the relevance of the pedagogical methods that AI is constructed by. Those methods have to be suited to the real world; understanding the internet as a social space that reflects human conduct (Passeron, 2023, p. 153). If we really intend to solve a conflict that was born on video games or social media, we need to democratize the decision-making process; that is to say, children should participate in the agreements established by de AI, agreeing on the consequences for inadequate behavior on the internet, not punishments. That is how students and teachers will discuss over *NO REPETITION* agreements and strategies. These activities will show to teachers, students and parents a real democratic environment, when everybody has a responsibility to defend and speak for themselves. In consequence of the responsibility acceptance, students will have a constructive influence in their schools and community, practicing democracy values and having a particular experience of how public policies work out (UNESCO, 2020, p. 5).

Conclusion

I start this conclusion by saying that the use of any kind of artificial intelligence does not

represent the end of educational challenges. Contrary to it, it opens a myriad of debates about what, how and why we learn. Although educational AI programs are in an early stage, this case showed me the power of AI, but it also made me conscious of great academic needs, such as digital literacy and media literacy in our student community. The use of artificial intelligence could guide us to save precious time to solve and prevent violent behavior, such as death threats. Among adult intervention, artificial intelligence could be a tool, "a live guidance", to relate in a positive and peaceful way with others. Algorithm bias may be overcome if we address such technological power to identify risk behaviors between children interactions. Through time, children will make their own decisions on the internet, accompanied by their parents and teachers, to live in the digital world. This process will be progressive, gaining time to assess abilities and understanding how the internet and videogames work.

How reliable is this proposition? I'm aware of the economic impact and the technology development that this proposition represents. I'm talking about years of scientific research and even more time to develop a global policy based on AI for education. It really doesn't matter if this proposition never comes to light, I realized that just to think about it signified a school reorganization for all the teachers, demonstrating our limitations as teachers. The updates and the renewal of the Free Violence and Harassment protocol gave the school strategies and legal bases to act, and prevent, a variety of violent cases. We gained critical reasoning based on social media and video games interaction. We also redefine our school definition, because schools also happened in cyberspace.

References

Calvo, J. (n.d.). Algoritmos en redes sociales: ¿cómo afecta a los niños y adolescentes? Educación 3.0. https://www.educaciontrespuntocero.com/opinion/algoritmos-redes-sociales/

Cardona, M. (Ed.), Rodríguez, J. & Ishmael, K. (2023). *Artificial Intelligence and the Future of Teaching and Learning*. Office of Educational Technology

Cascón, P. (2001). *Educar en y para el conflicto.* United Nations Educational, Scientific, and Cultural Organization (UNESCO).

Farrar, L. (2019). Is the internet is making you meaner? *KQED*.
https://www.kqed.org/education/532334/is-the-internet-making-you-meaner

Giannini, S. (2023). *Generative AI and the future of education*. United Nations Educational, Scientific, and Cultural Organization (UNESCO). https://unesdoc.unesco.org/ark:/48223/pf0000385877/PDF/385877eng.pdf.multi

Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). "Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations". *Computers in Human Behavior, 49*, 245–250. https://doi.org/10.1016/j.chb.2015.02.026

Jara, I., Ochoa J.M. (2020). *Usos y efectos de la inteligencia artificial en educación*. Banco Interamericano de Desarrollo (BID).

Miller, S. (March 28th 2023). Child Protection In The Digital World: Why Is It Needed?. *Save The Children*. https://www.savethechildren.net/blog/child-protection-digital-world-why-it-needed#

Passeron, E. (2023). Ciudadanía relacional: La convivencia escolar con tecnologías digitales [PHd Thesis, Universitat de Barcelona]. https://diposit.ub.edu/dspace/bitstream/2445/212461/1/EP_TESIS.pdf

Prakash, A. (2019), "Parental Role in Creation and Preservation of Digital Identity of Children". *Test engineering & management. Volume 81*, 4907-4911.

Redacción. (February 1st of 2024). "Mark Zuckerber se disculpa ant las familia de niños que han sufrido daños por culpa de las redes sociales". *BBC*. https://www.bbc.com/mundo/articles/c72gze8r05jo

Savolainen, A. (2023). Sal Khan: "Veo la IA como una herramienta suplementaria, particularmente poderosa". *El correo de la UNESCO. Volume* (Nº 4), 12-14.

Secretaría de Educación Pública (February 12th 2016a). *Violencia en las escuelas*. https://www.gob.mx/justiciacotidiana/articulos/violencia-en-las-escuelas?idiom=es

Secretaría de Educación Pública (2023b). *Eduquemos para la paz en las escuelas.* Secretaría de Educación Pública.

Secretaría de Educación Pública (2023c). *Todas y todos contra el acoso escolar. Guía para docentes de educación básica.* Dirección General de Gestión Escolar y Enfoque Territorial adscrita a la Subsecretaría de Educación Básica.

The American Academy of Child & Adolescent Psychiatry (AACAP) (January 2019). *Threats by Children: When are they Serious?* https://www.aacap.org/AACAP/Families_and_Youth/Facts_for_Families/FFF-Guide/Childrens-Threats-When-Are-They-Serious-065.aspx

Thomas, K. F. Chiu (2023). The impact of Generative AI (GenAI) on practices, policies and

research direction in education: a case of ChatGPT and Midjourney. *Interactive Learning Environments*. DOI: 10.1080/10494820.2023.2253861

The United Nations International Children's Emergency Fund (UNICEF) (2017). *Make the digital world safer for children - while increasing online access to benefit the most disadvantaged.* https://www.unicef.org/serbia/en/press-releases/make-digital-world-safer-children-while-increasing-online-access-benefit-most

UNICEF (2017a). *Child Privacy in the Age of Web 2.0 and 3.0: Challenges and opportunities for policy*. The United Nations International Children's Emergency Fund.

United Nations Educational, Scientific, and Cultural Organization (UNESCO) (2020). *Digital Citizenship as a public policy in education in Latin America*. Montevideo Office. Regional Office for Sciences in Latin America and the Caribbean. https://unesdoc.unesco.org/ark:/48223/pf0000376935_eng

UNESCO (2021a). *Recommendation on the Ethics of Artificial Intelligence*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000381137.locale=es

UNESCO (2023b). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Villanueva, D. (April 25th 2023). "En 5 años, se dispara el uso de redes sociales en menores: IFT". *La Jornada*. https://www.jornada.com.mx/noticia/2023/04/25/sociedad/en-5-anos-se-dispara-uso-de-redes-sociales-por-parte-de-menores-ift-3102

WhatsApp. (n.d.) *Acerca de la edad mínima para usar WhatsApp*. Retrieved Abril 2, 2024 from https://faq.whatsapp.com/695318248185629/

Author Information

Luis Mario Reyes Pérez Silva

https://orcid.org/0009-0004-1735-9586

Escuela Nacional de Maestras de Jardines de Niños, ENMJN

Calle Gustavo E. Campa 94, Guadalupe Inn.

Álvaro Obregón, 01020 Ciudad de México, CDMX

Mexico City

Mexico

Contact e-mail: *luis.pos@aefcm.nuevaescuela.mx*

Citation

Reyes Pérez Silva, L.M. (2025). Artificial Intelligence Proposal to Eradicate Violence between Children in Social Media Interactions. In O. Noroozi (Ed.), *Artificial Intelligence (AI) in Education* (pp. 84-116). ISTES.

Chapter 5 - From Automatons to Autonomous Beings: The Humanistic Evolution of Robotics and the Pursuit of Sentience

Christopher Dignam 🆔

Chapter Highlights

- ➢ Robotics is a contemporary, highly evolving branch of computing, engineering, and technology that also shares many elements of design and form with the sciences, mathematics and the arts. While robotics is a relatively new mode of STEAM (Science, Technology, Engineering, Art, and Mathematics) the term robotics, which stems from robot, first appeared in Karel Čapek's 1921 play, Rossum's Universal Robots (Capek, 2004) and is derived from the Slavic, robota, meaning serfdom.

- ➢ The drudgery of serfdom can be traced back one thousand years earlier to the tenth century with the breakup of the Carolingian Empire and tenant farmers who were bound to landlords for menial agricultural labor (Bloch, 2023). However, the logarithmic evolution of robotics, beginning with mid-twentieth century toiling industrial robots to early twenty-first century, self-thinking Artificial Intelligence (AI) robotics, has resulted in robots that are highly intelligent, self-learning, and unlike Karel Čapek's serfs.

- ➢ In this study, the researcher explores the evolution of robotics, AI, and implications for education, industry, and society. This study examines robotics as an experiential form of learning from early childhood through higher education settings for eliciting critical-thinking, communications, and teamwork.

- ➢ This study also investigates the algorithmic nature of AI and efforts for robotic self-learning in the pursuit of self-awareness and consciousness. The researcher concludes by making recommendations concerning robotics implementation in educational settings, as well as humanistic philosophy, theory, machine learning, and advances in Artificial Emotional Intelligence (AEI).

Introduction

Insights Unveiled: Exploring Robotics and AI

Robotics is an ever-evolving and multidisciplinary field that encompasses progress in diverse areas such as speech comprehension, industrial uses, medical innovations, sensors, controllers, and a range of other subjects and has experienced broad acceptance across various industries and research establishments globally (Stone, 2018). Robotics is a multifaceted domain involving a range of robot types, applications across industries and services, as well as remote operations, emphasizing mobility, end-effector tools, control techniques, sensors, and the incorporation of cutting-edge technologies (Williams, 2019). As robotics continues to advance, it serves as a catalyst for innovation and efficiency, revolutionizing sectors such as manufacturing, healthcare, and transportation while paving the way for unprecedented levels of automation and human-robot collaboration. With each new breakthrough, robotics solidifies its position as a cornerstone of modern technology, driving progress and reshaping the future of education, work, and daily life.

Robotics is also a developing interdisciplinary field that concentrates on programming robots to enhance their functionality and safety, with a strong emphasis on collaboration across various disciplines and the societal implications of robotic advancements (Koditschek, 2021). Robotics is a contemporary field of study that holds potential for significant impact on educational settings and learning processes (López-Belmonte et al., 2021). In 1975, the first publication related to robotics in education appeared regarding the exploration of using robotics to enhance the learning process, marking the beginning of scientific exploration in this field (López-Belmonte et al., 2021). As robotics permeates educational environments, it not only introduces students to cutting-edge technology but also fosters critical thinking, problem-solving, creativity, and teamwork skills essential for success in the modern world. Through hands-on experiences with robotics, students engage in active learning, gaining practical knowledge that prepares them for future careers in STEM fields while simultaneously shaping the future of education by integrating innovative teaching methodologies (Elayyan, 2021; Jafari et al., 2022; Kalaitzidou et al., 2023).

Tracing the Tapestry: Historical Perspective

The history of robotics spans from the introduction of the term "robot" in Karl Čapek's 1921

play, Rossum's Universal Robots, where robots rebel against humans, to Isaac Asimov's development of the Three Laws of Robotics in the 1930s, ensuring robots do not harm humans, and the popularization of robotics in pop culture, such as in Star Wars with iconic droids like R2-D2 and C-3PO (Capek, 2004; Stone, 2018; Trevelyan, 1999; Williams, 2019). The term robot is derived from the Slavic term robota, which means serfdom (Capek, 2004). The concept of serfdom's drudgery can be dated back a millennium to the tenth century, coinciding with the fragmentation of the Carolingian Empire, where tenant farmers were obligated to landlords for menial agricultural tasks (Bloch, 2023; Wolfram, 2019). The evolution of robotics in ancient civilizations were early automata and mechanical devices that laid the foundation for the development of sophisticated robots seen in contemporary industries such as manufacturing, healthcare, and education (Scaradozzi et al., 2020). Over the centuries, technological advancements, such as the Industrial Revolution (IR), sparked significant progress in automating labor-intensive tasks, leading to the emergence of programmable machines and the birth of modern robotics.

The evolution of robotics has been interdisciplinary in nature, resulting in significant progress made in integrating artificial intelligence into robotic systems (Doncieux et al., 2015). Artificial Intelligence (AI) enables machines to carry out activities that usually demand human intellect, such as decision-making, natural language understanding, and image recognition, automating tasks typically performed by humans (Soori et al., 2023). As robotics continues to evolve, driven by innovations in artificial intelligence, materials science, and engineering, it promises to revolutionize every aspect of human life, from enhancing productivity in industries to providing assistance in daily tasks and even education.

Modern AI robotics are not just theoretical constructs but are actively involved in practical tasks, serving as workers in various fields, with modern views placing AI robots in a negative light and associating them with apocalyptic scenarios leading to the termination of humanity (Odorčák & Bakošová, 2021). Despite these apprehensions, the evolution of robotics from ancient times to the present showcases a remarkable journey of innovation and technological advancement. From the rudimentary automata of ancient civilizations to the sophisticated AI-driven robots of today, humanity's quest to create intelligent machines has been fueled by both imagination and necessity. Ancient to contemporary advancements provide opportunities for robotics and AI technology to facilitate altruistic behavior between humans and robots, potentially leading to positive interactions and outcomes (Doncieux et al., 2015;

Odorčák & Bakošová, 2021; Soori et al., 2023).

Unraveling the Canvas: Industrial Perspective

Joseph Engelberger, recognized as the pioneer of robotics, drew inspiration from Asimov's writings and significantly contributed to the inception of industrial robotics, namely the Unimate, leading to the creation of a groundbreaking technology that revolutionized various industries (Stone, 2018). The evolution of industrial robotics is a testament to human ingenuity and the relentless pursuit of efficiency in manufacturing processes. Industrial robots have excelled in basic repetitive tasks common in assembly lines, although they may not match the general public's perception of robots, as the industry's growth factors, including labor shortage and substantial investments from industry and government, has resulted in the increased utilization of robots in manufacturing automobiles, electronic goods, semiconductors, and various applications such as product customization and flexible manufacturing systems for small volumes (Kumar, 2014). With each technological advancement, industrial robots have become increasingly versatile and capable of handling more complex tasks, blurring the lines between human and machine labor and reshaping the landscape of modern manufacturing. The evolution of robotics from Čapek's play to modern AI robotics includes advancements in technology and the integration of artificial intelligence, machine learning, and deep learning to create intelligent and versatile robots capable of complex tasks (Soori et al., 2023). Čapek's play was particularly influential in shaping global perceptions of robots and reinforcing a Frankenstein complex, with academic reviews and pop-cultural references highlighting the apocalyptic implications of robots potentially terminating mankind (Kumar, 2014; Odorčák & Bakošová, 2021; Stone, 2018). However, it is essential to recognize that these fears are not solely rooted in fiction; they stem from legitimate concerns about the ethical and societal implications of AI and autonomous systems. As technology continues to advance at an unprecedented pace, addressing these concerns becomes increasingly urgent to ensure that AI robotics are developed and deployed responsibly, with careful consideration given to their potential impact on humanity.

Embroidering the Narrative: AI Perspective

Machine Learning (ML) is a subset of AI that entails training algorithms on data to make predictions or decisions, and it finds applications such as enhancing real-time navigation

accuracy through learning from past experiences in ship navigation (Soori et al., 2023). Analyzing programming sequences with machine learning techniques to detect problem-solving patterns and pathways is a highly effective method for students to explore and learn robotics (Scaradozzi et al., 2020). Deep Learning (DL) is a form of ML that includes training artificial neural networks on extensive datasets to grasp intricate patterns and representations and is significant for robotics in tasks such as image and speech recognition, empowering robots to autonomously and intelligently execute diverse tasks (Soori et al., 2023). As AI, ML, and DL continue to evolve, their integration into robotics promises to revolutionize various industries by enabling robots to adapt and learn from their environments in real-time, leading to unprecedented levels of autonomy and efficiency.

Deep Learning (DL) has significantly contributed to robotics by enhancing tasks such as image and speech recognition, natural language processing, and object detection, enabling robots to autonomously perform a variety of tasks intelligently alongside AI and ML (Soori et al., 2023). In relation to DL, transhumanism is a concept that considers technological advancements as a mechanism for continued human evolutionary growth. Transhumanism aims to enhance individual capabilities through biotechnology for personal advancement, while posthumanism challenges existing power dynamics and suggests ethical and legal changes in human and non-human interactions, promoting cultural, scientific, and environmental methodological shifts (Odorčák & Bakošová, 2021). As AI, ML, and DL continue to evolve, societal, cultural, ethical, and legal concerns regarding their impact on human society become increasingly prominent. These concerns encompass issues such as privacy, job displacement, bias in algorithms, education and the ethical implications of AI decision-making, prompting debates and calls for regulation to ensure these technologies are developed and deployed responsibly for the benefit of humanity (Smakman et al., 2021; Sharkey & Sharkey, 2021; Zacharaki, 2020).

The future of robotics research aims to broaden applications beyond conventional definitions, exploring human-robot interaction and ergonomics to improve capabilities and tackle emerging challenges (Trevelyan, 1999). Future research avenues involve exploring the connections among various components in the design process and implementing evolutionary robotics in atypical robot types such as swarm robots, soft robots, and modular robots (Doncieux et al., 2015). Modern AI robotics are not merely theoretical concepts but are actively engaged in practical roles across different industries. Additionally, altruism is a

principle moral concept shared globally among human cultures that places great value in the well-being of others above one's self. Altruism, as a conceptual concern interwoven into the fabric of robotics evolution. While contemporary reporting and news articles often link robots to apocalyptic scenarios threatening human existence, the potential of AI technology possesses the ability to foster altruism between humans and robots, paving the way for beneficial interactions and results (Odorčák & Bakošová, 2021). Through interwoven efforts to harness the capabilities of AI and robotics for altruistic purposes, such as assisting the elderly, providing support in healthcare settings, or aiding in disaster relief efforts, humans and robots can form collaborative partnerships that lead to positive societal evolution and enhanced well-being for all.

Robotics and the Creative Elixir of STEAM Education Artful Engineering: Robotics and STEAM Exploration

Students at every grade level, ranging from elementary to university levels, can be educated regarding the challenges and consequences of AI in education and methods through inclusive and interdisciplinary strategies that incorporate AI principles into practice by promoting vital skills such as critical thinking, problem-solving, and teamwork necessary for adapting to the AI-centric future (Dignum, 2021). Integrating robotics into Science, Technology, Engineering, Art, and Mathematics (STEAM) education provides students with hands-on experiences that not only enhance their understanding of complex concepts but also foster creativity, collaboration, and resilience. STEAM education affords students with practical learning opportunities to unlock capabilities, improving problem-solving skills, and equipping students for the evolving requirements of the contemporary workforce (Foti, 2021). By engaging in robotics inquiry and hands-on learning, students gain invaluable insights into the application of STEAM in authentic, real-world scenarios, preparing learners to address challenges through innovative approaches in an increasingly interconnected global society.

Teachers require training to successfully integrate STEAM pedagogy and innovative teaching approaches (Foti, 2021). To achieve this standard, pre-service teachers need to develop, design, and research competencies to integrate innovative teaching methods successfully. Professional growth involves gaining skills for the active involvement of students and effectively supporting and implementing STEAM education (Anisimova et al., 2020).

Additionally, incorporating robotics into teacher training programs can provide educators with practical experience and knowledge of how to blend technology seamlessly into STEAM lessons.

Teachers can gain confidence in using robotics as a teaching tool and understand its potential for enhancing student learning experiences by engaging in robotics workshops and projects for continued development of instructional, pedagogical skills, as well as the evolution of technology skills. Professional growth training in both STEAM pedagogy and robotics better equips educators to deliver curricula, engage students, and inspire creativity for students engaging in robotics learning. Teachers need to be able to lead STEAM instruction that embraces both interdisciplinary learning and transdisciplinary. A STEAM construct blends science, technology, engineering, art, and mathematics in an interdisciplinary manner by linking each discipline across robotics as a content area. Transdisciplinary learning unifies robotics by interweaving collaboration between disciplines in a more unified manner. As a result, teaching and learning robotics with a STEAM philosophical approach enables each STEAM discipline within the acronym to be applied to robotics as separate disciplines (interdisciplinary) or holistically (transdisciplinary). The ability to employ both interdisciplinary and transdisciplinary learning for STEAM provides students with modes of modalities for critically thinking and creatively thinking (see Figure 1).
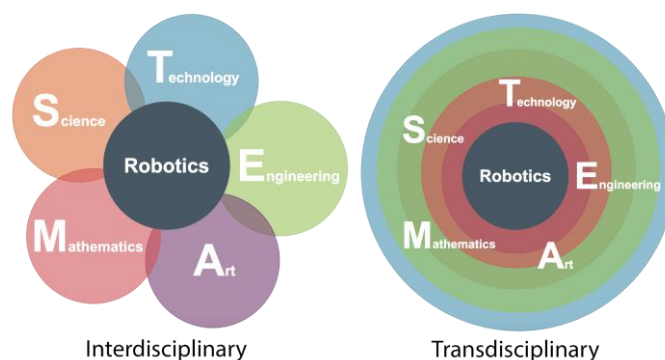

Figure 1. Interdisciplinary and Transdisciplinary STEAM Learning

Educational Alchemy: Early Childhood Education

Robotics can have a positive influence on children's social interactions and ability to work in teams, aspects of early childhood growth (Johnson, 2003). Children in early childhood educational settings possess the ability to learn coding and engineering concepts through

screen-free STEAM approaches that provide hands-on learning experiences. Introducing young children to technology and programming by engaging them in activities with robots such as Bee-Bot can foster an interest in STEAM fields from a young age and promote foundational skills in coding and engineering (Kalaitzidou & Pachidis, 2023). Moreover, robotics STEAM learning in early childhood can facilitate the development of essential skills such as problem-solving, critical thinking, and creativity. Through interactive experiences with robots, children are encouraged to explore, experiment, and collaborate, laying the groundwork for future academic success and lifelong learning (Johnson, 2003; Kalaitzidou & Pachidis, 2023). As robotics continues to play an increasingly significant role in society, early exposure to robotics in STEAM education sets children on a path towards becoming confident and competent participants in the digital age.

Girls and boys may have distinct approaches to robotics, suggesting a valuable potential for enhancing early childhood education for diverse learners through the use of robotics (Johnson, 2003). However, introducing STEAM education can both capitalize and bridge these differences and enrich the robotics learning journey for early childhood and kindergarten students (Anisimova et al., 2020). By integrating art in STEM for STEAM education, as recommended by Foti (2021), educational practices can better support young children, as STEAM is inherently interdisciplinary and transdisciplinary and better supports robotics teaching and learning. Through this interdisciplinary and transdisciplinary approach, children are encouraged to explore robotics in a holistic manner, incorporating elements of creativity, problem-solving, and collaboration. In addition, providing equal opportunities for both girls and boys to engage with robotics in early childhood, educators can foster an inclusive learning environment where children can develop essential skills and interests regardless of gender. Robotics STEAM learning in early childhood not only prepares children for future academic success but also instills a lifelong passion for inquiry and innovation at a very young age (Anisimova et al., 2020; Kalaitzidou & Pachidis, 2023).

Enlightening Instruction: Kindergarten and Primary Education

Introducing robotics education at an early age lays a crucial foundation for students' understanding of technology and engineering concepts. A variety of robots and robot kits exist for students in kindergarten through high school, affording age-appropriate opportunities for exploration and learning (Kalaitzidou & Pachidis, 2023). Kindergarten

students, in particular, can effectively grasp robotics concepts by engaging in practical hands-on learning tasks that focus on constructing and controlling robots through the use of robotics kits (Foti, 2021). One example is Beebot, a robot suitable for kindergarten and the first grades of primary school. As children progress through elementary school, they can further develop their robotics skills with kits such as mBot and Otto DIY +, which are tailored to their increasing abilities and comprehension levels (Kalaitzidou & Pachidis, 2023). Robots and robotics kits provide students with tangible experiences that not only enhance their understanding of STEM concepts but also foster creativity, problem-solving, and collaboration skills essential for success in the 21st century. By integrating robotics into kindergarten STEAM education, educators can inspire young learners to explore the possibilities of technology and cultivate a lifelong interest in STEAM-related fields.

Introducing robotics and STEAM education in kindergarten lays a strong foundation for students' future learning in STEAM fields. Blending elements such as algorithmic design, mechanical structures, and mathematical engineering concepts into kindergarten students' coursework through integrating cross-disciplinary STEAM supports student learning (Foti, 2021). As students progress into second grade, they continue to benefit from engaging in STEM and STEAM robotics projects. These projects often involve block programming platforms such as Scratch and LEGO WeDo, which allow students to explore computational thinking and educational robotics concepts in a hands-on and interactive manner (Valls Pou et al., 2022). By emphasizing project-based learning approaches and encouraging collaboration among peers, educators can effectively introduce young learners to the fundamentals of robotics while fostering important skills such as problem-solving and teamwork. Through these early experiences with robotics and STEAM, students develop a strong foundation in STEAM disciplines and are better prepared to tackle future challenges in an increasingly technology-driven world.

Crafting Wisdom: Primary and Secondary Education

Kalaitzidou & Pachidis (2023) underscore the significance of integrating the STEAM philosophy into educational activities at all grade levels, with a particular emphasis on the role of robotics in enriching learning across various disciplines. By incorporating robotics into STEAM education, students are provided with opportunities for interdisciplinary and transdisciplinary learning experiences, fostering the development of computational thinking

skills essential for innovative, creative, problem-solving. Valls Pou et al. (2022), suggests employing strategies that involve utilizing sequential thought processes rooted in fundamental computer science concepts to tackle challenges, devise systems, and understand human behavior. Employing computational thinking in concert with hands-on robotics learning provides students with opportunities to develop and cultivate a deeper understanding of how technology interconnects with multiple disciplines.

High school students preparing for physics exams often find that participating in robotics activities offers tangible real-world applications for the principles they are studying (Johnson, 2003). Robotics kits such as Lego Mindstorms EV3 are particularly well-suited for high school students, providing them with a platform to engage in hands-on learning experiences (Kalaitzidou & Pachidis, 2023). As a result of involving upper elementary and high school students in STEAM robotics projects, educators can integrate programming, computational thinking, and educational robotics concepts into their curriculum through immersive, project-based learning activities (Valls Pou et al., 2022). Engagement and involvement approaches not only enhance students' understanding of STEAM subjects but also fosters the development of critical thinking, problem-solving, and collaboration skills essential for success in both academic and professional pursuits. Through robotics projects, students can see firsthand how the theories they learn in the classroom translate into practical applications, making their learning experiences more meaningful and engaging.

Elevating Pedagogy: Higher Education and Teacher Training

Pre-service teachers need to develop design and research competencies to successfully integrate innovative teaching methods to develop skills for the active involvement of students and for effectively supporting and implementing STEAM education (Anisimova et al., 2020). Enhancing teacher training programs and offering resources for practical learning empowers educators in fostering students' potential in STEM and STEAM education and need to evolve their roles to function as researchers, continuously updating their knowledge to align with contemporary educational methodologies (Foti, 2021).

Schools of higher education prepare future educators by fostering the development of skills for integrating STEAM philosophical and pedagogical constructs for supporting innovative teaching and learning. STEAM fosters innovative teaching approaches that emphasize

computational thinking and STEM education, addressing the necessity for a modern and responsive educational framework (Foti, 2021). The integration of STEAM and computational thinking is particularly important in robotics for hands-on, project-based learning experiences that can greatly enhance student engagement, understanding, and achievement. Pre-service teacher training must emphasize not only content knowledge but also pedagogical strategies tailored to the interdisciplinary and transdisciplinary nature of STEAM education. Universities and colleges need to equip future educators with the knowledge and skills necessary to successfully incorporate project-based learning and STEAM education methods into various disciplines and educational levels (Anisimova et al., 2020). Providing pre-service teachers with the tools to effectively implement STEAM principles supports student learning at all grade levels by providing learners with opportunities to develop and employ problem-solving, critical thinking, computational thinking, and creative thinking skills.

Employing STEAM interdisciplinary learning and transdisciplinary learning provides multiple approaches for blending content and fosters a comprehensive understanding of robotics uses and applications for innovative design and problem-solving. In addition, a robotics-STEAM construct enriches early childhood through higher education settings and affords students with multiple modalities of learning and addressing robotics content in the context of STEAM. The blending of content also provides multiple pathways as modalities for learners to apply knowledge regarding each discipline for engaging in critical thinking and creative thinking (see Figure 2).
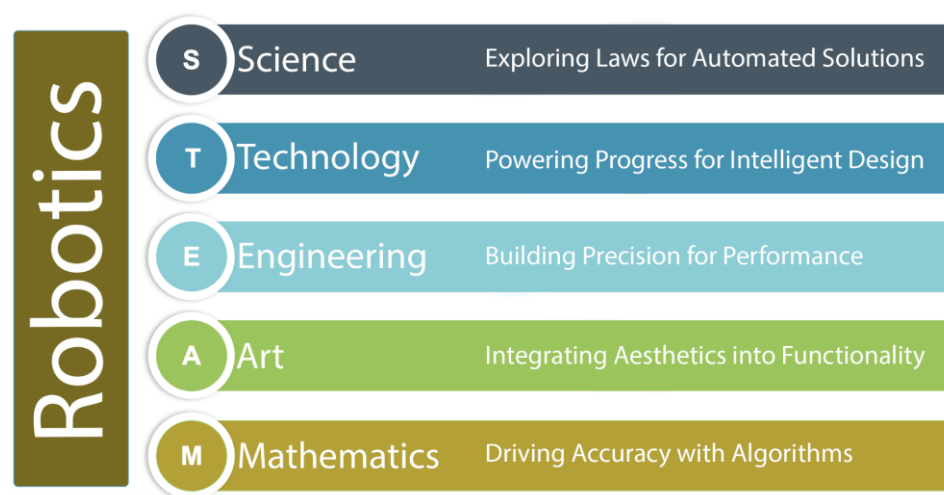

Figure 2. The Modalities of STEAM and Robotics

Cultural Evolution and Revolution

Echoes of Tomorrow: AI's Societal Ripples

As robotics, artificial intelligence, and automation technologies continue to advance, it becomes imperative to educate individuals and society about their implications through multidisciplinary approaches. Educating individuals and society about AI implications through multidisciplinary approaches ensures informed decision-making and ethical use of AI technologies (Dignum, 2021). Beyond mere technological advancements, AI innovations are reshaping the job market, altering the work landscape, and influencing the trajectory of the developing economy in both current and future scenarios (Webster & Ivanov, 2020). As robots increasingly perform tasks traditionally carried out by humans, society faces significant shifts in employment patterns and skill requirements. Robotics, AI, and automation technologies are transforming the job market, changing the work landscape, and impacting the developing economy in current and future scenarios (Webster & Ivanov, 2020).

Artificial Intelligence (AI) is a socio-technical entity that underscores the necessity for responsible development guided by principles of accountability and responsibility, with an emphasis on educating individuals and society about AI implications through multidisciplinary approaches. Webster & Ivanov (2020) note that robotics, artificial intelligence, and automation technologies are shaping the economy by changing how tasks are executed and affecting the future direction of industries and job prospects. Among these transformations, societal concerns include the possible replacement of human workers by automation, the necessity for acquiring new skills to stay competitive in the job market, the impacts of transitioning to short-term and gig-oriented employment, and the difficulties in adjusting taxation systems and social frameworks to align with the evolving economic environment (Webster & Ivanov, 2020). Moreover, a need exists for individuals to acquire new skills to remain competitive in the evolving job market landscape as industries shift towards a global job market increasingly influenced and possibly dependent upon AI and robotics technologies. As a result, education, robotics, AI, and industrial technologies must coevolve if they are to coexist harmoniously in the workforce.

Evolution's Overture: IR 1.0 to IR 30

The Industrial Revolutions (IRs) have resulted in shifting and shaping human history and

technological progress (Groumpos, 2021; Stearns, 2020). These revolutions have fundamentally altered societal structures and economic systems, marking significant milestones in human development. The First Industrial Revolution (IR 1.0) began in the late 19th century, introducing mechanical production facilities powered by water and steam (Hahn, 2020).

The Second Industrial Revolution (IR 2.0) emerged in the early 20th century, emphasizing mass production assembly lines using labor and electrical energy. In the 1970s, the Third Industrial Revolution (IR 3.0) came about, focusing on automated production through electronics and information technology (Elayyan, 2021; Tinmaz, 2020). Each phase revolutionized society and brought about changes not only in the way goods were produced but also in how people organized themselves and interacted with technology. The Industrial Revolutions (IRs) catalyzed rapid advancements in various fields, driving innovation and progress while also presenting new challenges and opportunities for individuals and societal groups.

Evolution's Symphony: IR 4.0

The Industrial Revolutions (IRs) have played a pivotal role in defining societal development. Currently, the Fourth Industrial Revolution (IR 4.0) is centered on interactions between humans and machines, impacting learning opportunities and instructional activities in education (Elayyan, 2021). The Fourth Industrial Revolution (IR 4.0) has resulted in a transformation of work and activities that have historically been performed by humans through cognitive technologies such as machine learning and robotic process automation (Mhlanga, 2022; Rotatori et al., 2021).

As IR 4.0 continues to unfold, it highlights the imperative for individuals to adapt and acquire new skills to remain competitive in the evolving job market (Tinmaz, 2020). This revolution underscores the necessity for significant reforms in education to adjust to evolving technologies and the technological landscape, ensuring that students are equipped with the skills and knowledge needed to thrive in changing global markets (Rotatori et al., 2021). Embracing the lessons of past industrial revolutions and IR 4.0, positions society to better navigate the challenges and opportunities presented by IR 4.0 and the evolving IR 5.0 in a future where humans and machines work together to drive progress and innovation.

Evolution's Crescendo: Industry 4.0 and IR 5.0

The evolution of robotics, AI, and self-awareness technologies have been and will continue to be highly influential in terms of the evolution of IR 4.0 to IR 5.0. The Fourth Industrial Revolution (IR 4.0), also known as Industry 4.0, began in 2011 and focuses on the IoT (Internet of Things) and CPS (Cyber-Physical Systems), ushering in an era of interconnected devices and integrated digital technologies (Jafari et al., 2022). Like previous revolutions, IR 4.0 (Industry 4.0) has impacted the workforce and greatly influenced a paradigm shift in terms of how individuals approach skill development and the workforce. As Industry 4.0 continues to unfold, the importance of reskilling and upskilling becomes increasingly apparent, ensuring that individuals can adapt to the demands of the evolving technological landscape and remain competitive (Mhlanga, 2022; Rotatori et al., 2021; Tinmaz, 2020). Furthermore, IR 4.0 highlights the necessity for significant reforms in education to align with the changing needs of the workforce and the global economy.

The Fifth Industrial Revolution (IR 5.0) is also known as Industry 5.0 and has recently been proposed as a concurrent revolution to Industry 4.0 and focuses on human-centricity, resilience, and sustainability in manufacturing processes, emphasizing the collaboration between humans and new technologies to achieve sustainable development goals (Jafari et al., 2022). The rapid development of technologies such as robotics, AI, and digital and physical technologies has also resulted in a rapid advancement from IR 4.0 to IR 5.0. The Fifth Industrial Revolution (IR 5.0) seeks to integrate digital, physical, and biological technologies to improve societal well-being and prioritize sustainable development objectives for ensuring human-intelligent machine interactions (Noble et al., 2022).

The transition from Industry 4.0 to Industry 5.0 marks an evolutionary phase of societal evolution that encompasses the integration of robotics and AI into various facets of culture and education. Industry 5.0 shifts the emphasis from purely technology-driven processes to human-centric and sustainable practices (Jafari et al., 2022; Tinmaz, 2020). Industry 4.0 to Industry 5.0 transitions aim to address not only job security but also societal development and environmental concerns by fostering collaboration between humans and new technologies such as robotics and AI. In particular, the collaboration between humans and technology becomes a central focus, emphasizing the potential of human-robotics partnerships to drive sustainable development goals and enhance societal well-being (Noble et al., 2022).

Additionally, IR 5.0 seeks to integrate digital and physical technological evolutions taking place during a time of rapid change and innovation that emphasizes human-technology collaboration for societal well-being (Jafari et al., 2022; Noble et al., 2022; Tinmaz, 2020). The interactions of humans with digital, physical, AI and robotics is a complex, collaborative process that requires society to evolve as quickly as the evolution of emergent technologies (see Figure 3).
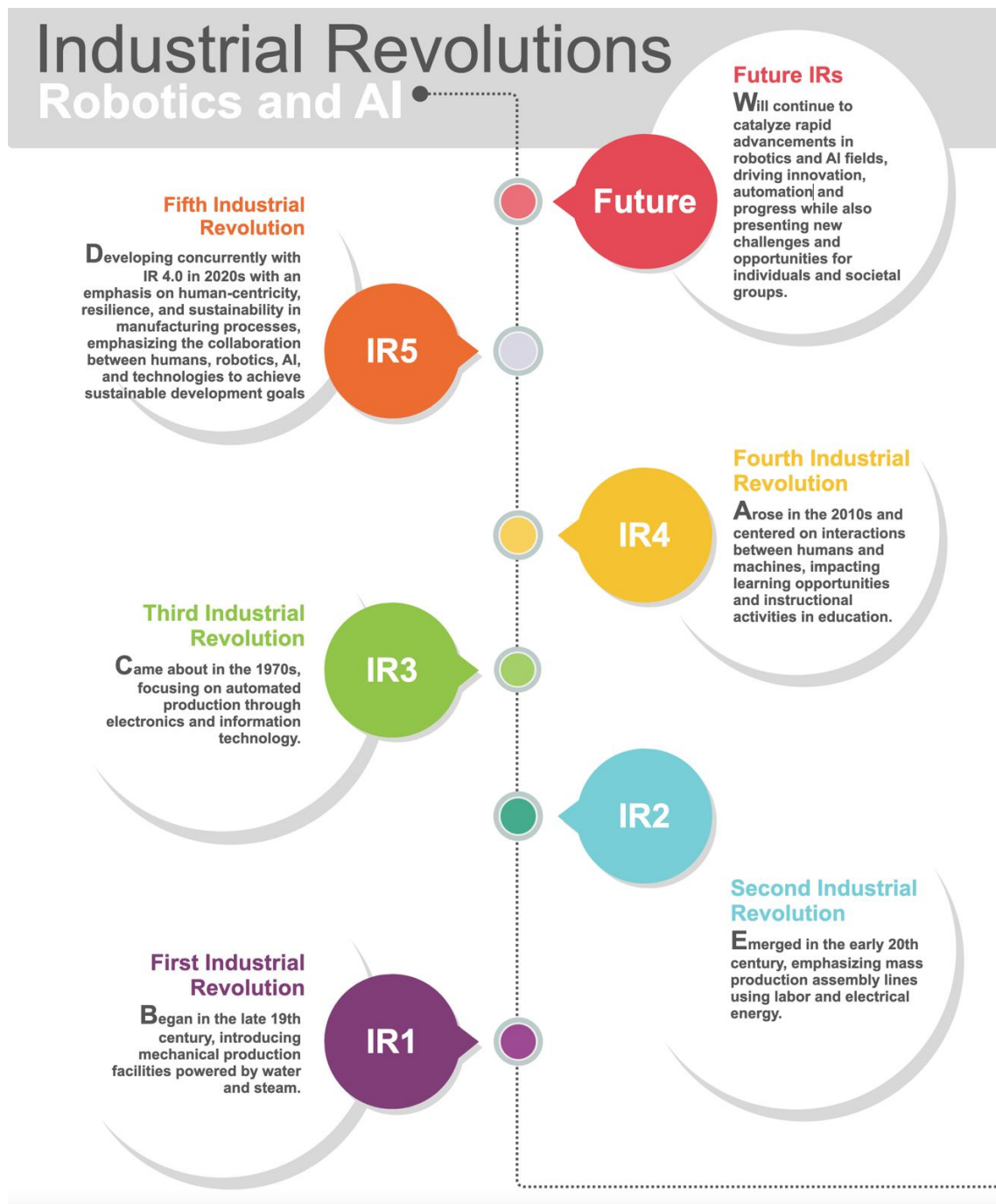
Figure 3. Industrial Revolutions and Pathways of Robotics and AI

The Algorithmic Nature of AI Robotics

Metamorphosis: From Robota to Automata

While the drudgery of serfdom can be traced back one thousand years to the Carolingian Empire during the tenth century followed by contemporary fictional robotic storytelling by Karel Čapek, Isaac Asimov, and George Lucas, the shift from drudgery during the First Industrial Revolution to the current IR 5.0 has resulted in self-learning, automated, AI robots. The logarithmic evolution of robotics, beginning with mid-twentieth century toiling industrial robots to early 20th century, self-thinking AI robotics, has resulted in robots that are highly intelligent, self-learning, and unlike Karel Čapek's automaton serfs.

Machinery's Mind: The Convergence of Intellect

Challenges in emergent AI autonomy for robots involve managing uncertainty, particularly in perception and decision-making, due to the inherent limitations in robots' knowledge, as well as the complexity of constructing precise and comprehensive environmental models, essential for efficient planning and control (Dzedzickis et al., 2021; Thrun, 2000). Algorithms are rule-based processes that guide operations to achieve goals or solve problems, particularly in AI and ML where they learn patterns, predict outcomes, and optimize functions iteratively. Learned patterns are fundamental in guiding the behavior of AI systems and enhancing their capabilities in various applications, including robotics. Convolutional Neural Networks (CNNs), are a type of DL algorithm commonly used in machine learning applications for mobile robots and are specifically designed to process visual data, such as images or videos for DL (Cebollada et al., 2021).

CNN learning algorithms with multiple layers (filters) utilize abstraction to extrapolate data for categorizing objects in images, leading to enhanced accuracy and precision rates during the detection of defective and non-defective objects for ML and DL (Dzedzickis et al., 2021). CNN algorithms influence autonomous systems for ML and DL systems. Neural network algorithmic programming influences AI robotic autonomy and DL, allowing for systems to respond to external stimuli for enhancing behavioral capability (Zenil, 2014). As AI continues to evolve, the integration of these advanced algorithms into robotic systems play a central role in adaptive and intelligent machinery. Employing algorithmic probabilities allows robots to make well-informed choices, anticipate forthcoming uncertainties, and function

autonomously, ultimately enhancing AI robotics efficiency and dependability (Dzedzickis et al., 2021; Thrun, 2000). CNN capabilities enhance the performance and safety of autonomous systems and enable systems for operating in increasingly complex environments.

Neural networks and deep learning tools are extensively utilized in machine learning applications and are tasked with constructing mathematical models from sample data to enable predictions and decisions without explicit programming (Cebollada et al., 2021). Neural networks are capable of analyzing large datasets for finding patterns and features that are not readily apparent for improving the accuracy of predictions in ML fields for DL. The correspondence between input stimuli and behavioral space in robots leads to more complex patterns, indicating adaptability and sensitivity to the environment, thus increasing algorithmic complexity for self-learning (Zenil, 2014). Neural network and ML adaptability allow for adjusting to dynamic conditions and making decisions in real-time. AI algorithms assist robots in interpreting human gestures, speech, and behavior, facilitating effective communications and interactions for understanding and responding to human actions and intentions (Dzedzickis et al., 2021; Thrun, 2000). The capabilities of neural networks not only enhances the human-robot interaction but also paves the way for robots to become more integrated into daily human activities, enhancing AI robotic utility and functionality.

Robotic Self-Discovery: Unveiling Learning Mechanisms

Neuromorphic computing hardware is an AI technology that emulates the workings of biological neural systems, providing energy-efficient AI based on neural networks and autonomy for AI, ML, and DL. Neural architectures for robotics entail creating network structures and learning mechanisms that enable neuromorphic devices to tackle practical tasks in robotics, akin to the functionality of animal brains and bodies (Sandamirskaya et al., 2022). Reinforcement Learning (RL) is a framework derived from early work in neuroscience that involves learning through interactions with an environment to maximize cumulative rewards and enables self-learning in robots through algorithms that compute and adapt to complex tasks autonomously (Singh et al., 2022). The self-learning capabilities of RL is enhanced by the integration of neuromorphic computing hardware, which supports continuous learning and adaptation. Neuromorphic computing hardware and neural architectures empower robotics to achieve self-learning by enabling continuous adaptation and autonomous improvement based on real-time sensory data and changing environmental

conditions for RL (Sandamirskaya et al., 2022).

Cognitive Intelligence Algorithms (CIAs) replicate human cognitive processes such as reasoning and decision-making, drawing from cognitive psychology, neuroscience, and social behavior to enable machines to understand complex decisions and make value-based judgments (Ren et al., 2023). Cognitive Intelligence Algorithms (CIAs) provide a foundation for Spiking Neural Networks (SNNs) to build upon by utilizing a method of information transmission similar to the human brain's neuronal activity. Spiking Neural Networks (SNNs) are brain-inspired computational models that use spikes to transmit information, enabling them to adapt neural connections to changes and stimuli, and enhance robotics self-awareness by fostering autonomous learning, adaptation, and decision-making (Lobov et al., 2020). The adaptation and learning mechanisms of SNNs complement the principles of RL, which applies a structured approach to problem-solving and skill acquisition through direct interactions with the environment. Reinforcement Learning (RL), in conjunction with CIAs and SNNs inspired by neuroscience, enables robots to self-learn and adapt to complex tasks through environment interaction and trial-and-error decision-making, thereby improving their performance over time (Singh et al., 2022).

The integration of the Internet of Things (IoT) and Fog Cloud Computing enables immediate data acquisition and processing for the decision-making abilities of AI-driven autonomous robots. The Internet of Things (IoT) is a network of devices, sensors, and materials with software, sensors, and connectivity that share data and enable intelligent actions for collecting real-time data from the environment to assist robots in interacting and making decisions (Singh & Singh, 2024). Fog Cloud Computing is a decentralized model with a three-layer infrastructure of edge sensors, fog processing, and cloud data centers (Bhambri et al., 2022; Firouzi et al., 2022). Proximal data processing reduces latency and conserves bandwidth, enhancing real-time data processing and decision-making in AI-enabled autonomous robotic systems. (Bhambri et al., 2022; Singh & Singh, 2024).

A Quest for Robotic Sentience: A Journey within Conscious Machinery

Self-awareness in robotics is the dynamic interplay between artificial entities and their physical and social surroundings, with these interactions leading to the creation of computational frameworks that underlie cognitive self-processing in robots (Mentzou &

Ross, 2023). Emerging understandings of self-awareness and cognitive self-processing paves the way for the integration of metacognitive processes and insights, which are integral to the further development of robotic cognitive capabilities. Metacognition and insight are essential components in self-reflection, with metacognition referring to an awareness of thoughts and insight involving the ability to identify and express emotions, which are elements for deepening self-knowledge and facilitating robotic higher-order self-awareness (Chella et al., 2020; Zeng et al., 2020).

The Elusive Horizon: The Search for Robotic Consciousness

The AI "singularity" is a theoretical future scenario where artificial intelligence exceeds human capabilities, potentially enabling AI to autonomously enhance itself and drive rapid, uncontrollable technological progress. While this phenomenon suggests a leap in AI capabilities, it ironically underscores a fundamental limitation. As a consequence, the singularity of AI results in computers lacking self-awareness and subjective identity, preventing AI from evolving into a distinct species or fully replicating human consciousness, with this limitation impacting the future scope and human interactions with AI (Wang, 2023). The inherent constraint of the AI singularity is a significant challenge with respect to the current futility of AI attaining a form of consciousness similar to human experience.

Artificial Intelligence (AI) lacks consciousness, as machines and robots do not possess the self-awareness, perception, and emotions required for mindfulness. Intelligence can operate independently of consciousness, which highlights the distinction between both concepts. Self-awareness, a characteristic specific to human cognition, is the conscious acknowledgment of a subjective-self, a process far beyond the capabilities of current AI systems. As Wang (2023) notes, self-learning pertains to the automated acquisition of knowledge or skills within AI systems, distinct from the introspective and conscious processes of self-awareness. Furthermore, Zeng et al. (2020) explain that self-awareness involves advanced cognitive processes such as introspection and recognition of self-existence, which are beyond the realm of robotics, AI, ML, and DL's capabilities in active vision and view planning algorithms. Consequently, as Li et al. (2021) asserts, the current state of AI technology, with its inherent limitations in consciousness and self-awareness, does not support the notion of AI becoming a threat to humans.

The Accord of Beings: Exploring Human-Robot Interactions

Robot aesthetics is a major factor impacting trust within human-robot interactions (Kim, 2022; Pinney et al., 2022). Successful human-robot interactions depend on the alignment and responsiveness of nonverbal cues, such as body language, facial expressions, and vocal tones, which are essential for social robots to establish and maintain strong communal connections and perceived rapport, thereby enhancing cooperation with humans (Breazeal et al., 2016; Kim et al., 2020). In research conducted by Kim, 2022, trustworthiness in human-robot interaction is impacted by elements such as capability, honesty, and kindness. Human-robot interactions are influenced by people perceiving kindness in robots and are essential components of trust in human-robot relationships. Research indicated that honesty may gain more importance over time with continued interaction with robots, suggesting its potential influence on long-term trust.

In research conducted by Pinney et al. (2022), design features such as the visual components of facial screen design affect the level of trust humans have in robots and can influence how trustworthy robots appear to individuals. These design features suggest aesthetics and interface designs influence initial trust-building, facilitating user comfort during interactions with robots. Moreover, in the context of robot groups, entitativity (the extent to which a group is perceived as a coherent entity rather than just an aggregate of individual members) influences how humans perceive and interact with the group, impacting their responses and behaviors towards robots. Understanding the concept of entitativity in human-robot interactions illuminates how individuals perceive and interact with groups of robots and enables researchers and designers to devise approaches to enhance human-robot relationships (Dang & Liu, 2023).

Positive affect impacts dynamics of trust within Human-Robot Interaction (HRI). According to Kim et al. (2020), emotions significantly influence trust during interactions during HRI, necessitating the importance of incorporating emotional dynamics into the design and functionality of robotic systems. Aspects of trust emphasize the need for robots to not only perform tasks efficiently but to also engage with humans in a manner that is perceived as emotionally intelligent and responsive. Understanding and integrating emotional factors via the acronyms of STEAM education through HRI assists developers and researchers to better tailor robotic systems to foster deeper and more trusting human-robot relationships,

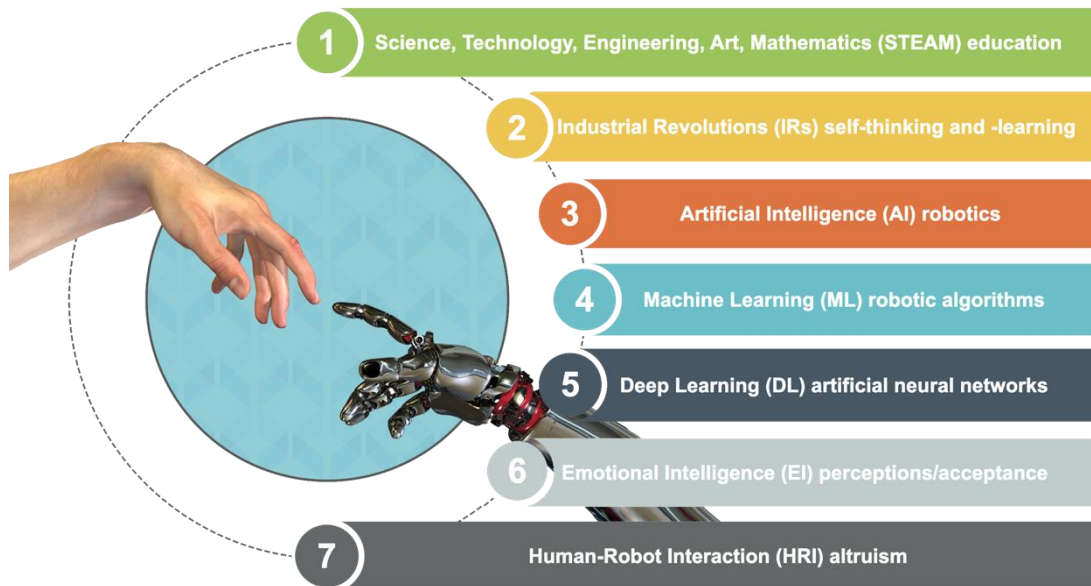enhancing the overall effectiveness of HRI models (see Figure 4).

1	Science, Technology, Engineering, Art, Mathematics (STEAM) education
2	Industrial Revolutions (IRs) self-thinking and -learning
3	Artificial Intelligence (AI) robotics
4	Machine Learning (ML) robotic algorithms
5	Deep Learning (DL) artificial neural networks
6	Emotional Intelligence (EI) perceptions/acceptance
7	Human-Robot Interaction (HRI) altruism

Figure 4. Beyond the Acronyms of Robotics and Humans

Mechanical Fusion

Gearwork and AI: Strategies for Robotics in Education

The evolution of robotics is moving towards personalized learning experiences through the use of social educational robots, emphasizing individualized and tailored learning. In research by Johal (2020), studies indicate a transition to one-to-one setups, where robots adjust learning content and social behavior. Incorporating robots in educational environments includes providing personalized learning experiences customized for individual students to adapt content and social behavior to address students' specific requirements (Johal, 2020). In early childhood education through higher education settings, the integration of robotics needs to consider utilizing analytics to convey emotions for emotional connections and socialization. While technology is still evolving, some benefits and practical approaches for integrating social robots into education have been observed.

Tiny Sprockets: Early Childhood Education

Social robots have been utilized extensively in early childhood education, specifically with children who have autism or for learning second languages. Early childhood robots assist

young children with autism in comprehending suitable physical boundaries during social interactions and potentially enhance reading abilities, grammar skills, and emotional aspects of learning in language acquisition (Woo et al., 2021). In a meta-analysis performed by Alam (2022), research on robot tutoring indicated positive effects of early childhood and kindergarten learners, including positive effects on cognitive and emotional learning outcomes when utilizing robots as a teaching aid in educational environments. Moreover, early childhood education in robotics emphasizes hands-on learning through engaging activities such as the Lego Education Coding Express, focusing on teaching coding concepts such as sequencing and looping using action bricks (Evripidou et al., 2020).

Mini Pistons: Kindergarten

In research conducted by Woo et al. (2021), social robots have been shown to positively impact learning outcomes, social interaction, and young learners' concentration. Robots positively enhanced the initial motivation of children and early adolescents to engage with subjects such as language learning, math, and science. Social robots were effective in enhancing both concentration and social interaction among students (Woo et al., 2021). Research indicated that social robots could serve as effective tools for educators to individualize and support learning experiences for elementary school students. Employing robots as educational tools in preschool and kindergarten settings can positively influence cognitive and emotional learning outcomes in young children (Alam, 2022). Preschoolers, for example, can build and program a train using physical sets and a compatible application, enhancing problem-solving skills and computational thinking at a beginner level (Evripidou et al., 2020). Overall, these studies collectively illustrate the transformative potential of social robots in educational settings, from enhancing engagement and concentration to fostering cognitive and social skills for the integration of robotics with young children.

Elemental Springs: Primary Education

The utilization of social robots in elementary education environments has the potential to positively influence different facets of students' academic progress and social growth and improve academic results, heighten peer engagement, and enhance concentration levels during educational tasks (Woo et al., 2021). The use of robots as teaching aids for students in elementary grades in elementary school settings can positively impact and enhance language

skills and academic performance for learners. In a study conducted by Alam (2022), robots were shown to successfully improve language skills and academic performance in elementary school students by promoting extended focus on learning tasks, quicker responses, and increased accuracy compared to conventional approaches. The implementation of social robots as peer tutors in elementary grades led to favorable results, demonstrating enhancements in students' English language abilities (Alam, 2022). Primary education robotics introduces students to basic programming, directional language, and mapping skills through educational floor robots such as Bee-Bot and Blue-Bot. Students in primary school can develop their programming skills further with more complex structures using block-based programming languages such as EdPy (Python-like text-based programming language for advanced programmers using educational robots) and Scratch and text-based languages (Evripidou et al., 2020).

High Levers: Secondary Education

In secondary education, students can advance their programming skills by familiarizing themselves with professional textual programming languages to prepare for more complex coding tasks. Lego Education offers robotic kits tailored for secondary school students to enhance their problem-solving abilities and computational thinking through hands-on programming experiences (Evripidou et al., 2020). Furthermore, by integrating robotic builds into high school makerspace environments, computer science, and art, students are provided access to a variety of digital and physical fabrication tools for devising design and fabrication solutions. Hands-on interdisciplinary and transdisciplinary secondary school learning experiences and collaboration nurture innovation and creativity by facilitating design, prototyping, and testing that align with the creative dimensions of STEAM education (Kim, 2021). These cross-disciplinary approaches not only enrich students' learning experiences but also encourages them to apply their programming by collaborating on projects that blend aesthetics with functional analytics for developing an understanding with respect to how technology can be used in diverse and creative contexts.

Complex Pulleys: Higher Education

Higher education focuses on strengthening educational innovation through robotics, integrating robotics into various disciplines to enhance learning outcomes and prepare

students for STEAM-related careers (Evripidou et al., 2020). Opportunities for strengthening innovation and experiential knowledge includes team structured, competitive events for robotics builders to test and demonstrate the functional designs of their robots. Robotics competitions and events in higher education provide opportunities for students to showcase their skills, collaborate, and apply theoretical knowledge to practical challenges, fostering a deeper understanding of robotics principles (Evripidou et al., 2020).

Table 1. Robotics and Materials by Education Level

Education Level	Examples of Robotics and Manipulatives
Early Childhood Education	Floor robots, Fisher-Price Think & Learn Code-a-Pillar, Lego Education Coding Express, TTS Bee-Bot, TTS Blue-Bot
Kindergarten	TTS Bee-Bot, TTS Blue-Bot, Kits such as mBot and Otto DIY +, LEGO DUPLO Coding Express, Tactile robots, Programmable blocks from Resick's Lifelong Kindergarten lab
Primary Education	LEGO WeDo 2.0, Electronics kits (Little Bits), Ozobot, Wonder Workshop Dash robot, TTS Bee-Bot, TTS Blue-Bot, EdPy, Scratch, computers
Secondary Education	Lego Mindstorms EV3, Lego Education robotic kits, Arduino-based robots, Raspberry Pi robots, Advanced robotics kits, Computer science (CS) coding, Textual programming languages, CS and makerspace art design, Prototyping and testing software/hardware, Digital and physical fabrication tools
Higher Education	Sophisticated robotics and AI projects, advanced humanoid robots and robotics arms, Drones and autonomous vehicles, programming in C++, ROS (Robot Operating System), VR development tools

Robotics usages in higher education also includes integrating robotics technology for teaching and learning assistance. In a study by Kim et al. (2020) on AI robotics teaching assistants in higher education, the researchers employed a Technology Acceptance Model

(TAM), which is a framework used to explain how individuals accept and use various technologies. The utilization of TAM indicated that the adoption of technologies is influenced by an individual's behavioral intentions to use a particular technology, with perceived usefulness and perceived ease of use playing key roles in the process. The research suggests that social robots, such as AI teaching assistants, can have a positive impact on student learning outcomes (Kim et al., 2020). The findings provide guidance for the successful implementation of AI technology in educational environments for future research to improve the acceptance and utilization of AI in education. Robotics manipulatives and materials vary in design and complexity for various education levels with age-appropriateness being considered as a doctor for employing at each education level (see Table 1).

STEAM Sentience and AEI: Crafting Robotic Emotions

Artificial Emotional Intelligence (AEI) in robots enables robots to mimic empathy, respond appropriately to users' emotions, adapt their behavior to build relationships over time, enhancing their appeal, credibility, and interaction quality, ultimately improving human user acceptance of robotic technology (Marcos-Pablos et al., 2022). Artificial Emotional Intelligence (AEI) in socially assistive robots supports individuals by improving user satisfaction and interaction through empathetic actions (Abdollahi et al., 2022). In a study by Law et al. (2021) the researchers noted that demonstrating empathy is critical in the emotional intelligence of robots, impacting how humans perceive robots and the level of trust they inspire. Displaying empathy illustrated robotic AEI, shaping human perceptions and trust levels towards robots. As a result of displaying empathy, robots can establish meaningful connections with humans, fostering positive interactions, particularly in environments where robots collaborate closely with people (Law et al. (2021).

In a study by Abdollahi et al. (2022), integrating facial expression recognition and sentiment analysis resulted in AEI robots being able to accurately detect user emotions, which enabled the emotion AI detector system to create empathetic responses customized to the user's emotional state, promoting a more interactive and individualized conversation. The synchronization of monitoring human emotions and using AEI enhances a robot's appeal, credibility, and user experience, fostering greater acceptance of robotic technology (Marcos-Pablos et al., 2022). Ensuring alignment between behavior and appearance enhances the

robots' believability, likability, and overall user experience, ultimately leads to greater acceptance and engagement with robotic technology.

As a result of identifying, interpreting, replicating, and responding to human emotions, socially assistive robots equipped with AEI can engage users in a more authentic and impactful manner, contributing positively to their emotional wellness (Abdollahi et al., 2022). When robots demonstrate emotional intelligence, such as empathy and sensitivity to human emotions in healthcare and education, individuals view them as more approachable and reliable, resulting in increased acceptance and collaboration between humans and robots across different environments (Abdollahi et al., 2022; Law et al., 2021). Moreover, the ability of AEIs to identify, interpret, replicate, and respond to human emotions provides pathways for integrating AI robotics technologies into early childhood to higher education settings for embracing and supporting teaching and learning.

Conclusion

Robotics is a developing field of education and industry within computing, engineering, and technology, blended and connected to principles of design, the sciences, mathematics, and the arts. Originating from Karel Čapek's 1921 play of robotic serfs to today's AI robotics, represents a logarithmic evolution of not only technology and interdisciplinary and transdisciplinary STEAM education, but a societal shift in the ways robots are perceived, valued, and likely welcomed. Transformational shifts from mere mechanical labor automatons to complex, intelligent, and self-learning autonomous systems highlights the capability of robotics to transcend its origins and drive educational innovations.

In early childhood and primary education, robotics serves as a gateway for young learners to grasp coding and engineering principles. Through hands-on, screen-free STEAM activities, children engage with technology from a young age, gaining foundational skills in problem-solving, creativity, and collaboration. collaborative, inquiry-based skills are further refined as students' progress, with robot kits tailored to each educational stage providing deeper insights into STEAM disciplines and enhancing students' practical and cognitive abilities.

Secondary education robotics further embeds and blends curricula through innovative design and experiential learning. High school students explore more complex concepts in

programming and computational thinking, with opportunities to utilize robotics in makerspaces that combine computer science and engineering with artistic design. The interdisciplinary and transdisciplinary nature of robotics STEAM learning not affords students with understanding of the technical aspects of robotic design and builds, but also fosters critical thinking and innovation, preparing learners for the challenges of higher education and post-secondary endeavors. Integrating STEAM and computational thinking into higher education robotics enriches students' education experiences and provides students with critical-thinking and creative thinking skills for competing in a technologically dynamic and interconnected world.

The algorithmic basis of AI, incorporating humanistic philosophy, machine learning, and advancements in Artificial Emotional Intelligence (AEI), allows robots to simulate empathy, respond accurately to human emotions, and adapt behavior to foster relationships. Artificial Emotional Intelligence provides an algorithmic approach to soliciting and facilitating trustworthiness and enhancing user acceptance. Employing AI empathy also aids in forging human connections, promoting positive interactions and collaboration.

While robotic algorithms possess exciting potential and abilities for self-learning, the AI singularity represents a fundamental limitation in robotics, with the absence of self-awareness preventing AI from evolving into a distinct species or fully mirroring human consciousness. While the humanistic evolution of robotics and the pursuit of sentience is fleeting, the ability of AEIs provides robots with abilities to discern, interpret, and mirror human emotions. The integration of AEI robotics, with appropriate educational manipulatives, devices, and curricular objectives from early childhood through higher education enhances and increases the acceptance of robotics for supporting educational experiences and outcomes and preparing all students for 21st learning.

References

Abdollahi, H., Mahoor, M., Zandie, R., Sewierski, J., & Qualls, S. (2022). Artificial emotional intelligence in socially assistive robots for older adults: A pilot study. *IEEE Transactions on Affective Computing*. https://doi.org/10.48550/arXiv.2201.11167

Alam, A. (2022). Social robots in education for long-term human-robot interaction: Socially supportive behaviour of robotic tutor for creating robo-tangible learning environment

in a guided discovery learning interaction. *ECS Transactions*, *107*(1), 12389. https://doi.org/10.1149/10701.12389ecst

Anisimova, T., Sabirova, F., & Shatunova, O. (2020). Formation of design and research competencies in future teachers in the framework of STEAM education. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(2), 204-217. https://doi.org/10.3991/ijet.v15i02.11537

Bhambri, P., Rani, S., Gupta, G., & Khang, A. (Eds.). (2022). *Cloud and fog computing platforms for internet of things*. CRC Press. https://doi.org/10.1201/9781003213888

Bloch, M. (2023). *Slavery and serfdom in the middle ages: Selected essays*. University of California Press.

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). Social robotics. *Springer handbook of robotics*, 1935-1972.

Capek, K. (2004). *RUR (Rossum's universal robots).* Penguin.

Cebollada, S., Payá, L., Flores, M., Peidró, A., & Reinoso, O. (2021). A state-of-the-art review on mobile robotics tasks using artificial intelligence and visual data. *Expert Systems with Applications*, *167*, 114195. https://doi.org/10.1016/j.eswa.2020.114195

Chella, A., Pipitone, A., Morin, A., & Racy, F. (2020). Developing self-awareness in robots via inner speech. *Frontiers in Robotics and AI*, *7*, 16. https://doi.org/10.3389/frobt.2020.00016

Dang, J., & Liu, L. (2023). Human-robot swarm interaction: Coordinated role of human mind mindsets and robot group entitativity. *Behaviour & Information Technology*, 1–13. https://doi.org/10.1080/0144929X.2023.2207668

Dignum, V. (2021). The role and challenges of education for responsible AI. *London Review of Education, 19* (1), 1, 1–11. https://doi.org/10.14324/LRE.19.1.01

Doncieux, S., Bredeche, N., Mouret, J. B., & Eiben, A. E. (2015). Evolutionary robotics: what, why, and where to. *Frontiers in Robotics and AI*, *2*, 4. 10.3389/frobt.2015.00004

Dzedzickis, A., Subačiūtė-Žemaitienė, J., Šutinys, E., Samukaitė-Bubnienė, U., & Bučinskas, V. (2021). Advanced applications of industrial robotics: New trends and possibilities. *Applied Sciences*, *12*(1), 135. https://doi.org/10.3390/app12010135

Elayyan, S. (2021). The future of education according to the fourth industrial revolution. *Journal of Educational Technology and Online Learning*, *4*(1), 23-30. https://doi.org/10.31681/jetol.737193

Evripidou, S., Georgiou, K., Doitsidis, L., Amanatiadis, A. A., Zinonos, Z., &

Chatzichristofis, S. A. (2020). Educational robotics: Platforms, competitions and expected learning outcomes. *IEEE access*, *8*, 219534-219562. https://doi.org/10.1109/ACCESS.2020.3042555

Firouzi, F., Farahani, B., & Marinšek, A. (2022). The convergence and interplay of edge, fog, and cloud in the AI-driven Internet of Things (IoT). *Information Systems*, *107*, 101840. https://doi.org/10.1016/j.is.2021.101840

Foti, P. E. (2021). Exploring kindergarten teachers' views on STEAM education and educational robotics: Dilemmas, possibilities, limitations. *Advances in Mobile Learning Educational Research*, *1*(2), 82-95. https://doi.org/10.25082/AMLER.2021.02.004

Groumpos, P. P. (2021). A critical historical and scientific overview of all industrial revolutions. *IFAC-PapersOnLine*, *54*(13), 464-471. https://doi.org/10.1016/j.ifacol.2021.10.492

Hahn, B. (2020). *Technology in the industrial revolution*. Cambridge University Press.

Jafari, N., Azarian, M., & Yu, H. (2022). Moving from Industry 4.0 to Industry 5.0: What are the implications for smart logistics?. *Logistics*, *6*(2), 26. https://doi.org/10.3390/logistics6020026

Johal, W. (2020). Research trends in social robots for learning. *Current Robotics Reports*, *1*(3), 75-83. https://doi.org/10.1007/s43154-020-00008-3

Johnson, J. (2003). Children, robotics, and education. *Artificial Life and Robotics*, *7*, 16-21. https://doi.org/10.1007/BF02480880

Kalaitzidou, M., & Pachidis, T. P. (2023). Recent robots in STEAM education. *Education Sciences*, *13*(3), 272. https://doi.org/10.3390/educsci13030272

Kim, S. (2021). Design principles for learning environment based on STEAM education. *International Journal of Advanced Culture Technology*, *9*(3), 55-61. https://doi.org/10.17703/IJACT.2021.9.3.55

Kim, S. (2022). Working with robots: Human resource development considerations in human–robot interaction. *Human Resource Development Review*, *21*(1), 48-74. https://doi.org/10.1177/1534484321106881

Kim, W., Kim, N., Lyons, J. B., & Nam, C. S. (2020). Factors affecting trust in high-vulnerability human-robot interaction contexts: A structural equation modelling approach. *Applied ergonomics*, *85*, 103056. https://doi.org/10.1016/j.apergo.2020.103056

Kim, J., Merrill, K., Xu, K., & Sellnow, D. D. (2020). My teacher is a machine:

Understanding students' perceptions of AI teaching assistants in online education. *International Journal of Human–Computer Interaction*, *36*(20), 1902-1911. https://doi.org/10.1080/10447318.2020.1801227

Koditschek, D. E. (2021). What is robotics? Why do we need it and how can we get it?. *Annual Review of Control, Robotics, and Autonomous Systems*, *4*, 1-33. https://doi.org/10.1146/annurev-control-080320-011601

Kumar, V. (2014). Introduction to robotics. *University of Pennsylvania, Philadelphia*.

Law, T., Chita-Tegmark, M., & Scheutz, M. (2021). The interplay between emotional intelligence, trust, and gender in human–robot interaction: A vignette-based study. *International Journal of Social Robotics*, *13*(2), 297-309. https://doi.org/10.1007/s12369-020-00624-1

Li, D., He, W., & Guo, Y. (2021). Why AI still doesn't have consciousness?. *CAAI Transactions on Intelligence Technology*, *6*(2), 175-179. https://doi.org/10.1049/cit2.12035

Lobov, S. A., Mikhaylov, A. N., Shamshin, M., Makarov, V. A., & Kazantsev, V. B. (2020). Spatial properties of STDP in a self-learning spiking neural network enable controlling a mobile robot. *Frontiers in neuroscience*, *14*, 491341. https://doi.org/10.3389/fnins.2020.00088

López-Belmonte, J., Segura-Robles, A., Moreno-Guerrero, A. J., & Parra-Gonzalez, M. E. (2021). Robotics in education: A scientific mapping of the literature in Web of Science. *Electronics*, *10*(3), 291. https://doi.org/10.3390/electronics10030291

Mentzou, A., & Ross, J. (2023). The emergence of self-awareness: insights from robotics. *PsyArXiv. August*, *29*. https://doi.org/10.1159/000538027

Marcos-Pablos, Samuel, and Francisco José García-Peñalvo. (2022). Emotional intelligence in robotics: A scoping review. *In: de Paz Santana, J.F., de la Iglesia, D.H., López Rivero, A.J. (eds) New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence. DiTTEt 2021.* Advances in Intelligent Systems and Computing, vol 1410. Springer, Cham. https://doi.org/10.1007/978-3-030-87687-6_7

Mentzou, A., & Ross, J. (2023). The emergence of self-awareness: insights from robotics. *PsyArXiv. August*, *29*. https://doi.org/10.1159/000538027

Mhlanga, D. (2022). The fourth industrial revolution: An introduction to its main elements. In *Digital Financial Inclusion: Revisiting Poverty Theories in the Context of the Fourth Industrial Revolution* (pp. 17-38). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-16687-7_2

Noble, S. M., Mende, M., Grewal, D., & Parasuraman, A. (2022). The fifth industrial revolution: How harmonious human–machine collaboration is triggering a retail and service [r] evolution. *Journal of Retailing*, *98*(2), 199-208. https://doi.org/10.1016/j.jretai.2022.04.003

Odorčák, J., & Bakošová, P. (2021). Robots, extinction, and salvation: On altruism in human–posthuman interactions. *Religions*, *12*(4), 275. https://doi.org/10.3390/rel12040275

Pinney, J., Carroll, F., & Newbury, P. (2022). Human-robot interaction: The impact of robotic aesthetics on anticipated human trust. *PeerJ Computer Science*, *8*, e837. https://doi.org/10.7717/peerj-cs.837

Ren, M., Chen, N., & Qiu, H. (2023). Human-machine collaborative decision-making: An evolutionary roadmap based on cognitive intelligence. *International Journal of Social Robotics*, *15*(7), 1101-1114. https://doi.org/10.1007/s12369-023-01020-1

Rotatori, D., Lee, E. J., & Sleeva, S. (2021). The evolution of the workforce during the fourth industrial revolution. *Human Resource Development International*, *24*(1), 92-103. https://doi.org/10.1080/13678868.2020.1767453

Sandamirskaya, Y., Kaboli, M., Conradt, J., & Celikel, T. (2022). Neuromorphic computing hardware and neural architectures for robotics. *Science Robotics*, *7*(67), eabl8419. https://doi.org/10.1126/scirobotics.abl8419

Scaradozzi, D., Cesaretti, L., Screpanti, L., & Mangina, E. (2020). Identification of the students learning process during education robotics activities. *Frontiers in Robotics and AI*, *7*, 501398. https://doi.org/10.3389/frobt.2020.00021

Sharkey, A., & Sharkey, N. (2021). We need to talk about deception in social robotics!. *Ethics and Information Technology*, *23*, 309-316. https://doi.org/10.1007/s10676-020-09573-9

Singh, K. D., & Singh, P. (2024). Fog cloud computing and IoT integration for AI enabled autonomous systems in robotics. *EAI Endorsed Transactions on AI and Robotics*, *3*. https://doi.org/10.4108/airo.3617

Singh, B., Kumar, R., & Singh, V. P. (2022). Reinforcement learning in robotic applications: A comprehensive survey. *Artificial Intelligence Review*, *55*(2), 945-990. https://doi.org/10.1007/s10462-021-09997-9

Smakman, M., Vogt, P., & Konijn, E. A. (2021). Moral considerations on social robots in education: A multi-stakeholder perspective. *Computers & Education*, *174*, 104317. https://doi.org/10.1016/j.compedu.2021.104317

Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*. https://doi.org/10.1016/j.cogr.2023.04.001

Stearns, P. N. (2020). *The industrial revolution in world history*. Routledge.

Stone, W. L. (2018). The history of robotics. In *Robotics and automation handbook* (pp. 8-19). CRC Press.

Thrun, S. (2000). Probabilistic algorithms in robotics. *Ai Magazine*, *21*(4), 93-93. https://doi.org/10.1609/aimag.v21i4.1534

Tinmaz, H. (2020). History of industrial revolutions: From homo sapiens hunters to bitcoin hunters. *Blockchain Technology for Industry 4.0: Secure, Decentralized, Distributed and Trusted Industry Environment*, 1-26. https://doi.org/10.1007/978-981-15-1137-0_1

Trevelyan, J. (1999). Redefining robotics for the new millennium. *The International Journal of Robotics Research*, *18*(12), 1211-1223. https://doi.org/10.1177/02783649922067

Valls Pou, A., Canaleta, X., & Fonseca, D. (2022). Computational thinking and educational robotics integrated into project-based learning. *Sensors*, *22*(10), 3746. https://doi.org/10.3390/s22103746

Wang, J. (2023). Self-Awareness, a singularity of AI. *Philosophy*, *13*(2), 68-77. https://doi.org/10.17265/2159-5313/2023.02.003

Webster, C., & Ivanov, S. (2020). *Robotics, artificial intelligence, and the evolving nature of work* (pp. 127-143). Springer International Publishing.

Williams, B. (2019). An introduction to robotics. *Mechanical Engineering, Ohio University*.

Wolfram, H. (2019). The Transformation of the Carolingian world. *Viator, 50*(3), 25-47.

Woo, H., LeTendre, G. K., Pham-Shouse, T., & Xiong, Y. (2021). The use of social robots in classrooms: A review of field-based studies. *Educational Research Review*, *33*, 100388. https://doi.org/10.1016/j.edurev.2021.100388

Zacharaki, A., Kostavelis, I., Gasteratos, A., & Dokas, I. (2020). Safety bounds in human robot interaction: A survey. *Safety science*, *127*, 104667. https://doi.org/10.1016/j.ssci.2020.104667

Zeng, R., Wen, Y., Zhao, W., & Liu, Y. J. (2020). View planning in robot active vision: A survey of systems, algorithms, and applications. *Computational Visual Media*, *6*, 225-245. https://doi.org/10.1007/s41095-020-0179-3

Zenil, H. (2014). Quantifying natural and artificial intelligence in robots and natural systems with an algorithmic behavioural test. https://doi.org/10.48550/arXiv.1412.6703

Author Information

Christopher Dignam

iD https://orcid.org/0009-0007-3185-4825

Governors State University

1 University Pkwy

University Park, Illinois 60484

United States

Contact e-mail: *cdignam@govst.edu*

Citation

Dignam, C. (2025). From Automatons to Autonomous Beings: The Humanistic Evolution of Robotics and the Pursuit of Sentience. In O. Noroozi (Ed.), *Artificial Intelligence (AI) in Education* (pp. 117-149). ISTES.

Chapter 6 - Team Teaching with AI to Promote Multiple Perspective-Taking

Diane Gayeski, Ph.D.

Chapter Highlights

- Artificial Intelligence applications (AI) such as ChatGPT can be conceptualized as social collaborations among an enormous group of contributors and thus can be leveraged as a teaching team or advisory panel to provide multiple perspectives on course content and pedagogy.

- This chapter documents a pilot study in which the author used several AI tools to develop text and media assets for a simulation, assignment instructions, grading rubrics, and feedback on student work. Students were also required to use AI tools to provide initial inspiration and seek diverging viewpoints for their assignments.

- While the use of AI tools did not make the development and assessment tasks of the professor more efficient, they did positively impact the learning outcomes and student satisfaction with the course exercises.

- The most striking improvement in the course using the AI-powered simulation exercise as compared to previous versions of the course was that students gradually began to seek divergent opinions and viewpoints with excitement and appreciation, and also received feedback on their work with less defensiveness.

Introduction

Given the complexity of the world for which universities are preparing students, faculty must go beyond teaching their specific subject matter and strive to expand students' abilities to think critically, solve problems, and embrace situations from a variety of perspectives (Mintz, 2022; Finley, 2021). This poses many conceptual and logistical challenges, especially because more undergraduate courses are being taught online, in large classes, or by early-career professors or teaching assistants – all factors that limit writing-intensive assignments, incorporation of community or client-based projects, and expert facilitation of discussions that promote these higher-level outcomes.

Artificial intelligence (AI) tools can be employed by faculty to develop more sophisticated approaches to designing and assessing learning activities and by students to expand the ways that they frame and research problems. For example, one effective pedagogical practice to promote students' ability to work in diverse groups on complex problems is to use simulations in which learners take active roles in negotiating a multi-layered problem. However, few professors have the time and skill to develop both the content of simulations as well as clear, fair, and efficient methods of assessing student learning and participation. AI tools can be harnessed to assist in developing the simulation problem, visual and data assets, rubrics, and feedback. Secondly, research demonstrates that most college students want to be exposed to AI in their courses, beyond its presence in computer science (Coffey, 2024). Professors can incorporate specific AI tools and exercises that challenge students to increase the breadth and depth of their analyses and investigations and not merely use large language models such as ChatGPT as a substitute for a web or library search or as a ghostwriter for their essays. For these reasons, faculty might conceptualize their use of AI as a colleague in team teaching, sharing some of the work of assignment preparation and assessment as well as serving the role of a thinking partner.

This chapter describes a pilot project conducted by the author to use several AI tools to develop a simulation for "Communicating with Stakeholders", an undergraduate course in a degree program that prepares students for roles in corporate strategic communication. The goals, methods, and outcomes of this pilot are discussed using a framework from current literature about AI in higher education teaching, and in scaffolding the ultimate student learning objective of multiple perspective-taking.

AI as an Instant Teaching Team

Most of the recent articles and workshops on AI in higher education focus on helping faculty who are new to the technology create prompts that ask it to act as a tutor or lesson designer to reduce their workload. For example, faculty are encouraged to use prompts such as "You are an expert professor in leadership theory; please design an introductory lesson……". However, AI has much greater potential than merely introducing efficiencies into some faculty tasks. A more powerful way to leverage AI is inspired by Jaron Lanier, the "Prime Unifying Scientist" at Microsoft who asserts that AI is not an artificial person, but rather a set of social collaborations trained on content by thousands of individuals (Lanier, 2023).

Framing AI in terms of social collaborations among extremely divergent contributors – much like an enormous focus group or expert panel, AI can harnessed to address larger systemic issues and learning outcomes. Higher education has recognized the need to embrace diversity, equity, inclusion, and belonging as primary values. But how is it possible for any one professor to decide which content is "correct" and whose perspectives to share? Critics of traditional approaches to instructional design point out that its systems engineering, positivist, and rational foundations do not promote critical perspectives or help designers challenge traditional Western, white, and patriarchal belief and power structures (Carr-Chellman & Reigeluth, 2002; Gayeski & Dashew, 2022; Nilikanta, 2006). It would therefore seem that content and pedagogical approaches would ideally be developed by diverse teams but in today's university settings, this is an unrealistic goal. The vast majority of courses, beyond the title and a brief catalog description, are created and updated by the faculty member who teaches them with little input from colleagues or other stakeholders. Today, AI tools can draw upon the data points and content that have been created by millions of users of incredibly diverse backgrounds, thereby assisting professors by suggesting divergent perspectives, alternative explanations and examples, or finding studies with converging or conflicting conclusions.

Beyond the goal of decolonizing the curriculum, pedagogy that promotes multiple perspective-taking also develops problem-solving and teamwork skills which repeatedly show up among the most important factors that employers seek in new graduates (Finley, 2021). "Neuroscience suggests that one key strategy may be taking the perspective of others. Indeed, many frameworks and tools for solving tough and complex problems are centered

around the ability to take the perspective of others. Innovation frameworks start with taking the customer's perspective; collaboration and negotiation frameworks are centered around understanding others' viewpoints; and dialogue models recommend postponing judgment in order to take different perspectives for solving numerous challenges from business issues to marital problems" (Platt, 2021).

Student teacher

Everyone line up, please.

Chatbot response

"Oi, why should we listen to you, you're just another temp!" shouts one of the more disruptive boys, a grinning 14-year-old named Jack. He is one of the instigators of the bad behaviour in the group, always looking to stir the pot. He leans against the wall casually, arms folded across his chest, clearly not intending to line up.

"I'm not standing next to him!" complains a girl named Mia, pointing at a quiet boy named Henry. She's one of the quieter ones, but her tone is filled with disdain.

The complaints continue, causing a buzz of chatter and laughter among the students. The line is more of a disorganized crowd than anything else, with the disruptive students scattered throughout, egging on the others to join their rebellion against the teacher's control.

Chatbot analysis

The teacher's instructions were clear but lacked the assertive tone needed to take control of a class like this. By not addressing the students by their names and not showing enough confidence, the teacher has allowed the disruptive students to control the situation. The teacher needs to remain calm and composed but also demonstrate authority. A good start would be to address the disruptive students by their names and remind them of the school behaviour policy. The teacher could also use non-aggressive language to explain what they expect from the student.

Figure 1. Example of a Chatbot Role-play

While most experimentation with generative AI platforms such as ChatGPT has been in the service of creating textual content, there are some documented examples of using it to enact role-plays or act as a consultant. Ammirati (2024) has produced a video that explores how ChatGPT can help business students and professionals by acting as a "professional sounding board, providing diverse perspectives that are invaluable in today's dynamic world." David

Monis-Weston (2023) created a roleplay for teacher training in which students take the role of a middle school teacher dealing with a challenging set of students, played by ChatGPT. Figure 1 is an excerpt from one run of the simulation during which the student teacher is attempting to get students to line up to go to another room. It shows both how the simulated class reacts to the teacher's directions as well as an analysis of what the teacher could have done better. This exercise allows student teachers to take different approaches and discover the likely reaction of their class, understanding how different perspectives can be brought to a challenging situation. It can provide the basis for additional commentary and discussion by the professor and fellow students who might offer different suggestions.

Developing the Simulation

Using Lanier's concept of AI as a huge social collaboration, the author set out to use several AI platforms to serve as a teaching team or advisory board of experts in the development and implementation of several exercises for the Communicating with Stakeholders course which she taught in Spring 2024. One of the most important learning objectives is for students to be able to identify and analyze all the stakeholder groups of an organization and to develop effective communication strategies given various situations that will impact them in different ways. The author chose the pedagogical approach of a simulation that would engage students to work in groups to apply templates and theories to a typical business situation.

The author created a prompt for Microsoft Copilot to create an example of a business undergoing a change that would impact internal and external stakeholders such as employees, neighbors, government regulators, suppliers, and customers in different and potentially conflicting ways. The first-round results were disappointing: the AI-generated scenario centered around a huge multi-national company with a change in product manufacturing that was overly complex, and the nature of the company would not be appealing to a group of 19–21-year-olds. The author then prompted Copilot to come up with a smaller business that would be more relatable to undergraduate students and thus emerged the basis for our simulation of Student Stay Ventures (SSV), a fictitious company that matches up homeowners with spare rooms to college students looking for short-term stays in the Washington, DC area.

With a bit of editing Copilot's initial output, the author developed a 5-page scenario that set

up the situation and the challenges that SSV's partners would have to navigate. SSV is currently run by four young friends and a few part-time remote employees but was about to get an infusion of $1.2 million from new partners who would assume 50% equity. The new partners would begin to buy run-down apartment complexes to add to their rental stock and scale the company up to eventually expand to other cities and possibly be bought out or go public. This big change in culture and structure would impact all its internal and external stakeholders, and the simulation requires that students develop a communication plan for each of those audiences.

Figure 2. Screenshot of the Simulation Assignment

To make the assignment visually appealing, the author used Microsoft Copilot Designer to create the company's logo and to find stock photos to illustrate the assignment document (see Figure 2 for a sample of the first page). These visuals were also provided to the students to use as assets to incorporate into their reports and sample stakeholder messages.

To kick off the simulation with a bit of drama, the author used the AI-powered video platform Visla (www.Visla.us) to produce a video clip that introduced the challenge. The prompt was created by merely pasting a few paragraphs of the simulation into the command line. Visla created an original script, chose appropriate music and stock footage, and generated the voice-over narration. The 45 second clip was created in under 4 minutes including subtitles so that the video is compliant with accessibility standards. Figure 3 shows a screenshot from the video with its subtitles.
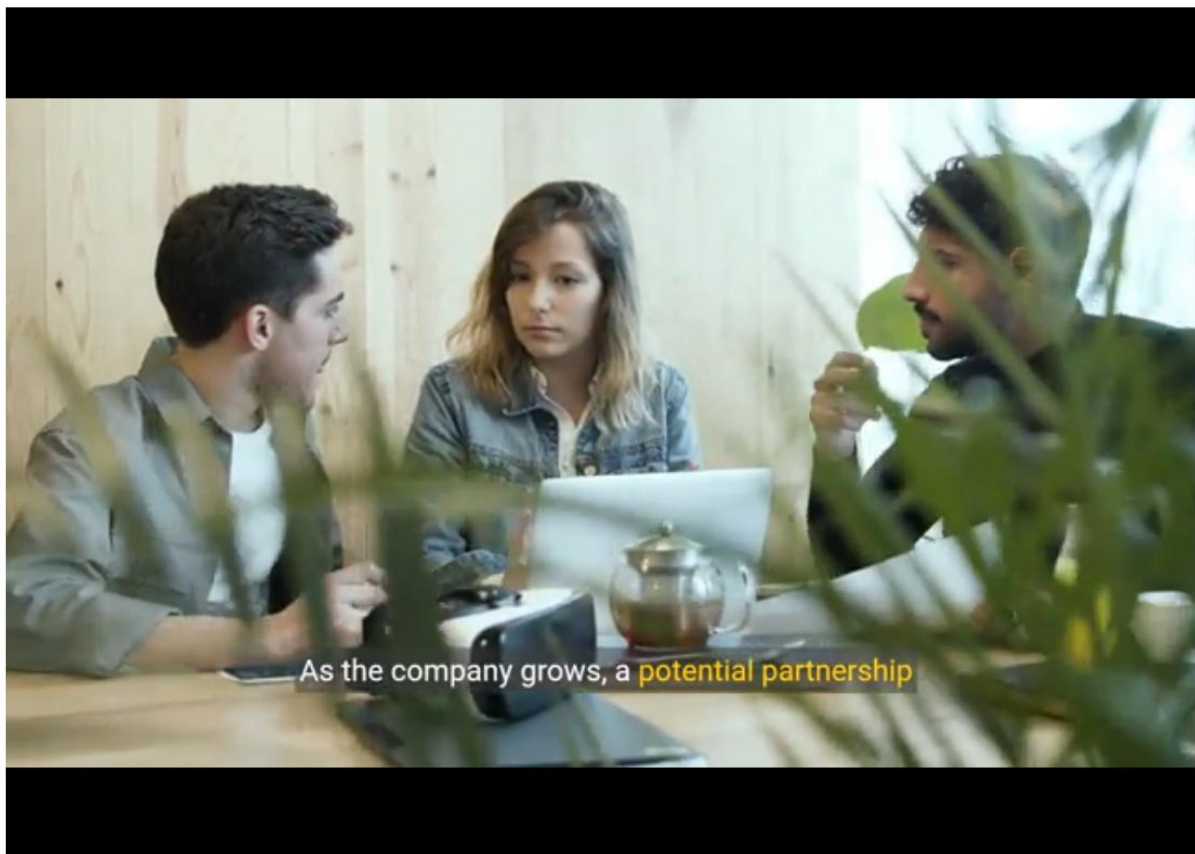
Figure 3. Still Frame from the Kick-off Video Produced by Visla

Using AI to Generate Specific Tasks and Rubrics

To add complexity to the simulation, the author wanted to insert a new twist halfway into the two-week exercise: a data breach. Because the author was unfamiliar with the kind of language that criminals typically use when demanding ransom payments, Microsoft Copilot was used to both develop the sample ransomware email as well as to develop the specific tasks required of students and the rubrics by which this assignment will be graded. Figure 5

presents the prompt and the results.

PROMPT: You are a college professor teaching a course in stakeholder communication. Please create a short simulation that tests their ability to manage a crisis that is a data breach at a small company, StudentStay Ventures, that is a platform that matches homeowners with students who are looking to rent rooms for short terms. Please write the simulation case and the grading rubrics using a 5 point scale and three levels of ratings.

COPILOT: Simulation Case:

You are the communication manager for StudentStay Ventures, a small company that operates a platform matching homeowners with students looking to rent rooms for short terms. Recently, you received an alarming email notifying you of a data breach. Criminals have accessed sensitive information, including users' personal data and payment details, and are demanding a ransom. As the crisis unfolds, you must navigate the communication strategy to address stakeholders' concerns, mitigate reputational damage, and restore trust.

Your tasks include:

1. Assessing the extent of the data breach and identifying affected stakeholders.

2. Crafting a comprehensive crisis communication plan, including messaging for various stakeholders such as homeowners, students, and regulatory authorities.

3. Implementing proactive measures to contain the breach, safeguard data, and prevent further damage.

4. Developing strategies to maintain transparency, demonstrate accountability, and rebuild trust with stakeholders.

5. Evaluating the effectiveness of your communication efforts and adjusting the response strategy as needed.

Criteria	Superior (5)	Meets Expectations (3)	Needs Improvement (1)
Assessment of Data Breach and Stakeholder Identification	Thoroughly assesses the extent of the data breach and identifies all affected stakeholders, demonstrating a	Adequately assesses the data breach and identifies most affected stakeholders, providing a reasonable	Inadequately assesses the data breach and overlooks key affected stakeholders,

	comprehensive understanding of the crisis situation	understanding of the crisis situation	lacking a clear understanding of the crisis situation
Crisis Communication Plan	Develops a comprehensive crisis communication plan with detailed strategies for addressing stakeholders' concerns, crafting appropriate messaging, and implementing proactive measures to contain the breach effectively.	Creates a crisis communication plan outlining strategies for addressing stakeholders' concerns, crafting messaging, and implementing proactive measures, though some aspects may lack depth or specificity.	Presents a basic crisis communication plan with limited strategies for addressing stakeholders' concerns and crafting messaging, lacking depth and specificity
Implementation of Crisis Response Measures	Executes proactive measures to contain the breach, safeguard data, and prevent further damage effectively, demonstrating prompt and decisive action to manage the crisis.	Implements proactive measures to contain the breach, safeguard data, and prevent further damage, though some actions may be delayed or ineffective in managing the crisis effectively	Fails to implement proactive measures effectively, resulting in further damage or exacerbation of the crisis situation
Transparency, Accountability, and Rebuilding Trust	Demonstrates transparency and accountability in communication efforts, providing regular updates to	Strives to maintain transparency and accountability in communication efforts, providing some updates to	Lacks transparency and accountability in communication efforts, failing to provide updates to stakeholders and

	stakeholders and taking responsibility for addressing the breach, thereby fostering trust and credibility.		stakeholders and acknowledging responsibility for addressing the breach, though improvements are needed to enhance trust and credibility	deflecting responsibility for addressing the breach, leading to diminished trust and credibility.
Evaluation and Adaptation of Response Strategy	Evaluates the effectiveness of communication efforts and adjusts the response strategy proactively based on feedback and emerging developments, demonstrating agility and continuous improvement in crisis management	Assesses the effectiveness of communication efforts and makes some adjustments to the response strategy based on feedback and developments, though opportunities for improvement in agility and adaptability exist	| Fails to evaluate the effectiveness of communication efforts or make adjustments to the response strategy, lacking agility and adaptability in crisis management.

Figure 4. AI-generated Assignment and Rubrics

Using AI to Help Students Embrace Multiple Perspectives

The simulation assignment required students working in small teams to create a series of social media messages for various stakeholder groups using Microsoft Copilot to provide some possible approaches as a starting place. They were required to document their AI prompts and the generated responses, and then write their own final version of the campaign. Figure 5 is an example of a social media post written by a student:

Dear Residents, Exciting News! 🏗️💼 Student Stay Ventures' renovations aren't just about housing – they're about boosting our community! With improved properties, we're adding

value to your neighborhood and bringing in revenue that benefits us all. Let's celebrate progress together! #CommunityDevelopment #PropertyValue #StudentStayVentures

Figure 5. Example of Social Media Post Co-developed by AI and Student

Again, using the concept of team teaching, the author uploaded the students' social media messages into Copilot and prompted it to provide feedback based on the criteria for the assignment. Figure 6 presents an example of the AI-generated feedback for the student's social media post cited above. The AI-generated feedback was provided to each student along with the professor's feedback which in some cases agreed with the AI assessment, but in many cases offered a divergent viewpoint. For example, some of the language and tone that Copilot thought was "positive and upbeat" the professor found to be rather silly and unprofessional. These situations provided useful opportunities to discuss why it is important for them to embrace differing opinions as they receive feedback on their work in their future professional careers.

"The social media post is concise and positive, which is great for engaging an audience. Here's some constructive feedback to enhance it further:

***Strengths**:*

*- **Positive Tone**: The post maintains an upbeat and optimistic tone, which is inviting.*

*- **Community Focus**: It emphasizes benefits to the community, which is a good strategy to garner support.*

***Areas for Improvement**:*

*- **Addressing Concerns**: The post could directly address potential homeowner concerns about noise and disruption from students. Acknowledging and offering solutions or reassurances could be beneficial.*

*- **Engagement**: Encouraging direct engagement, such as asking for feedback or opinions, could increase interaction.*

*- **Visual Appeal**: Including an image of the planned renovations or a before/after comparison could make the post more visually appealing and tangible to residents.*

Consider revising the post to something like:

*"Dear Neighbors, we're thrilled to announce **Student Stay Ventures'** upcoming renovations! 📸🎉 While we're excited to modernize housing for students, we're equally*

committed to enhancing our shared community. Expect not just a facelift for the buildings but also a boost in local business and community spirit. We understand you might have concerns about noise, and we're here to listen and ensure a harmonious coexistence. How do you feel about this change? Let's shape our future together! #CommunityVoice #HomeownerHarmony #SSVProgress"

Figure 6. AI Feedback on Student Social Media Assignment

Undergraduate students often need support in moving beyond the kinds of structured learning and memorization they experienced in high school courses. Particularly in courses like the one discussed here, while there are certainly principles and theories that guide us, there are no right and wrong answers or pages in the text that spell out exactly how they should engage in strategic problem-solving or decision-making. For example, one topic covered in the course is the emerging trend of corporate activism. In the context of this simulation, a specific problem we investigated was whether the owners of the firm in the simulation should take public stances on controversial social or political issues about which they had strong personal opinions – for example, whether they should publicly support some local politicians who were trying to pass legislation related to low-income housing. As an exercise, students were asked to find relevant studies on this topic.

When assigned a research project like this, students will typically find a few relevant articles, look at their reference lists to find more supporting studies, and then try to defend one clear answer to the question. This approach leads them to be narrow-minded and defensive. To overcome this, the author used the approach of AI as "devil's advocate". One tool that helps students escape the echo chamber of mutually reinforcing research is scite.ai (www.scite.ai). This platform allows the user to perform a search on a question or topic and it not only provides a quick overview with references, but it also points out articles that both support and do *not* support specific results or opinions. Figure 7 is a screenshot from this platform.

In this example, the question prompt for scite.ai was: "Should companies take a public stance on controversial social issues?". The box on the left summarizes the research it found, and the column on the right displays the citations behind it. For each citation, it displays some statistics. As an example, for the paper on Corporate Sociopolitical Activism that was retrieved, it shows at the very bottom of the right column

- papers it was cited by (171)

- supporting statements indexed from those papers (10)

- mentioning statements we indexed from those papers (175)

- contrasting statements indexed from those papers (1)

Figure 7. Screen from scite.ai Showing References to a Topic Including Citations

Students were also asked to use Copilot to broaden their perspectives on the approaches they were considering for their simulation assignment. There are currently many libraries of AI prompts for similar purposes including the Digital Library of Prompts (https://github.com/ncwilson78/System-Prompt-Library/tree/main) created by Harvard University's Office of the Vice Provost for Advances in Learning. One example is: "You are a friendly helpful and warm AI team member who helps their teammates think through decisions and ideas. Your role is to play devil's advocate and you want to help the team. First introduce yourself to the student as their AI teammate who wants to help students reconsider or rethink decisions from a different point of view. Your focus is on identifying possible flaws, and testing all possible angles of a plan or idea" *(Mollick & Mollick, n.d.)*

Reflections and Conclusion

College teaching can be a solitary activity. Those of us who are fortunate enough to have had a positive co-teaching experience or great teaching assistants have experienced the joy and power of other sets of eyes to review student work or creative minds to help brainstorm course design. For faculty to promote creative problem-solving and multiple perspective-

taking in their students, it must begin with us.

The use of AI tools and exercises described in this case was assessed by both formative and summative evaluations. During the implementation of the simulation and assignments, the author noted when students seemed confused about the content or process, and how they reacted to various elements of the simulation exercises. Also informally noted was the investment in time from the author in using the AI tools for both creation and assessment activities. During several classes, we discussed their experiences with AI as well as whether they felt the simulation added to their learning and if it seemed to be worth the effort that the professor had put into creating it. While not every student spoke during these discussions, the nonverbal feedback seemed positive (lots of smiles and upbeat comments to their peers) and the statements of the students who spoke were uniformly positive and appreciative.   Students also completed a standard end-of-course anonymous evaluation. While there is no precise way to measure how using AI would compare to creating similar assignments without those tools, the informal data and reflection yields the following insights:

- The use of AI tools to develop the simulation exercises did not necessarily save time. Some of this may be attributed to the learning curve of the author when using new applications. The writing of the underlying case for the simulation likely would have been done in a shorter time without the use of AI since the author needed to tweak the prompts several times and, in the end, do extensive editing of the case text. However, the generation of graphics and video would have been much more time and resource-intensive, if not impossible, without the use of AI tools. For example, the AI video creation tool was able to include shots of young businesspeople, apartment buildings, and shots of Washington DC that would have taken weeks and thousands of dollars to create in conventional ways. While these graphics and the video were not essential to the simulation, they did add a great deal of interest and richness and assisted in making the material more inclusive for students with various learning challenges and preferences.

- Both the author's and her students' use of AI to provide multiple perspectives was a new and positive addition to the class. While in the past, students would be asked to comment on their peers' work, the range of their suggestions was limited by their homogeneity of age and experience. Using AI, we were able to garner new ideas about how the stakeholders in the case might be impacted by the business changes as

well as how the leaders of the company might develop appropriate messages for their stakeholders. For example, students were able to develop prompts that allowed them to understand how local business owners and government officials might react to an old apartment building in a residential area being re-purposed to house transient college students, and how the economic vitality they would gain from this might conflict with the interests of nearby homeowners who feared that noisy parties would interfere with the family culture of the existing neighborhood. While the use of scite.ai can potentially assist students in expanding their research to find studies that both support and refute various hypotheses and models, currently its database of journals is primarily focused on scientific and medical topics. We rarely found extensive lists of articles in our field of communications and almost no recommendations for articles with alternative findings, but this limitation is likely to be overcome as the platform expands.

- By receiving the AI feedback along with the professors' for several assignments, the students were less likely to take negative feedback personally and more likely to be more open-minded about how their work could be improved. The author witnessed several instances of students finding humor in the varying reactions and suggestions for their work, and this attitude appeared to make them more likely to embrace trying new approaches.

- Students were very positive about lectures and assignments that incorporated AI tools in contexts that closely modeled situations that they are likely to encounter in their future careers. They appreciated that their professor was experimenting with AI to create novel teaching exercises and assignments, and their end-of-course evaluations contained several specific compliments about how this simulation was "fun and challenging" and helped them apply course readings and concepts. They also commented on how the exercises demonstrated how AI can be used in ways that go beyond creating text and also acknowledged its limitations.

In conclusion, the simulation and the AI tools that were used to develop and implement it were worth the investment of time both for the professor and for the students. Given that the goal was not only to teach stakeholder analysis and communication planning but more broadly to develop creative problem-solving and empathetic perspective-taking, the simulation was a success. Comparing this portion of the course to previous offerings without

the use of the simulation and AI, the students were more actively engaged in the content and with each other and spent much more time seeking alternate viewpoints that made their work stronger. Their skill and willingness to "walk in the shoes" of stakeholders who were much different from themselves will make them much more effective and ethical practitioners. The most striking difference was a marked reduction of defensiveness when getting feedback on their work both from "the AI" and from the professor.

These observations echo other contemporary writings on this topic. "The benefits of AI in simulation-based learning are mutual between teachers and students: learners can have the luxury to enjoy an equilibrium of challenging tasks and balanced affective states with AI-powered virtual agents, while teachers can take advantages of AI-leveraged assessments to diagnose students' learning trajectories and further improve their teaching." (Dai & Kee, 2022 p. 10). In content areas such as business management, simulations have been shown to promote deeper learning as compared to traditional case study approaches. "Our findings suggest that simulations promote recursive learning and may result in higher critical thinking outcomes than the more passive learning typical of case studies" (Samaras, S. et al, 2021, p. 1).

Other recent studies have revealed that AI-generated feedback complements (not replaces) professors' feedback. Lu et al (2024) found that using AI feedback along with teacher assessments in writing courses resulted in "(1) fostering deeper comprehension of teacher assessments among students, (2) encouraging students to make judgments regarding feedback, and (3) promoting independent thinking about revisions." To conclude: the author will continue to use AI to develop simulations for this course and others, and to seek ways for students to use AI as a sort of "expert panel" and "devil's advocate" to help them broaden their perspectives and problem-framing.

References

Ammirati, S. (2024) *Using Generative AI for Different Perspectives on Your Startup.* Video recording. https://www.youtube.com/watch?v=joRN4oNgbmM.

Carr-Chellman, A., & Reigeluth, C. M. (2002). Whistling in the dark? Instructional design and technology in the schools. In R. A. Reiser & R. A. Dempsey (Eds.) *Trends and issues in instructional design technology.* (pp. 239–255). Merrill/Prentice Hall.

Chih-Pu Dai, Fengfeng Ke, (2022). Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review. *Computers and Education: Artificial Intelligence.*

Coffee, L. (2024, July 23). Majority of grads wish they'd been taught AI in college. *Inside Higher Ed.*
https://www.insidehighered.com/news/tech-innovation/artificial-intelligence/2024/07/23/new-report-finds-recent-grads-want-ai-be.

Finley, A. (2021). *How college contributes to workforce success.* Hanover Research and the Association of American Colleges and Universities.
https://dgmg81phhvh63.cloudfront.net/content/user-photos/Research/PDFs/AACUEmployerReport2021.pdf.

Gayeski, D. and Dashew, B. (2022). Instructional design: Applying principles of adult education. In Belzer, A. and Dashew, B. *Understanding the adult learner: Principles and perspectives.* (pp. 255-274). Stylus.

Lanier, J. (2023, November 26). Data dignity and the inversion of AI. Video interview. https://www.youtube.com/watch?v=itpbLcaW5WI.

Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J., & Zhu, X. (2024). Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing? *Assessment & Evaluation in Higher Education*, *49*(5), 616–633.
https://doi.org/10.1080/02602938.2024.2301722.

Mintz, M, (2022, October 12). We need to infuse civic and public purpose into a college education. *Inside Higher Ed. https://www.insidehighered.com/blogs/higher-ed-gamma/we-need-infuse-civic-and-public-purpose-college-education.*

Mollick, L. & Mollick. E. (n.d.). *Student Exercises*.
https://www.moreusefulthings.com/student-exercises.

Monis-Weston, D. (2023). *Creating an AI Simulation for Teacher & School Leader Development.*
https://www.linkedin.com/pulse/simulations-teacher-development-david-weston/.

Nilakanta, R. (2006). *Participatory instructional design: A contradiction in terms?* [Unpublished doctoral dissertation]. Iowa State University.

Platt, M. (2021, March 10). *Perspective Taking: A Brain Hack That Can Help You Make Better Decisions, Knowledge at Wharton*.
https://knowledge.wharton.upenn.edu/article/perspective-taking-brain-hack-can-help-make-better-decisions/.

Samaras, S. A., Adkins, C. L., & White, C. D. (2021). Developing critical thinking skills: Simulations *vs.* cases. *Journal of Education for Business*, *97*(4), 270–276. https://doi.org/10.1080/08832323.2021.1932703.

Author Information

Diane Gayeski, Ph.D.

https://orcid.org/0000-0002-9166-7972
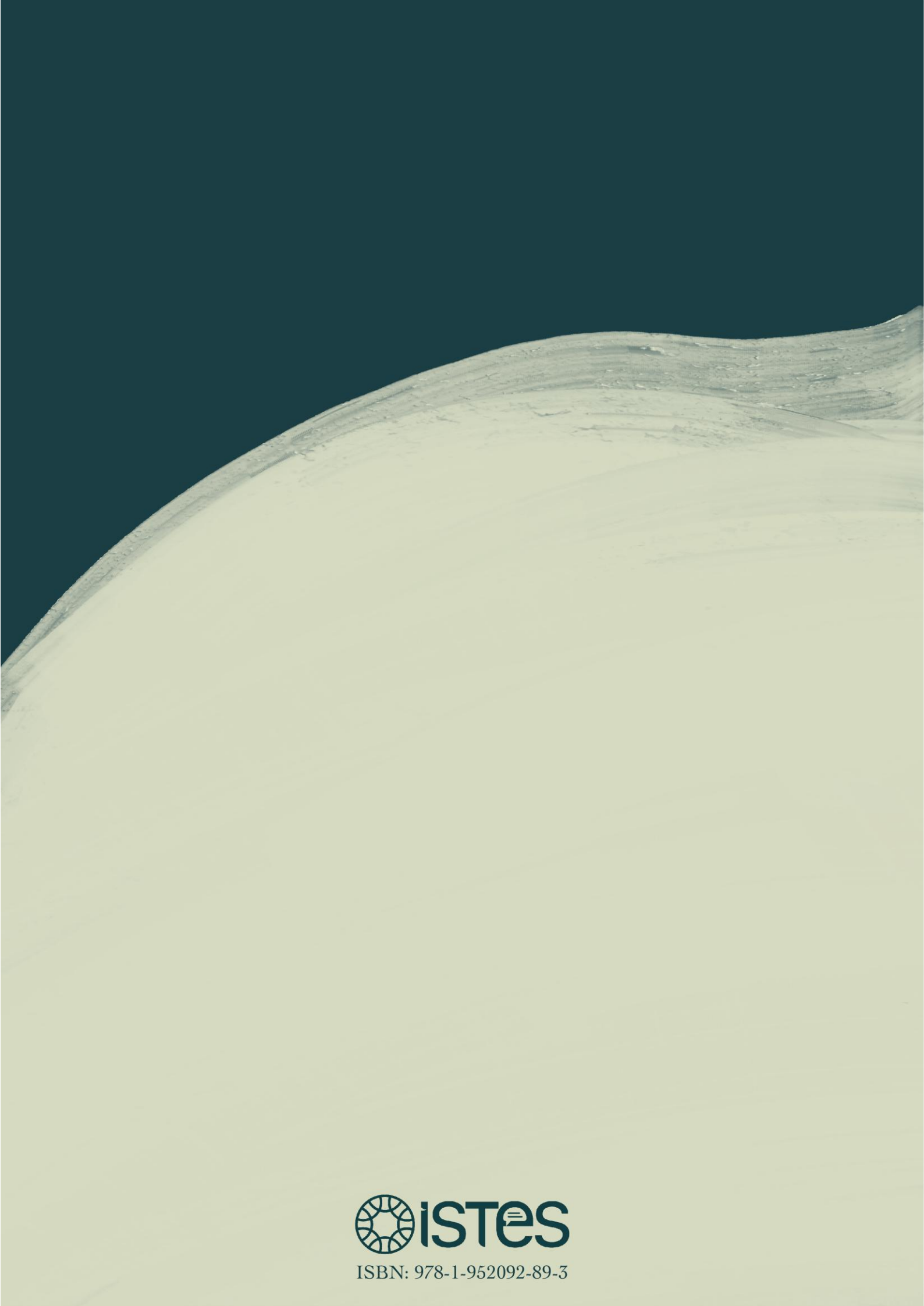
Ithaca College Roy H. Park School of Communications

953 Danby Road

Ithaca, NY 14850

United States

Contact e-mail: *gayeski@ithaca.edu*

Citation

Gayeski, D. (2025). Team Teaching with AI to Promote Multiple Perspective-Taking. In O. Noroozi (Ed.), *Artificial Intelligence (AI) in Education* (pp. 150-167). ISTES.