Promises and challenges of generative artificial intelligence for human learning

4 authors:

Lixiang Yan
Monash University (Australia)
47 PUBLICATIONS   **849** CITATIONS

SEE PROFILE

Ziwen Teuber
University of Luxembourg
43 PUBLICATIONS   **255** CITATIONS

SEE PROFILE

Samuel Greiff
Technische Universität München
389 PUBLICATIONS   **9,223** CITATIONS

SEE PROFILE

Dragan Gasevic
The University of Edinburgh
778 PUBLICATIONS   **25,605** CITATIONS

SEE PROFILE

Promises and challenges of generative artificial intelligence for human learning

Lixiang Yan[1], Samuel Greiff[2,3*], Ziwen Teuber[2], Dragan Gašević[1*]

[1]Faculty of Information Technology, Monash University, 20 Exhibition Walk, Clayton, 3168, Victoria, Australia.
[2]Department of Behavioral & Cognitive Science, University of Luxembourg, 11, porte des sciences, Esch, 4366, Luxembourg.
[3]Department of Educational Psychology, Goethe-University Frankfurt, Theodor-W.-Adorno-Platz 6, Frankfurt, 60323, Germany.

*Corresponding author(s). E-mail(s): samuel.greiff@gmail.com; dragan.gasevic@monash.edu;
Contributing authors: lixiang.yan@monash.edu; ziwen.teuber@uni.lu;

Abstract

Generative artificial intelligence (GenAI) holds the potential to transform the delivery, cultivation, and evaluation of human learning. This Perspective examines the integration of GenAI as a tool for human learning, addressing its promises and challenges from a holistic viewpoint that integrates insights from learning sciences, educational technology, and human-computer interaction. GenAI promises to enhance learning experiences by scaling personalised support, diversifying learning materials, enabling timely feedback, and innovating assessment methods. However, it also presents critical issues such as model imperfections, ethical dilemmas, and the disruption of traditional assessments. Cultivating AI literacy and adaptive skills is imperative for facilitating informed engagement with GenAI technologies. Rigorous research across learning contexts is essential to evaluate GenAI's impact on human cognition, metacognition, and creativity. Humanity must learn with and about GenAI, ensuring it becomes a powerful ally in the pursuit of knowledge and innovation, rather than a crutch that undermines our intellectual abilities.

Keywords: Generative Artificial Intelligence, Human Learning, AI Agent, Large Language Models, Diffusion Models

1 Main

Human learning is a journey that shapes minds, fosters innovation, and builds the foundations of society. Beyond merely acquiring knowledge and skills, learning is a path towards fostering critical thinking, creativity, collaboration, and social cohesion. By nurturing the ability to question, analyse, and innovate, learning empowers individuals to navigate complex challenges and contribute to societal progress. While education encompasses formalised systems that structure learning processes, learning represents the dynamic and personal process that occurs within this framework (Table 1).

The history of human learning presents a narrative of continuous evolution and adaptation to technological breakthroughs. For example, the printing press democratised access to knowledge and opened the opportunity of learning to many, while the Internet and digital technologies transformed information dissemination and collaborative learning across time and space. In this continuum of innovation, recent advancements in artificial intelligence (AI) present another transformative opportunity to rethink learning processes and educational methodologies [1].

Generative AI (GenAI) technologies, such as large language models (LLMs) and diffusion models (Table 2), have shown promise in automating various learning tasks [2], delivering feedback on human efficacy [3], outperforming average students in reflective writing [4], innovating conversational assessments [5], creating dynamic learning resources [6], and supporting multimedia learning [7]. However, these technologies also present challenges and ethical considerations that could outweigh their benefits [2, 8]. One major concern is the digital divide, where unequal access to these powerful technologies can exacerbate existing inequalities in learning opportunities [9]. Additionally, overreliance on GenAI may negatively impact learners' agency, critical thinking, and creativity, warranting caution [10].

Consequently, it is essential to balance technological advancement and human-centred values in learning. This perspective paper aims to delve into the promises and challenges of advancing human learning in the age of GenAI. By integrating human-centred theories of learning and instruction, we emphasise the importance of designing AI-driven educational tools that prioritise the needs of learners in contemporary societies. We elaborate on how this technology can transform learning and teaching practices while remaining critical of the ethical and practical challenges it poses, contributing to a future research agenda for investigating human-AI interaction and the adoption of GenAI as a tool for learning (Fig. 1).

2 Promises

GenAI promises to transform human learning by scaling personalised support, diversifying learning resources, enabling timely feedback, and innovating assessment methods. The realisation of these promises depends on the roles and interactions GenAI has with learners and educators (Fig. 2). Specifically, GenAI technologies can act as cognitive facilitators within learners' Zone of Proximal Development, providing adaptive support at scale. GenAI can also enrich learning experiences by assisting in the creation of diverse multimedia resources. In feedback, GenAI systems offer timely and multimodal insights, surpassing traditional methods in depth and efficiency. For

Learning impacts	Key components	Learners	Educators	Researchers	Policymakers	Technologists
Learning support	Scale personalised and adaptive support	●	●	●		●
	Learn within Zone of Proximal Development	●	●	●		
Learning resource	Efficient learning content creation		●	●		●
	Diverse multimedia learning resources		●	●		●
Learning feedback	Timely, specific, and constructive feedback		●	●		●
	Multimodal feedback delivery		●	●		●
Learning assessment	Adaptive and authentic assessments		●	●		●
	Enhanced real-world simulation		●	●		●
GenAI's imperfections	Hallucinations and inaccuracies			●		●
	Instability and biases			●		●
Ethical dilemma	Transparency and privacy concerns			●	●	●
	Equality and beneficence				●	●
Disruption of assessment	Human-AI hybrid cognition assessment		●	●		
	AI-induced performance illusion	●	●	●		
AI literacy	Basic understanding of AI systems	●	●	●	●	
	Critically evaluate AI-generated content	●	●	●	●	
Evidence-based decision making	Robust learning impacts evidence			●		
	Multi-party collaborative efforts	●	●	●	●	●
Methodological rigour	Evidence quality appraisal instruments			●		
	Long-term learning impact evaluations			●		

Learning impacts — Promises | Challenges | Needs. **Responsible stakeholders**: Learners, Educators, Researchers, Policymakers, Technologists.

Fig. 1 Overview of the impacts of generative artificial intelligence on human learning. The left side of the figure lists various learning impacts, which are categorised into promises (green), challenges (red), and needs (blue). The middle column presents key components associated with each learning impact. These components detail specific aspects that need to be addressed or leveraged to use generative AI as a tool for learning. The matrix on the right shows the five main groups involved in implementing these key components: learners, educators, researchers, policymakers, and technologists. The dots in each column indicate that the relevant group needs to make a substantive contribution to achieving the goals of the key component in the corresponding row.

assessment, GenAI enables adaptive and authentic evaluations using generative agents and multimodal models. The following sections explore each of these promises, illustrating their potential to transform the delivery, cultivation, and evaluation of human learning.

2.1 Learning Support

The unique contribution of GenAI, particularly LLMs, to learning support lies in its scalability and adaptability. GenAI can function as a master teacher at scale, providing personalised and adaptive support to a wide range of learners across various subjects and languages. Unlike conventional intelligent tutoring systems that require extensive knowledge engineering to design rule-based responses [11], GenAI can achieve superior and more naturalistic interactions, such as personalised feedback, adaptive questioning, and conversational engagement, with minimal prior training. These enhanced interactions facilitate more effective and intuitive tutoring experiences, making the learning process more engaging and tailored to individual student needs [8]. This capability holds the potential to democratise access to high-quality learning support, making it accessible to learners globally.
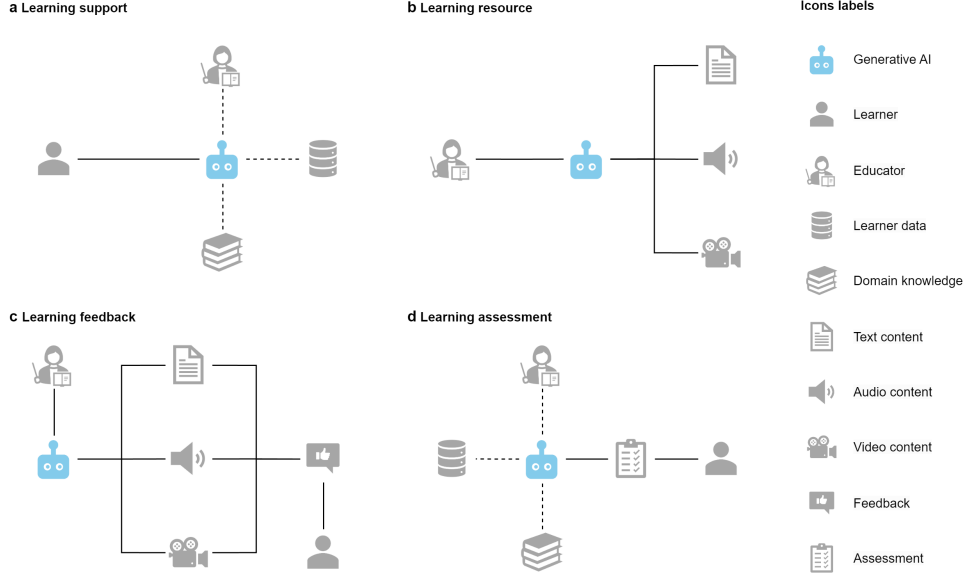
GenAI's role aligns with Vygotsky's sociocultural theory of learning, where more knowledgeable others guide learners within their Zone of Proximal Development [12]. By integrating novel technologies like ChatGPT into intelligent tutoring systems, GenAI can offer personalised and adaptive support based on each learner's unique knowledge [13]. These language models have demonstrated remarkable proficiency in processing semantic and contextual information [14], a key aspect of their effectiveness as a tool for learning. By accurately interpreting and responding to the linguistic and contextual nuances in learners' queries, LLMs ensure that the learning experience is interactive and thought-provoking. Rather than merely dispensing solutions, they can be used to encourage learners to engage cognitively with the material. This engagement is achieved by prompting students to think critically, unpack problems, and understand underlying concepts.

A representative case of how GenAI can support learning comes from Khan Academy's "Khanmigo" chatbot, powered by GPT-4 and designed to assist with mathematical queries [15]. Khanmigo exemplifies the shift from providing direct answers to a more nuanced, guided learning approach that offers constructive feedback and step-by-step instruction. For example, when students present Khanmigo with a problem on fractions, it guides them through the underlying concepts of denominators and numerators, encouraging them to apply these concepts to solve the problem through a series of guiding questions. Khanmigo functions as a facilitator, aligning with the principles of inquiry-based learning [16], a human-centred learning theory that emphasises the importance of active learning through inquiry. This theory encourages students to ask questions, explore, and engage deeply with the learning material to develop deep knowledge. This iterative methodology reflects Vygotsky's emphasis on the importance of the learning journey over the destination by fostering deep conceptual comprehension and retention [12, 16]. By engaging learners in a dialogic process, GenAI-driven systems such as Khanmigo aim to enhance learners' critical thinking and problem-solving skills.

Despite the promising design of systems like Khanmigo and students' positive attitudes towards using such technologies for personalised learning support [17], it is important to acknowledge the current limitations in empirical evidence regarding their short- and long-term impacts on learning outcomes [18]. Emerging evidence indicates that the impact on learning engagement, agency, and performance can paint a complicated and mixed picture (e.g., lack of learning gains after removing GenAI supports) [19, 20]. Therefore, further research is needed to substantiate GenAI's long-term benefits to human learning. This includes conducting longitudinal and randomised controlled studies that compare the effectiveness of GenAI tutoring with conventional rule-based tutoring systems over several academic terms and across different subjects to contextualise its impacts within various educational settings.

2.2 Learning Resource

Effective learning relies on the quality and diversity of resources, yet developing high-quality materials is often time-consuming and resource-intensive. GenAI promises to ease this burden by creating diverse and engaging content, fostering curriculum innovation and enhancing learning experiences. Studies on human-AI collaboration indicate

a Learning support **b** Learning resource

Icons labels

- Generative AI
- Learner
- Educator
- Learner data
- Domain knowledge
- Text content
- Audio content
- Video content
- Feedback
- Assessment

c Learning feedback **d** Learning assessment

Fig. 2 Examples of human-AI interactions in human learning. a, Learners receive personalised and adaptive support from generative AI tutors, which are co-designed with educators and have access to prior learner data and domain knowledge. **b,** Educators use generative AI to create multimodal learning resources, incorporating text, audio, and video content. **c,** Educators collaborate with generative AI to deliver multimodal feedback to learners. **d,** Generative AI agents use input requirements from educators, prior learner data, and domain knowledge to create assessment activities that evaluate learners.

that co-creating content with GenAI can meet diverse learning needs, providing students with relevant and accessible materials to support their individual paths efficiently and creatively [21–23]. For instance, early explorations have shown that GPT-4 can automatically generate instructional materials, such as explanations of programming concepts, examples, and quiz questions, thereby boosting learner engagement and satisfaction [24]. Additionally, GPT-4 has demonstrated proficiency in generating college-level biology questions for lower levels of Bloom's Taxonomy (e.g., remember and understand) but struggles with higher levels (e.g., apply and create) [25]. These findings suggest that while GenAI can produce learning resources, educators' expertise remains crucial for ensuring the accuracy, relevance, and pedagogical soundness of the material. This highlights the need for a human-AI collaborative approach to create meaningful resources that meet diverse learning objectives and learner needs.

GenAI can also enrich learning resources by generating interactive activities, multimedia content, and real-world problem-solving scenarios. Text-to-image models like Stable Diffusion, Midjourney, and DALL-E [26, 27] enable educators to create visual learning materials from textual content. These tools can foster students' creative thinking by engaging them in activities such as using AI to generate images. For instance, students can create imaginative visuals with AI and write inspired diaries based on these images, a practice found to reduce gender disparities in interest in art during Science, Technology, Engineering, the Arts and Mathematics (STEAM) classes [28].

This innovative approach has also been shown to enhance primary school students' extrinsic motivation, problem-solving awareness, critical thinking, and learning performance in ancient Chinese poetry [29]. Similarly, text-to-video generation tools like Runway's Gen-3 Aphla and OpenAI's Sora can support educators in creating video narratives from textual content, further diversifying learning modalities. This capability is particularly valuable for teaching students with specific disabilities, such as providing multisensory instruction to students with dyslexia [30]. A preliminary study found no significant differences in learning gains and perceived experiences between GenAI-generated videos with synthetic instructors and traditional recorded instructor videos, suggesting that GenAI could make high-quality learning resources more accessible globally [31].

By offering a range of pedagogical possibilities through efficient and diversified resource development, GenAI can help educators create more dynamic and engaging learning environments. This enables learners to interact with content in more informed, creative, and personalised ways. Such an approach aligns with constructivist learning principles, which emphasise the importance of learners actively constructing knowledge through exploration and interaction [32]. However, more research is needed to balance this integration of GenAI in developing resources for human learning [21, 33], such as determining the optimal level of automation versus human control, the extent of expert validation required, and the degree of alignment with learning objectives.

2.3 Learning Feedback

Another promise of GenAI in supporting human learning is its potential to provide timely, specific, and constructive feedback, a key element of high-quality instruction and essential for effective learning [34–36]. Providing detailed feedback regularly is laborious and time-consuming, adding to educators' workloads, especially since students perceive timely feedback as the most effective [37]. GenAI can assist by analysing student work and delivering instant, personalised feedback with minimal prior training. For example, a recent study found that ChatGPT generates more in-depth and fluent feedback, coherently summarising students' performances compared to human educators [3]. This AI-generated feedback also includes process-focused elements, which are more effective in shaping learning strategies [36].

Emerging studies show similar benefits in various learning contexts, such as formative feedback in secondary school essay writing [38, 39], programming assignments in introductory computer science courses [40, 41], and collaborative second language writing [42]. GenAI-generated feedback has led to enhanced task performance and positive experiences [39, 40, 42]. Additionally, chatbots powered by GenAI models with natural and visual language understanding capabilities (e.g., GPT-4 with Vision and Gemini 1.0 Pro) can help students navigate and comprehend insights from learning analytics dashboards [43], which combine data, analytics, and visualisations to provide feedback on learning processes and outcomes [44]. These chatbots could facilitate dialogic feedback, which is associated with improved learner productivity and engagement [45].

GenAI could also expand feedback delivery beyond text and graphics to include narrated audio and video, addressing the scalability challenges of these formats and

leveraging their benefits for enhanced feedback efficiency and student engagement [46]. For example, by combining 3D diffusion models [47] and text-to-speech models [48], educators can create digital avatars to convey feedback through a narrated voice rather than text alone. This diversity in feedback modalities can increase engagement and effectiveness [46]. Prior research indicates that audio and video feedback is often perceived as more personal and dynamic, enhancing understanding and engagement compared with traditional written feedback [49–51]. The integration of GenAI technologies promises to facilitate timely and multimodal feedback, providing more informative feedback and fostering improved effectiveness and engagement in the learning process. However, it is essential to evaluate the cost and benefits, as these models, especially video generation models, require high computational power, potentially widening the inequality in learning opportunities.

2.4 Learning Assessment

GenAI is transforming the assessment of learning, shifting from traditional, often onerous methods to more adaptive and authentic processes [52]. Central to this shift is GenAI's potential to create personalised and adaptive assessments, enhancing the understanding of each student's needs and progression. This is enabled by advancements in generative agents – autonomous and adaptive AI entities that operate independently, pursuing objectives without continuous user interaction, as exemplified by tools like AutoGen (preprint) [53]. These agents exhibit human-like cognitive and metacognitive abilities, including task planning, situational assessment, progress monitoring, and collaborative efforts among agents. For instance, a group of 25 generative agents in a dynamic sandbox environment successfully organised and conducted a Valentine's Day celebration based on a single user input [54]. By leveraging a similar agent architecture, encompassing observation, planning, and reflection, and integrating these with process-centred methodologies (e.g., modelling self-regulated learning from learners' digital traces [55]) from the field of learning analytics [56, 57], learning scientists and researchers can develop generative agents capable of autonomously evaluating human learning. These agents can identify areas of knowledge deficiency and provide tailored learning resources and adaptive assessments.

Recent educational technology studies highlight the potential of automated assessments through multi-agent frameworks that leverage multiple LLM agents. These GenAI systems are being used to grade coding assignments in online learning [58], conduct cognitive assessments to identify students' strengths and weaknesses according to Bloom's taxonomy in e-learning environments [59], and assess educators' mathematical content knowledge for professional development programs [60]. These applications demonstrate strong potential for generalisability, precision, and dependability.

GenAI also holds promise for advancing authentic assessments [52]. It can enhance learning tasks in both virtual and physical simulations to more accurately mirror real-world situations, making assessments more meaningful and contextualised. Previous studies have shown the effectiveness of combining LLMs with knowledge graphs to create virtual standard patients, aiding the training and evaluation of medical students' diagnostic skills [61]. Knowledge graphs are structured representations that integrate diverse data sources, providing a comprehensive understanding of a domain [62]. When

7

used with LLMs, they can simulate complex learning and assessment scenarios requiring critical thinking and problem-solving skills, such as in driving education [63], programming education [64], and laboratory safety courses [65]. Integrating multimodal generative models, such as GPT-4 Vision for text and image generation, Meta's Voicebox for audio creation from text, and generative adversarial networks for digital avatar production, can further enhance the authenticity of simulated assessment environments [66]. These enhancements allow students to interact naturally and perform procedural actions as if they were in real professional settings, a concept proven effective in virtual internships [67] and healthcare simulations [68]. However, much effort is required to develop valid and reliable behavioural and psychological indicators in these novel assessment settings to accurately capture genuine human learning.

3 Challenges

Amidst GenAI's promises, formidable challenges confront learners and educators alike and raise critical moral and ethical concerns about integrating such technology into human learning. These challenges involve GenAI technologies' imperfections, the ethical dilemmas of transparency, privacy, equality, and beneficence, and the disruption of assessment practices. The following sections elaborate on each of these challenges.

3.1 GenAI's Imperfections

As GenAI technologies become increasingly integrated into learning support, resource generation, feedback, and assessment, it is imperative to address the risks posed by hallucinations [69]. Hallucinations occur when there are mismatches in training data or complexities in language generation tasks, resulting in outputs that may not align with factual information [70]. The probabilistic nature of LLMs and diffusion models further limits their utility due to inherent instabilities and potential biases in their training data [71]. For instance, ChatGPT often fails tasks easily solved by humans, such as reasoning tasks requiring real-world knowledge, logic, mathematical calculations, and distinguishing between factual and fictive information. Consequently, it sometimes provides fabricated facts [72] (preprint). These inaccuracies can undermine GenAI's reliability as a learning tool, potentially outweighing its promises (Section 2).

Emerging studies indicate that hallucinations in GenAI can occur with non-negligible frequency, increasing with the complexity and specificity of queries posed to the AI [70]. GenAI may perform reasonably well with generic questions (e.g., What are Newton's laws of motion?) but is more prone to errors with nuanced, context-specific, time-sensitive, or highly technical information [73]. The lack of transparency in GenAI's decision-making process complicates identifying when and why these hallucinations occur [70, 74]. Relying solely on GenAI for learning content creation and curriculum development without validation could introduce inaccuracies, misleading both educators and students. Similarly, GenAI-generated feedback or assessments based on incorrect information could misguide students' learning processes, leading to misconceptions or a lack of understanding of key concepts.

Addressing these challenges requires an interdisciplinary effort. Educators should adopt a balanced and proactive approach, teaching learners to critically evaluate AI-generated content by cross-referencing with reliable sources, questioning plausibility, and recognising signs of hallucination. These steps are essential for cultivating AI literacy [75], as discussed further in Section 4.1. Additionally, designing and optimising the interface of educational technologies to highlight potential hallucinations requires collaboration among learning scientists, human-computer interaction researchers, and technology providers [74, 76]. Such a collaborative approach is essential to empower learners to deal with the imperfections of GenAI both intrinsically, by developing critical thinking skills, and extrinsically, by leveraging improved technological interfaces that signal potential inaccuracies.

3.2 Ethical Dilemmas

Adopting GenAI to support human learning raises several ethical issues, notably in areas such as transparency, privacy, equality, and beneficence. A key concern is the transparency of AI-generated solutions, as highlighted in a recent systematic literature review [2]. The review found that a majority (92%) of GenAI tools currently used for supporting learning practices, particularly those based on LLMs, are transparent only to AI experts, not to educators, students, or other stakeholders. This lack of transparency is problematic as it obscures the understanding of AI functionalities and potential flaws from those directly impacted by these technologies [77]. The primary cause of this transparency gap is the absence of human-in-the-loop elements in prior research, such as involving educators and students in the development and evaluation of GenAI-powered educational technologies. This aligns with the growing emphasis on developing explainable and human-centred AI, underscoring the essential role of stakeholder involvement in crafting impactful and meaningful educational technologies [78, 79].

To achieve personalisation in learning support, resource generation, feedback, and assessment using GenAI, learners' personal data must be provided to these models. However, privacy concerns can reduce learner participation [80, 81]. These concerns are prominent due to the lack of clear consent strategies and data protection measures surrounding GenAI in supporting human learning [2]. Using learner-generated data without explicit consent or adequate anonymisation raises serious issues about exposing sensitive information [82]. For instance, researchers conducted a divergence attack on ChatGPT, compromising its security and causing it to output original training data containing personally identifiable information [83] (preprint). Although OpenAI has addressed this vulnerability, potential data breaches from unforeseen attacks remain a concern [84, 85]. This issue is particularly troubling given the resources required for GenAI to unlearn information once private data has been used for model training, especially for large, commercial, and proprietary models [84].

Regarding equality, there is an evident disparity in language representation and accessibility of GenAI models. While advancements have been made in non-English languages for LLMs and speech diffusion models [14, 48], the predominance of English-based AI solutions perpetuates a bias towards Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies [2]. This imbalance raises concerns about the

global applicability and fairness of these technologies, potentially intensifying existing inequalities and the digital divide in learning opportunities [86].

Finally, beneficence is a critical ethical principle that must be addressed. Several studies highlight the risks of underperforming or biased AI models, which can negatively impact human learning and perpetuate systemic biases, such as gender, racial, and social class biases [87, 88]. Strategies like balanced sampling and cautionary labelling have been proposed [89, 90], but the opaque nature of many generative models makes ensuring fairness and accuracy challenging, potentially violating the principle of beneficence [91]. While model alignment is often implemented to prevent GenAI from producing toxic content, recent evidence suggests that adversarial attacks using specific prompts can undermine these measures [85]. Such attacks could facilitate cheating, promote biased views, or expose students to offensive language [92]. These issues could disrupt learning, compromise safety and inclusivity, and cause psychological harm, eroding trust in educational technologies. These ethical challenges underscore the need for rigorous and multifaceted ethical considerations in deploying GenAI and the urgency of establishing regulations, such as the EU AI Act [93].

3.3 Disruption of Assessment

GenAI poses significant challenges to conventional learning assessment methodologies. Traditionally, assessments have focused on evaluating learning products, such as essays, to measure outcomes [52]. However, GenAI's ability to produce high-quality, human-like responses calls into question the validity of these approaches [94]. A central issue is distinguishing between a learner's work and AI-generated output. GenAI, particularly LLMs like ChatGPT and Llama 3, can generate responses that closely mimic human reasoning and writing styles, making it difficult to discern the origin of the work [4].

A performance paradox arises when tasks are completed with GenAI assistance. A recent randomised controlled experiment found that while GenAI tools can help students achieve better performance, removing this support significantly lowers their performance [19]. This suggests that GenAI may create an illusion of improved learning without developing essential skills, such as self-regulated learning. Thus, we must ask: Who and what are we actually assessing? This dilemma extends beyond detecting AI-generated content to reconsidering the purpose of assessment in learning.

The challenge is further compounded when considering the learning process itself. GenAI's ability to interact with computational systems means even the learning process can be imitated or augmented by AI. Preliminary work on multimodal GenAI agents [95] (preprint) has shown these agents can operate smartphone applications, generating digital trace data while executing user requests. This AI-generated data could impede existing learning analytic methods that rely on such data to model the learning process [96]. This issue blurs the line between human cognition and AI-augmented cognition [97], complicating the assessment of skills traditionally seen as exclusively human, such as critical thinking, problem-solving, and creativity [94].

Consequently, we must reconsider the purpose of learning assessment across different educational stages. Assessing human cognition and metacognition remains essential for K-12 education, as young learners continue developing fundamental skills. In higher

education, prioritising the evaluation of human-AI hybrid cognition and metacognition could be crucial for preparing learners for an AI-integrated workforce [98]. This shift demands rethinking assessment strategies to accommodate the collaborative nature of learning in the presence of AI.

4 Needs

Within GenAI's promises and challenges, three pivotal needs must be addressed for effective integration into human learning: cultivating AI literacy among learners and educators, prioritising evidence-based decision-making, and ensuring methodological rigour in research using GenAI. These needs aim to foster a balanced integration that enhances human abilities and ensures a synergistic relationship between GenAI and human development.

4.1 AI Literacy

Cultivating AI literacy is essential to ensuring the effective, responsible, and ethical use of GenAI technologies to support human learning [75, 99]. This need extends beyond learners to include educators, policymakers, and administrators, who are integral to the design, delivery, and facilitation of learning experiences. AI literacy encompasses a basic understanding of how AI systems function but also an intimate awareness of their potential impact, ethical considerations, and limitations [75]. The absence of AI literacy can lead to severe consequences. For instance, the New York Times reported that a lawyer using ChatGPT for a court filing was unaware of fabricated citations generated by the AI, resulting in a breach of professional ethics and legal standards [100]. One must ask: What if educators unknowingly provided students with AI-generated learning resources that contained fabricated content? Such actions could erode trust and integrity in education systems, misleading students and compromising their learning quality.

These concerns highlight the critical need to cultivate AI literacy. A recent study indicates that human users often prefer AI-generated content for its comprehensiveness and well-articulated language style, despite its inaccuracies [101]. As GenAI's propensity to hallucinate remains challenging to address at the foundational model level [70], understanding its limitations and identifying potential pitfalls will be crucial for preparing individuals to live, learn, and work with GenAI in the 21st century. This requires adopting AI literacy models, practices for their development, and measurement approaches. Institutions, policymakers, and researchers must focus on AI literacy as a key learning objective to ensure that educators, students, administrators, and even parents are not merely consumers of AI technology but also informed participants in its evolution and application.

4.2 Evidence-Based Decision Making

The integration of GenAI into human learning promises to enhance experiences and outcomes (as highlighted in Section 2). However, adopting these technologies requires a commitment to evidence-based decision-making. This necessitates a collaborative

effort among researchers, practitioners, and policymakers to generate robust evidence guiding the effective and responsible use of AI in learning practices. By working together, these stakeholders can ensure GenAI deployment aligns with learning goals and supports the development of essential cognitive and metacognitive skills.

Encouraging the use of GenAI to support human learning requires a nuanced understanding of its benefits and limitations. For instance, while GenAI can improve the efficiency of information processing and retrieval, there is a risk of fluency bias, where learners may overestimate their understanding due to the ease of cognitive information processing [102, 103] (preprint). Similarly, reliance on GenAI for creative and problem-solving tasks could weaken these critical skills, fostering a dependency that may hinder innovation and original thought [104–106].

To mitigate these risks and maximise GenAI's benefits, it is imperative to foster partnerships among researchers, practitioners, and policymakers. These collaborations can produce evidence that informs learning and teaching practices, ensuring that GenAI enhances rather than replaces human cognitive, metacognitive, and creative processes. By prioritising evidence-based decision-making and stakeholder collaboration, we can leverage GenAI's advantages in educational environments while promoting deep learning, creativity, and problem-solving abilities among learners.

4.3 Methodological Rigour

Building on discussions of evidence-based decision-making, it is crucial to emphasise methodological rigour in applying GenAI technologies within human learning research. As these technologies evolve, human learning researchers and scientists must adapt and refine their methodologies to accurately assess the impact of these tools on teaching and learning processes. GenAI's capabilities, such as passing the United States Medical Licensing Exam [107], completing exams at the University of Minnesota Law School [108], and solving queries from Wharton School of Business tests [109], underscore its potential. However, the excitement must be tempered with caution to avoid overestimating effectiveness due to methodological shortcomings. A notable example is a preprint study claiming GPT-4, with prompt engineering, could achieve perfect scores in the MIT Mathematics, Electrical Engineering, and Computer Science curriculum [110]. This study [110], initially attracting widespread attention, was later retracted due to methodological concerns, including data set contamination, over-reliance on GPT-4 for accuracy assessment, and ambiguities in manual verification of results [111] (preprint). This incident underscores the need for rigorous methodological standards, likely requiring new approaches in evaluating GenAI technologies.

To address these challenges, it is essential to establish standards for appraising the quality of evidence on GenAI's impact on learning processes, outcomes, and experiences [18]. In the medical field, tools such as the Cochrane Risk of Bias Tool and ROBINS-I are used to assess study quality. Given the distinct methodological requirements introduced by GenAI, including various prompting engineering strategies and retrieval generation techniques, it is crucial to establish specific quality standards and evaluation tools. These requirements go beyond conventional methodologies used in human learning research. For example, using GenAI to generate physics practice questions might involve retrieval methods that limit the AI to sourcing content solely

on Newton's laws of motion and crafting prompts specifying complexity level, target student grade, and desired question format (e.g., multiple-choice, short answer, or problem-solving). By working collaboratively, the human learning research community can create a robust framework for evaluating evidence, ensuring a solid foundation for future policies and practices. This effort will enable researchers, practitioners, and policymakers to build on reliable, valid, and generalisable findings, fostering the responsible and effective integration of GenAI technologies into learning.

5 Conclusion and Future Directions

As we look toward the next decade, powerful AI tools are set to become integral to our society, transforming how we learn, work, and live [112]. GenAI technologies could permeate every aspect of human learning. Imagine students collaborating with AI agents designed to mimic certain personality traits to help students learn about leadership and teamwork, engaging in debates with digital twins of Socrates, Plato, and Aristotle to explore ancient Greek philosophy, learning impressionist painting techniques from a humanoid robotic mentor modelled after Claude Monet, and visualising Einstein's special theory of relativity in virtual realities. All this could occur while receiving personalised support from a GenAI tutor hosted on a wearable device. This integration necessitates a dual approach to learning: educating ourselves both about and with GenAI, while continuing to develop critical thinking, problem-solving, self-regulation, and reflective thinking skills. These skills are crucial for maintaining cognitive and metacognitive autonomy as AI becomes embedded in our daily lives.

Understanding the relationship between GenAI and human cognition, metacognition, and creativity is essential for maximising its potential as a learning tool. This understanding will enhance the effectiveness of AI-driven educational tools and ensure human ingenuity is preserved amidst technological advancement. Key research questions include: How can we promote human-AI interaction to maximise learner agency? What behavioural indicators can reliably capture cognitive and metacognitive processes during AI-assisted learning? How can we assess learning to reflect genuine knowledge and skill development rather than an AI-created performance illusion? What strategies can prevent over-reliance on AI, ensuring humans remain primary agents of critical thinking and problem-solving?

Educators are pivotal in integrating AI tools to enhance traditional teaching methods. We anticipate a shift in educators' roles, with GenAI reducing the burden of knowledge dissemination, allowing teachers to focus on deeper connections with students as mentors and facilitators. This transition requires educators to adopt new pedagogical paradigms that leverage AI to foster intellectual and emotional growth. They must become proficient in AI literacy, effectively integrate AI tools into their teaching, and remain vigilant about potential pitfalls, such as GenAI's imperfections and the risk of student over-reliance on AI. Balancing AI use with activities promoting human creativity, critical thinking, and social interaction is crucial to ensure AI augments rather than replaces human educators. Educational institutions must invest in ongoing professional development and support systems to help teachers manage techno-stress and workload burdens from adopting new technologies.

Policymakers and technology companies should consider: How can we ensure accountability for AI tools used in human learning, and who should be responsible for their outcomes? What ethical guidelines should govern AI tools in educational settings? How can we design and implement AI learning tools to promote equality and inclusivity?

We argue that human-centred theories of learning and instruction must be integrated with GenAI to ensure these technologies enhance rather than detract from human learning. This involves developing AI systems that support and elevate human cognitive capacities. By fostering a learning environment that harmonises technology with theoretical approaches, we can promote personal growth and the development of adaptive skills and knowledge needed to navigate the rapid changes in the age of AI. A united effort among researchers, policymakers, technology companies, and educators is essential to fully leverage GenAI's potential in advancing human learning. By addressing these critical questions and considerations, we can ensure that GenAI becomes a powerful ally in the pursuit of knowledge and innovation, rather than a crutch that undermines our intellectual abilities.

References

[1] Gašević D, Siemens G, Sadiq S. Empowering learners for the age of artificial intelligence. Computers and Education: Artificial Intelligence. 2023;4:100130.

[2] Yan L, Sha L, Zhao L, Li Y, Martinez-Maldonado R, Chen G, et al. Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology. 2023;00:1–23.

[3] Dai W, Lin J, Jin H, Li T, Tsai YS, Gašević D, et al. Can large language models provide feedback to students? A case study on ChatGPT. In: 2023 IEEE International Conference on Advanced Learning Technologies (ICALT). IEEE; 2023. p. 323–325.

[4] Li Y, Sha L, Yan L, Lin J, Raković M, Galbraith K, et al. Can large language models write reflectively. Computers and Education: Artificial Intelligence. 2023;4:100140.

[5] Yildirim-Erbasli SN, Bulut O. Conversation-based assessment: A novel approach to boosting test-taking effort in digital formative assessment. Computers and Education: Artificial Intelligence. 2023;4:100135.

[6] Mazzoli CA, Semeraro F, Gamberini L. Enhancing Cardiac Arrest Education: Exploring the potential use of MidJourney. Resuscitation. 2023;189.

[7] Vartiainen H, Tedre M. Using artificial intelligence in craft education: crafting with text-to-image generative models. Digital Creativity. 2023;34(1):1–21.

[8] Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? On opportunities and challenges of large language models

Table 1 Defining human learning concepts

Human learning. Human learning is the process through which individuals acquire new knowledge, skills, attitudes, or values.
Education. Education is the structured process of teaching and learning, typically occurring in institutional settings such as schools, universities, and training programs. It aims to develop learners' intellectual, social, and emotional capacities, preparing them for personal and professional success.
Constructivist learning. Constructivist learning is a theory that posits learners construct their own understanding and knowledge of the world through experiences and reflecting on those experiences. It emphasises active engagement, exploration, and the application of knowledge in meaningful contexts.
Inquiry-based learning. Inquiry-based learning refers to pedagogical approaches in which learners use scientific methods to build knowledge through formulating hypotheses, conducting experiments, and making observations to discover causal relations. It emphasises problem-solving, active participation, and self-directed learning, facilitating inductive and deductive reasoning. The knowledge gained is usually new to the learner, fostering deep understanding through exploration.
The zone of proximal development. The zone of proximal development represents the gap between what a learner can do on their own and what they can achieve with help from a skilled guide. It encompasses tasks that are currently beyond the learner's independent capabilities but can be completed with assistance and guidance.
Bloom's taxonomy. Bloom's Taxonomy is a framework used to categorise learning goals and objectives and provides a structured approach to designing learning content and assessing learning outcomes. It classifies cognitive skills into a hierarchy from basic to advanced: remember, understand, apply, analyse, evaluate, and create.
Learning analytics. Learning analytics involves the collection, measurement, and analysis of data about learners and their contexts to understand and optimise learning and the environments in which it occurs. It uses data-driven insights to inform educational decision-making and enhance learning outcomes.
Feedback. Feedback refers to responses related to a learner's performance or understanding, information to them about the correctness of task solutions or providing them with content-related or strategic assistance and information about their processing. The German psychologist Lipowsky identifies feedback as one of the nine hallmarks that characterise high-quality instruction.
Authentic assessment. Authentic assessment refers to evaluation methods that require learners to apply their skills and knowledge to real-world tasks and problems. It aims to assess learners' abilities in contexts that are relevant and meaningful, providing a more accurate measure of their competencies.
Intelligent tutoring system. An intelligent tutoring system is a computer-based system or tool created to mimic human tutoring. It provides learners with immediate, personalised instruction or feedback, often functioning autonomously without requiring direct intervention from a human teacher.
Digital twins. Digital twins are virtual replicas of physical entities, such as objects, systems, or processes. In education, digital twins can simulate real-world scenarios, providing learners with immersive and interactive experiences to enhance understanding and skill development.

Table 2 Glossary of artificial intelligence terms

Generative artificial intelligence. Generative artificial intelligence refers to AI systems designed to create new content, such as text, images, or music, by learning patterns from existing data. These systems can produce outputs that are novel and relevant, often mimicking human creativity.
Large language model (LLM). An LLM is a computational model known for its ability to perform general-purpose language generation and various natural language processing tasks like classification. LLMs acquire these abilities by learning statistical relationships from vast text datasets during an intensive self-supervised and semi-supervised training process. They can generate text by taking inputs and repeatedly predicting the next word, a form of generative AI.
Diffusion model. A diffusion model is a type of probabilistic model that generates data by simulating the gradual transformation of noise into coherent data points. These models use a series of iterative steps to refine random noise into structured outputs, such as images or text. Diffusion models have shown promise in generating high-quality, realistic data and are used in applications like image synthesis, text generation, and other creative tasks.
Knowledge graph. A knowledge graph is a structured representation of information that captures relationships between various entities, such as objects, events, or concepts. It organises data into nodes (representing entities) and edges (representing relationships between entities), enabling complex queries and inferences. Knowledge graphs are used in applications like search engines, recommendation systems, and natural language understanding to provide context-aware and semantically rich insights. They help in connecting and integrating diverse data sources, thus enhancing the ability of AI systems to understand and reason about the world.
AI Agent. An AI agent is an autonomous and adaptive AI entity that operates independently, pursuing objectives without continuous user interaction.
AI literacy. AI literacy involves understanding the fundamental concepts and capabilities of artificial intelligence, as well as its implications and ethical considerations. It encompasses the skills needed to interact with AI systems effectively and to critically evaluate their outputs.
Hallucinations. In the context of AI, hallucinations refer to instances where generative models produce outputs that are incorrect, nonsensical, or fabricated. These errors can occur due to the model's limitations or biases in the training data.
Divergence attack. A divergence attack is a method used to exploit weaknesses in AI models, causing them to deviate from their intended behaviour. This can result in the model generating harmful or unintended outputs, potentially exposing sensitive information or producing biased content.
Model alignment. Model alignment involves ensuring that AI systems' behaviours and outputs are consistent with human values and intended goals. It includes efforts to make AI systems safe, reliable, and ethical in their operations.

for education. Learning and individual differences. 2023;103:102274.

[9] Pontual Falcão T, Ferreira Mello R, Lins Rodrigues R.: Applications of learning analytics in Latin America. Wiley Online Library.

[10] Darvishi A, Khosravi H, Sadiq S, Gašević D, Siemens G. Impact of AI assistance on student agency. Computers & Education. 2023;p. 104967.

[11] Mousavinasab E, Zarifsanaiey N, R Niakan Kalhori S, Rakhshan M, Keikha L, Ghazi Saeedi M. Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. Interactive Learning Environments. 2021;29(1):142–163.

[12] Vygotsky LS, Cole M. Mind in society: Development of higher psychological processes. Harvard university press; 1978.

[13] Joksimovic S, Ifenthaler D, Marrone R, De Laat M, Siemens G. Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review. Computers and Education: Artificial Intelligence. 2023;p. 100138.

[14] Chang Y, Wang X, Wang J, Wu Y, Yang L, Zhu K, et al. A survey on evaluation of large language models. ACM Transactions on Intelligent Systems and Technology. 2024;15(3):1–45.

[15] Academy K.: Meet Khanmigo: Meet Khanmigo: Khan Academy's AI-powered teaching assistant & tutor. Accessed: 2023-12-14. https://www.khanmigo.ai/.

[16] Lee VS. What is inquiry-guided learning? New directions for teaching and learning. 2012;2012(129):5–14.

[17] Chan CKY, Hu W. Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education. 2023;20(1):43.

[18] Hennessy S, Cukurova M, Lewin C, Mavrikis M, Major L.: BJET Editorial 2024: A call for research rigour. Wiley Online Library.

[19] Darvishi A, Khosravi H, Sadiq S, Gašević D, Siemens G. Impact of AI assistance on student agency. Computers & Education. 2024;210:104967.

[20] Nie A, Chandak Y, Suzara M, Malik A, Woodrow J, Peng M, et al. The GPT Surprise: Offering Large Language Model Chat in a Massive Coding Class Reduced Engagement but Increased Adopters' Exam Performances. Center for Open Science; 2024.

[21] Molenaar I. Towards hybrid human-AI learning technologies. European Journal of Education. 2022;57(4):632–645.

[22] Ji H, Han I, Ko Y. A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. Journal of Research on Technology in Education. 2023;55(1):48–63.

[23] Yang KB, Lawrence L, Echeverria V, Guo B, Rummel N, Aleven V. Surveying teachers' preferences and boundaries regarding human-AI control in dynamic pairing of students for collaborative learning. In: Technology-Enhanced Learning for a Free, Safe, and Sustainable World: 16th European Conference on Technology Enhanced Learning, EC-TEL 2021, Bolzano, Italy, September 20-24, 2021, Proceedings 16. Springer; 2021. p. 260–274.

[24] Pesovski I, Santos R, Henriques R, Trajkovik V. Generative ai for customizable learning experiences. Sustainability. 2024;16(7):3034.

[25] Hwang K, Wang K, Alomair M, Choa FS, Chen LK. Towards Automated Multiple Choice Question Generation and Evaluation: Aligning with Bloom's Taxonomy. In: International Conference on Artificial Intelligence in Education. Springer; 2024. p. 389–396.

[26] Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. Learning transferable visual models from natural language supervision. In: International conference on machine learning. PMLR; 2021. p. 8748–8763.

[27] Chiu TK. The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney. Interactive Learning Environments. 2023;p. 1–17.

[28] Lee U, Han A, Lee J, Lee E, Kim J, Kim H, et al. Prompt Aloud!: Incorporating image-generative AI into STEAM class with learning analytics using prompt data. Education and Information Technologies. 2023;p. 1–31.

[29] Chen Y, Zhang X, Hu L. A progressive prompt-based image-generative AI approach to promoting students' achievement and perceptions in learning ancient Chinese poetry. Educational Technology & Society. 2024;27(2):284–305.

[30] Long L, MacBlain S, MacBlain M. Supporting students with dyslexia at the secondary level: An emotional model of literacy. Journal of Adolescent & Adult Literacy. 2007;51(2):124–134.

[31] Leiker D, Gyllen AR, Eldesouky I, Cukurova M. Generative AI for learning: investigating the potential of learning videos with synthetic virtual instructors. In: International conference on artificial intelligence in education. Springer; 2023. p. 523–529.

[32] Bada SO, Olusegun S. Constructivism learning theory: A paradigm for teaching and learning. Journal of Research & Method in Education. 2015;5(6):66–70.

[33] Tavakoli M, Faraji A, Molavi M, T Mol S, Kismihók G. Hybrid human-AI curriculum development for personalised informal learning environments. In: LAK22: 12th International Learning Analytics and Knowledge Conference; 2022. p. 563–569.

[34] Pardo A, Jovanovic J, Dawson S, Gašević D, Mirriahi N. Using learning analytics to scale the provision of personalised feedback. British Journal of Educational Technology. 2019;50(1):128–138.

[35] Lim LA, Gentili S, Pardo A, Kovanović V, Whitelock-Wainwright A, Gašević D, et al. What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course. Learning and Instruction. 2021;72:101202.

[36] Hattie J, Timperley H. The power of feedback. Review of educational research. 2007;77(1):81–112.

[37] Poulos A, Mahony MJ. Effectiveness of feedback: The students' perspective. Assessment & Evaluation in Higher Education. 2008;33(2):143–154.

[38] Steiss J, Tate T, Graham S, Cruz J, Hebert M, Wang J, et al. Comparing the quality of human and ChatGPT feedback of students' writing. Learning and Instruction. 2024;91:101894.

[39] Meyer J, Jansen T, Schiller R, Liebenow LW, Steinbach M, Horbach A, et al. Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. Computers and Education: Artificial Intelligence. 2024;6:100199.

[40] Zhang Z, Dong Z, Shi Y, Price T, Matsuda N, Xu D. Students' perceptions and preferences of generative artificial intelligence feedback for programming. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38; 2024. p. 23250–23258.

[41] Liang Z, Sha L, Tsai YS, Gašević D, Chen G. Towards the Automated Generation of Readily Applicable Personalised Feedback in Education. In: International Conference on Artificial Intelligence in Education. Springer; 2024. p. 75–88.

[42] Wiboolyasarin W, Wiboolyasarin K, Suwanwihok K, Jinowat N, Muenjanchoey R. Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. Computers and Education: Artificial Intelligence. 2024;6:100228.

[43] Yan L, Zhao L, Echeverria V, Jin Y, Alfredo R, Li X, et al. VizChat: Enhancing Learning Analytics Dashboards with Contextualised Explanations Using Multimodal Generative AI Chatbots. In: International Conference on Artificial Intelligence in Education. Springer; 2024. p. 180–193.

[44] Matcha W, Gašević D, Pardo A, et al. A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. IEEE transactions on learning technologies. 2019;13(2):226–245.

[45] Yang M, Carless D. The feedback triangle and the enhancement of dialogic feedback processes. Teaching in Higher Education. 2013;18(3):285–297.

[46] Dawson P, Henderson M, Ryan T, Mahoney P, Boud D, Phillips M, et al. Technology and feedback design. In: Learning, design, and technology: An international compendium of theory, research, practice, and policy. Springer; 2023. p. 695–739.

[47] Wang T, Zhang B, Zhang T, Gu S, Bao J, Baltrusaitis T, et al. Rodin: A generative model for sculpting 3d digital avatars using diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023. p. 4563–4573.

[48] Le M, Vyas A, Shi B, Karrer B, Sari L, Moritz R, et al. Voicebox: Text-guided multilingual universal speech generation at scale. Advances in neural information processing systems. 2024;36.

[49] McCarthy J. Evaluating written, audio and video feedback in higher education summative assessment tasks. Issues in Educational Research. 2015;25(2):153–169.

[50] Orlando J. A comparison of text, voice, and screencasting feedback to online students. American Journal of Distance Education. 2016;30(3):156–166.

[51] Henderson M, Phillips M. Video-based feedback on student assessment: Scarily personal. Australasian Journal of Educational Technology. 2015;31(1).

[52] Swiecki Z, Khosravi H, Chen G, Martinez-Maldonado R, Lodge JM, Milligan S, et al. Assessment in the age of artificial intelligence. Computers and Education: Artificial Intelligence. 2022;3:100075.

[53] Wu Q, Bansal G, Zhang J, Wu Y, Zhang S, Zhu E, et al. Autogen: Enabling next-gen llm applications via multi-agent conversation framework. arXiv preprint arXiv:230808155. 2023;.

[54] Park JS, O'Brien J, Cai CJ, Morris MR, Liang P, Bernstein MS. Generative agents: Interactive simulacra of human behavior. In: Proceedings of the 36th annual acm symposium on user interface software and technology; 2023. p. 1–22.

[55] Fan Y, van der Graaf J, Lim L, Raković M, Singh S, Kilgour J, et al. Towards investigating the validity of measurement of self-regulated learning based on trace data. Metacognition and Learning. 2022;17(3):949–987.

[56] Allen LK, Creer SC, Öncel P. Natural Language Processing as a Tool for Learning Analytics - Towards a Multi-Dimensional View of the Learning Process. In: Lang C, Siemens G, Wise AF, Gašević D, Merceron A, editors. The Handbook of Learning Analytics. 2nd ed. SoLAR; 2022. p. 46–53.

[57] Gašević D, Greiff S, Shaffer DW.: Towards strengthening links between learning analytics and assessment: Challenges and potentials of a promising new bond. Elsevier.

[58] Lagakis P, Demetriadis S. EvaAI: A Multi-agent Framework Leveraging Large Language Models for Enhanced Automated Grading. In: International Conference on Intelligent Tutoring Systems. Springer; 2024. p. 378–385.

[59] Shahzad R, Aslam M, Al-Otaibi S, Javed MS, Khan AR, Bahaj SA, et al. Multi-Agent System for Students Cognitive Assessment in E-Learning Environment. IEEE Access. 2024;.

[60] Yang K, Chu Y, Darwin T, Han A, Li H, Wen H, et al. Content Knowledge Identification with Multi-agent Large Language Models (LLMs). In: International Conference on Artificial Intelligence in Education. Springer; 2024. p. 284–292.

[61] Song W, Hou X, Li S, Chen C, Gao D, Sun Y, et al. An Intelligent Virtual Standard Patient for Medical Students Training Based on Oral Knowledge Graph. IEEE Transactions on Multimedia. 2022;.

[62] Ji S, Pan S, Cambria E, Marttinen P, Philip SY. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE transactions on neural networks and learning systems. 2021;33(2):494–514.

[63] Rehm J, Reshodko I, Børresen SZ, Gundersen OE. The Virtual Driving Instructor: Multi-Agent System Collaborating via Knowledge Graph for Scalable Driver Education. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38; 2024. p. 22806–22814.

[64] Jin H, Lee S, Shin H, Kim J. Teach AI How to Code: Using Large Language Models as Teachable Agents for Programming Education. In: Proceedings of the CHI Conference on Human Factors in Computing Systems; 2024. p. 1–28.

[65] Yang QF, Lian LW, Zhao JH. Developing a gamified artificial intelligence educational robot to promote learning effectiveness and behavior in laboratory safety courses for undergraduate students. International journal of educational technology in higher education. 2023;20(1):18.

[66] Thanh BN, Vo DTH, Nhat MN, Pham TTT, Trung HT, Xuan SH. Race with the machines: Assessing the capability of generative AI in solving authentic assessments. Australasian Journal of Educational Technology. 2023;39(5):59–81.

[67] Chesler NC, Ruis A, Collier W, Swiecki Z, Arastoopour G, Williamson Shaffer D. A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. Journal of Biomechanical Engineering. 2015;137(2):024701.

[68] Cant RP, Cooper SJ. Simulation-based learning in nurse education: systematic review. Journal of advanced nursing. 2010;66(1):3–15.

[69] Maynez J, Narayan S, Bohnet B, McDonald R. On Faithfulness and Factuality in Abstractive Summarization. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020. p. 1906–1919.

[70] Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. ACM Computing Surveys. 2023;55(12):1–38.

[71] Carlini N, Tramer F, Wallace E, Jagielski M, Herbert-Voss A, Lee K, et al. Extracting training data from large language models. In: 30th USENIX Security Symposium (USENIX Security 21); 2021. p. 2633–2650.

[72] Borji A. A categorical archive of chatgpt failures. arXiv preprint arXiv:230203494. 2023;.

[73] Chelli M, Descamps J, Lavoué V, Trojani C, Azar M, Deckert M, et al. Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis. Journal of Medical Internet Research. 2024;26:e53164.

[74] Sahoo NR, Saxena A, Maharaj K, Ahmad AA, Mishra A, Bhattacharyya P. Addressing Bias and Hallucination in Large Language Models. In: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024): Tutorial Summaries; 2024. p. 73–79.

[75] Ng DTK, Leung JKL, Chu SKW, Qiao MS. Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence. 2021;2:100041.

[76] Leiser F, Eckhardt S, Knaeble M, Maedche A, Schwabe G, Sunyaev A. From ChatGPT to FactGPT: A Participatory Design Study to Mitigate the Effects of Large Language Model Hallucinations on Users. In: Proceedings of Mensch Und Computer 2023. MuC '23. New York, NY, USA: Association for Computing Machinery; 2023. p. 81–90. Available from: https://doi.org/10.1145/3603555.3603565.

[77] Schneider J, Richner R, Riser M. Towards trustworthy autograding of short, multi-lingual, multi-type answers. International Journal of Artificial Intelligence in Education. 2023;33(1):88–118.

[78] Khosravi H, Shum SB, Chen G, Conati C, Tsai YS, Kay J, et al. Explainable artificial intelligence in education. Computers and Education: Artificial Intelligence. 2022;3:100074.

[79] Yang SJ, Ogata H, Matsui T, Chen NS. Human-centered artificial intelligence in education: Seeing the invisible through the visible. Computers and Education: Artificial Intelligence. 2021;2:100008.

[80] Short H. A critical evaluation of the contribution of trust to effective T echnology E nhanced L earning in the workplace: A literature review. British Journal of Educational Technology. 2014;45(6):1014–1022.

[81] Mutimukwe C, Viberg O, Oberg LM, Cerratto-Pargman T. Students' privacy concerns in learning analytics: Model development. British Journal of Educational Technology. 2022;53(4):932–951.

[82] Brown H, Lee K, Mireshghallah F, Shokri R, Tramèr F. What Does it Mean for a Language Model to Preserve Privacy? In: 2022 ACM Conference on Fairness, Accountability, and Transparency; 2022. p. 2280–2292.

[83] Nasr M, Carlini N, Hayase J, Jagielski M, Cooper AF, Ippolito D, et al. Scalable extraction of training data from (production) language models. arXiv preprint arXiv:231117035. 2023;.

[84] Winograd A. Loose-lipped large language models spill your secrets: The privacy implications of large language models. Harvard Journal of Law & Technology. 2023;36(2).

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

[86] Pugh SL, Subburaj SK, Rao AR, Stewart AE, Andrews-Todd J, D'Mello SK. Say What? Automatic Modeling of Collaborative Problem Solving Skills from Student Speech in the Wild. International Educational Data Mining Society. 2021;.

[87] Sha L, Rakovic M, Whitelock-Wainwright A, Carroll D, Yew VM, Gasevic D, et al. Assessing algorithmic fairness in automatic classifiers of educational forum posts. In: Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14–18, 2021, Proceedings, Part I 22. Springer; 2021. p. 381–394.

[88] Merine R, Purkayastha S. Risks and Benefits of AI-generated Text Summarization for Expert Level Content in Graduate Health Informatics. In: 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI). IEEE; 2022. p. 567–574.

[89] Sha L, Raković M, Das A, Gašević D, Chen G. Leveraging class balancing techniques to alleviate algorithmic bias for predictive tasks in education. IEEE Transactions on Learning Technologies. 2022;15(4):481–492.

[90] Sha L, Li Y, Gasevic D, Chen G. Bigger Data or Fairer Data? Augmenting BERT via Active Sampling for Educational Text Classification. In: Proceedings of the 29th International Conference on Computational Linguistics; 2022. p. 1275–1285.

[91] Wu J. Analysis and Evaluation of the Impact of Integrating Mental Health Education into the Teaching of University Civics Courses in the Context of Artificial Intelligence. Wireless Communications and Mobile Computing. 2022;2022.

[92] Tlili A, Shehata B, Adarkwah MA, Bozkurt A, Hickey DT, Huang R, et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments. 2023;10(1):15.

[93] European Parliment.: EU AI Act: first regulation on artificial intelligence. Accessed: 2023-12-24. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

[94] Mao J, Chen B, Liu JC. Generative Artificial Intelligence in Education and Its Implications for Assessment. TechTrends. 2023;p. 1–9.

[95] Yang Z, Liu J, Han Y, Chen X, Huang Z, Fu B, et al. AppAgent: Multimodal Agents as Smartphone Users. arXiv preprint arXiv:231213771. 2023;.

[96] Viberg O, Hatakka M, Bälter O, Mavroudi A. The current landscape of learning analytics in higher education. Computers in human behavior. 2018;89:98–110.

[97] Siemens G, Marmolejo-Ramos F, Gabriel F, Medeiros K, Marrone R, Joksimovic S, et al. Human and artificial cognition. Computers and Education: Artificial Intelligence. 2022;3:100107.

[98] Järvelä S, Zhao G, Heikkilä J, Järvenoja H, Mikkonen K, Kaleva S. Hybrid Intelligence–Human-AI Co-Evolution and Learning in Multirealities (HI). In: HHAI 2023: Augmenting Human Intellect. IOS Press; 2023. p. 392–394.

[99] Long D, Magerko B. What is AI literacy? Competencies and design considerations. In: Proceedings of the 2020 CHI conference on human factors in computing systems; 2020. p. 1–16.

[100] Weiser B.: Here's What Happens When Your Lawyer Uses ChatGPT. Accessed on 2023-12-01. Available from: https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html.

[101] Kabir S, Udo-Imeh DN, Kou B, Zhang T. Is stack overflow obsolete? an empirical study of the characteristics of chatgpt answers to stack overflow questions. In: Proceedings of the CHI Conference on Human Factors in Computing Systems; 2024. p. 1–17.

[102] Bjork RA, Dunlosky J, Kornell N. Self-regulated learning: Beliefs, techniques, and illusions. Annual review of psychology. 2013;64:417–444.

[103] Kabir S, Udo-Imeh DN, Kou B, Zhang T. Who answers it better? an in-depth analysis of chatgpt and stack overflow answers to software engineering questions. arXiv preprint arXiv:230802312. 2023;.

[104] Rafner J, Beaty RE, Kaufman JC, Lubart T, Sherson J. Creativity in the age of generative AI. Nature Human Behaviour. 2023;p. 1–3.

[105] Shneiderman B. Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human–Computer Interaction. 2020;36(6):495–504.

[106] UNESCO.: Generative Artificial Intelligence in Education: What are the Opportunities and Challenges? United Nations Educational, Scientific and Cultural Organization. Available from: https://www.unesco.org/en/articles/generative-artificial-intelligence-education-what-are-opportunities-and-challenges.

[107] Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLoS digital health. 2023;2(2):e0000198.

[108] Choi JH, Hickman KE, Monahan AB, Schwarcz D. ChatGPT goes to law school. J Legal Educ. 2021;71:387.

[109] Terwiesch C. Would Chat GPT3 get a Wharton MBA? A prediction based on its performance in the operations management course. Mack Institute for Innovation Management at the Wharton School, University of Pennsylvania. 2023;45.

[110] Zhang SJ, Florin S, Lee AN, Niknafs E, Marginean A, Wang A, et al. Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models. arXiv preprint arXiv:230608997. 2023;.

[111] Chowdhuri R, Deshmukh N, Koplow D.: No, GPT4 can't ace MIT. Accessed: 2023-12-23. https://bit.ly/No-GPT4-can-t-ace-MIT.

[112] Lorenz P, Perset K, Berryhill J.: Initial policy considerations for generative artificial intelligence. Accessed: 2024-07-09. https://www.oecd.org/en/publications/initial-policy-considerations-for-generative-artificial-intelligence_fae2d1e6-en.html.

Acknowledgments

Competing interests

The authors declare no competing interests.