nature human behaviour

Perspective

https://doi.org/10.1038/s41562-024-02004-5

Promises and challenges of generative artificial intelligence for human learning

Received: 25 February 2024

Accepted: 3 September 2024

Published online: 22 October 2024

Check for updates

Lixiang Yan [1], Samuel Greiff [2,3,4] ✉, Ziwen Teuber [2] & Dragan Gašević [1] ✉

Generative artificial intelligence (GenAI) holds the potential to transform the delivery, cultivation and evaluation of human learning. Here the authors examine 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 personalized 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. Thus, 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 effect on human cognition, metacognition and creativity. Humanity must learn with and about GenAI, ensuring that it becomes a powerful ally in the pursuit of knowledge and innovation, rather than a crutch that undermines our intellectual abilities.

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. Although education encompasses formalized systems that structure learning processes, learning represents the dynamic and personal process that occurs within this framework (see Box 1 for key definitions of human learning concepts).

The history of human learning presents a narrative of continuous evolution and adaptation to technological breakthroughs. For example, the printing press democratized access to knowledge and opened the opportunity of learning to many, whereas 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 (see Box 2 for key definitions of AI terms), 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, over-reliance on GenAI may negatively affect learners' agency, critical thinking and creativity, warranting caution[10].

Consequently, it is essential to balance technological advancement and human-centred values in learning. The aim of this Perspective

[1]Faculty of Information Technology, Monash University, Melbourne, Victoria, Australia. [2]Department of Behavioral and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg. [3]Department of Educational Psychology, Goethe-University Frankfurt, Frankfurt, Germany. [4]Present address: Centre for International Student Assessment (ZIB) & School of Social Sciences and Technology, Technical University of Munich, Munich, Germany. ✉e-mail: samuel.greiff@tum.de; dragan.gasevic@monash.edu

BOX 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 programmes. 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 that learners construct their own understanding and knowledge of the world through experiences and reflecting on those experiences. It emphasizes active engagement, exploration and the application of knowledge in meaningful contexts.

Inquiry-based learning. Inquiry-based learning refers to pedagogical approaches through which learners use scientific methods to build knowledge through formulating hypotheses, conducting experiments and making observations to discover causal relationships. It emphasizes 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 categorize 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 optimize 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, informing 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 identified feedback as one of the nine hallmarks that characterize 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, personalized 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.

is 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 emphasize the importance of designing AI-driven educational tools that prioritize 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 (Table 1).

Promises

GenAI promises to transform human learning by scaling personalized support, diversifying learning resources, enabling timely feedback and innovating assessment methods. The realization of these promises depends on the roles and interactions GenAI has with learners and educators (Fig. 1). 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 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.

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 personalized 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 personalized feedback, adaptive questioning and conversational engagement, with minimal previous 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 democratize access to high-quality learning support, making it accessible to learners globally.

GenAI's role aligns with Vygotsky's sociocultural theory of learning, whereby more knowledgeable others guide learners within their zone of proximal development[12]. By integrating new technologies such as ChatGPT into intelligent tutoring systems, GenAI can offer personalized and adaptive support based on each learner's unique knowledge[13]. These language models have demonstrated remarkable proficiency

BOX 2

Glossary of AI terms

GenAI. GenAI 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 new and relevant, often mimicking human creativity.

LLM. An LLM is a computational model known for its ability to perform general purpose language generation and various natural language processing tasks such as 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 GenAI.

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 such as 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 organizes data into nodes (representing entities) and edges (representing relationships between entities), enabling complex queries and inferences. Knowledge graphs are used in applications such as 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 AI, 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.

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 the 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 emphasizes 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 such as Khanmigo and students' positive attitudes towards using such technologies for personalized learning support[17], it is important to acknowledge the current limitations in empirical evidence regarding their short- and long-term effects on learning outcomes[18]. Emerging evidence indicates that the effects on learning engagement, agency and performance can paint a complicated and mixed picture (for example, 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 randomized controlled studies that compare the effectiveness of GenAI tutoring with conventional rule-based tutoring systems over several academic terms and across different subjects to contextualize its effects within various educational settings.

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 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 (for example, remember and understand) but struggles with higher levels (for example, apply and create)[25]. These findings suggest that

https://doi.org/10.1038/s41562-024-02004-5

Table 1 | Overview of the effects of GenAI on human learning

Theme	Learning effect	Key components	Learners	Educators	Researchers	Policymakers	Technologists
Promises	Learning support	Scale personalized and adaptive support	x	x	x		x
		Learn within zone of proximal development	x	x	x		
	Learning resource	Efficient learning content creation		x	x		x
		Diverse multimedia learning resources		x	x		x
	Learning feedback	Timely, specific and constructive feedback		x	x		x
		Multimodal feedback delivery		x	x		x
	Learning assessment	Adaptive and authentic assessments		x	x		x
		Enhanced real-world simulation		x	x		x
Challenges	GenAI's imperfections	Hallucinations and inaccuracies			x		x
		Instability and biases			x		x
	Ethical dilemma	Transparency and privacy concerns			x	x	x
		Equality and beneficence				x	x
	Disruption of assessment	Human–AI hybrid cognition assessment		x	x		
		AI-induced performance illusion	x	x	x		
Needs	AI literacy	Basic understanding of AI systems	x	x	x	x	
		Critically evaluate AI-generated content	x	x	x	x	
	Evidence-based decision-making	Robust evidence on learning effects			x		
		Multi-party collaborative efforts	x	x	x	x	x
	Methodological rigour	Evidence quality appraisal instruments			x		
		Long-term evaluation of learning effects			x		

although 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 such as 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 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 such as Runway's Gen-3 Alpha 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 to create more dynamic and engaging learning environments. This enables learners to interact with content in more informed, creative and personalized ways. Such an approach aligns with constructivist learning principles, which emphasize 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

https://doi.org/10.1038/s41562-024-02004-5

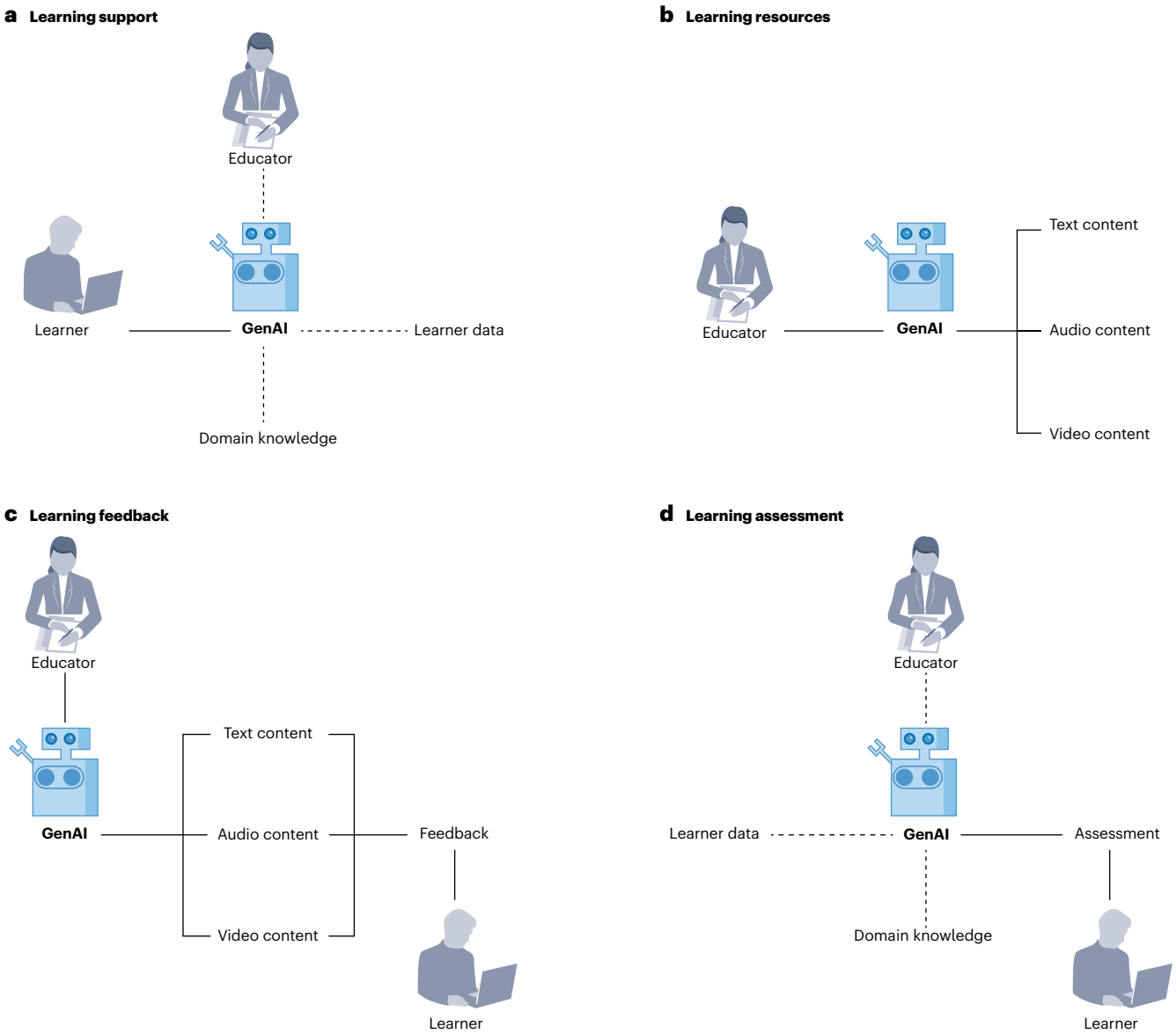

Fig. 1 | Examples of human–AI interactions in human learning. a, Learners receive personalized and adaptive support from GenAI tutors, which are co-designed with educators and have access to previous learner data and domain knowledge. **b**, Educators use GenAI to create multimodal learning resources, incorporating text, audio and video content. **c**, Educators collaborate with GenAI to deliver multimodal feedback to learners. **d**, GenAI agents use input requirements from educators, previous learner data and domain knowledge to create assessment activities that evaluate learners.

the optimal level of automation versus human control, the extent of expert validation required and the degree of alignment with learning objectives.

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 that is essential for effective learning[34–36]. Providing detailed feedback regularly is laborious and time consuming, adding to educators' workloads, especially as students perceive timely feedback as the most effective[37]. GenAI can assist by analysing student work and delivering instant, personalized feedback with minimal previous training. For example, a recent study found that ChatGPT generates more in-depth and fluent feedback, coherently summarizing students' performances compared with 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 (for example, GPT-4 with Vision and Gemini 1.0 Pro) can help students to navigate and comprehend insights from learning analytics dashboards[43], which combine data, analytics and visualizations 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 three-dimensional 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]. Previous 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 costs and benefits, as these models—especially video generation models—require high computational power, potentially widening the inequality in learning opportunities.

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 personalized 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 such as AutoGen[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 organized 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 (for example, 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 programmes[60]. These applications demonstrate strong potential for generalizability, 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 contextualized. 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 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 though they were in real professional settings—a concept that has 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 new assessment settings to accurately capture genuine human learning.

Challenges

Amid 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.

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 that are 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]. These inaccuracies can undermine GenAI's reliability as a learning tool, potentially outweighing its promises.

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 (for example, 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 the identification of 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 recognizing signs of hallucination. These steps are essential for cultivating AI literacy[75], as discussed further in the section 'AI literacy'. Additionally, designing and optimizing 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).

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 affected by these technologies[77]. The primary cause of this transparency gap is the absence of human-in-the-loop elements in previous 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 personalization 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 anonymization raises serious issues about exposing sensitive information[82]. For instance, researchers conducted a divergence attack on ChatGPT, compromizing its security and causing it to output original training data containing personally identifiable information[83]. 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 have 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. Although 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 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 affect human learning and perpetuate systemic biases, such as gender, racial and social class biases[87,88]. Strategies such as 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]. Although 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].

Disruption of assessment

GenAI poses substantial 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 such as 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 randomized controlled experiment found that although GenAI tools can help students to 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 that even the learning process can be imitated or augmented by AI. Preliminary work on multimodal GenAI agents[95] has shown that these agents can operate smartphone applications, generating digital trace data while executing user requests. These 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. Assessment of human cognition and metacognition remains essential for kindergarten to 12th grade (K–12; 5 to 18 years old) education, as young learners continue developing fundamental skills. In higher education, prioritizing the evaluation of human–AI hybrid cognition and metacognition could be crucial in 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.

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, prioritizing 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.

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 awareness of their potential effects, 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 indicated 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 twenty-first 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.

Evidence-based decision-making

The integration of GenAI into human learning promises to enhance experiences and outcomes (as highlighted in the 'Promises' section). 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 that 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, although 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]. 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 maximize 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 prioritizing 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.

Methodological rigour

Building on discussions of evidence-based decision-making, it is crucial to emphasize 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 effect of these tools on teaching and learning processes. GenAI's capabilities, such as passing the United States Medical Licensing Exam[107], completing examinations 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 study claiming that GPT-4, with prompt engineering, could achieve perfect scores in the MIT Mathematics and Electrical Engineering and Computer Science curricula[110]. This study[110], initially attracting widespread attention, was later retracted due to methodological concerns, including dataset contamination, over-reliance on GPT-4 for accuracy assessment and ambiguities in manual verification of the results[111]. This incident underscores the need for rigorous methodological standards, probably 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 effect 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 the complexity level, target student grade and desired question format (for example, 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 generalizable findings, fostering the responsible and effective integration of GenAI technologies into learning.

Conclusion and future directions

As we look towards 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 to 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 visualizing Einstein's special theory of relativity in virtual realities. All of this could occur while receiving personalized 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 maximizing its potential as a learning tool. This understanding will enhance the effectiveness of AI-driven educational tools and ensure that human ingenuity is preserved amid technological advancement. Key research questions include: how we can promote human–AI interaction to maximize learner agency; what behavioural indicators can reliably capture cognitive and metacognitive processes during AI-assisted learning; how we can assess learning to reflect genuine knowledge and skill development rather than an AI-created performance illusion; and what strategies can prevent over-reliance on AI, ensuring that 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 models 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 that 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 we can 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; and how we can 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 that GenAI 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 harmonizes 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. *Comput. Educ. Artif. Intell.* **4**, 100130 (2023).
2. Yan, L. et al. Practical and ethical challenges of large language models in education: a systematic scoping review. *Br. J. Educ. Technol.* **35**, 90–112 (2023).
3. Dai, W. et al. Can large language models provide feedback to students? A case study on ChatGPT. In *Proc. 2023 IEEE International Conference on Advanced Learning Technologies* 323–325 (IEEE, 2023).
4. Li, Y. et al. Can large language models write reflectively. *Comput. Educ. Artif. Intell.* **4**, 100140 (2023).
5. Yildirim-Erbasli, S. N. & Bulut, O. Conversation-based assessment: a novel approach to boosting test-taking effort in digital formative assessment. *Comput. Educ. Artif. Intell.* **4**, 100135 (2023).
6. Mazzoli, C. A., Semeraro, F. & Gamberini, L. Enhancing cardiac arrest education: exploring the potential use of Midjourney. *Resuscitation* **189**, 109893 (2023).
7. Vartiainen, H. & Tedre, M. Using artificial intelligence in craft education: crafting with text-to-image generative models. *Digit. Creat.* **34**, 1–21 (2023).
8. Kasneci, E. et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Diff.* **103**, 102274 (2023).
9. Falcão, T. P., Mello, R. F. & Rodrigues, R. L. Applications of learning analytics in Latin America. *J. Learn. Anal.* **51**, 871–874 (2020).
10. Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D. & Siemens, G. Impact of AI assistance on student agency. *Comput. Educ.* **210**, 104967 (2024).
11. Mousavinasab, E. et al. Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. *Interact. Learn. Environ.* **29**, 142–163 (2021).
12. Vygotsky, L. S. & Cole, M. *Mind in Society: Development of Higher Psychological Processes* (Harvard Univ. 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. *Comput. Educ. Artif. Intell.* **4**, 100138 (2023).
14. Chang, Y. et al. A survey on evaluation of large language models. *ACM Trans. Intell. Syst. Technol.* **15**, 1–45 (2024).
15. Meet Khanmigo: Khan Academy's AI-powered teaching assistant & tutor. *Khan Academy* https://www.khanmigo.ai/ (2023).
16. Lee, V. S. What is inquiry-guided learning? *New Dir. Teach. Learn.* **129**, 5–14 (2012).
17. 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).
18. Hennessy, S., Cukurova, M., Lewin, C., Mavrikis, M. & Major, L. BJET Editorial 2024: a call for research rigour. *Br. J. Educ. Technol.* **55**, 5–9 (2024).
19. Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D. & Siemens, G. Impact of AI assistance on student agency. *Comput. Educ.* **210**, 104967 (2024).
20. Nie, A. et al. The GPT surprise: offering large language model chat in a massive coding class reduced engagement but increased adopters exam performances. Preprint at arXiv https://doi.org/10.48550/arXiv.2407.09975 (2024).
21. Molenaar, I. Towards hybrid human–AI learning technologies. *Eur. J. Educ.* **57**, 632–645 (2022).
22. Ji, H., Han, I. & Ko, Y. A systematic review of conversational AI in language education: focusing on the collaboration with human teachers. *J. Res. Technol. Educ.* **55**, 48–63 (2023).
23. Yang, K. B. et al. Surveying teachers' preferences and boundaries regarding human–AI control in dynamic pairing of students for collaborative learning. In *Proc. 16th European Conference on Technology Enhanced Learning* 260–274 (Springer, 2021).
24. Pesovski, I., Santos, R., Henriques, R. & Trajkovik, V. Generative AI for customizable learning experiences. *Sustainability* **16**, 3034 (2024).
25. Hwang, K., Wang, K., Alomair, M., Choa, F.-S. & Chen, L. K. Towards automated multiple choice question generation and evaluation: aligning with Bloom's taxonomy. In *Proc. 25th International Conference on Artificial Intelligence in Education* 389–396 (Springer, 2024).
26. Radford, A. et al. Learning transferable visual models from natural language supervision. In *Proc. 38th International Conference on Machine Learning* 8748–8763 (PMLR, 2021).
27. Chiu, T. K. The impact of generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney. *Interact. Learn. Environ.* https://doi.org/10.1080/10494820.2023.2253861 (2023).
28. Lee, U. et al. Prompt Aloud!: incorporating image-generative AI into STEAM class with learning analytics using prompt data. *Educ. Inform. Technol.* **29**, 9575–9605 (2024).
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. *Educ. Technol. Soc.* **27**, 284–305 (2024).
30. Long, L., MacBlain, S. & MacBlain, M. Supporting students with dyslexia at the secondary level: an emotional model of literacy. *J. Adolesc. Adult Lit.* **51**, 124–134 (2007).
31. Leiker, D., Gyllen, A. R., Eldesouky, I. & Cukurova, M. Generative AI for learning: investigating the potential of learning videos with synthetic virtual instructors. In *Proc. 24th International Conference on Artificial Intelligence in Education* 523–529 (Springer, 2023).
32. Bada, S. O. & Olusegun, S. Constructivism learning theory: a paradigm for teaching and learning. *J. Res. Method Educ.* **5**, 66–70 (2015).
33. Tavakoli, M., Faraji, A., Molavi, M., Mol, S. T. & Kismihók, G. Hybrid human–AI curriculum development for personalised informal learning environments. In *Proc. 12th International Learning Analytics and Knowledge Conference* 563–569 (ACM, 2022).
34. Pardo, A., Jovanovic, J., Dawson, S., Gašević, D. & Mirriahi, N. Using learning analytics to scale the provision of personalised feedback. *Br. J. Educ. Technol.* **50**, 128–138 (2019).
35. Lim, L.-A. et al. What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course. *Learn. Instr.* **72**, 101202 (2021).
36. Hattie, J. & Timperley, H. The power of feedback. *Rev. Educ. Res.* **77**, 81–112 (2007).
37. Poulos, A. & Mahony, M. J. Effectiveness of feedback: the students' perspective. *Assess. Eval. High. Educ.* **33**, 143–154 (2008).
38. Steiss, J. et al. Comparing the quality of human and ChatGPT feedback of students' writing. *Learn. Instr.* **91**, 101894 (2024).
39. Meyer, J. et al. Using llms to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Comput. Educ. Artif. Intell.* **6**, 100199 (2024).
40. Zhang, Z. et al. Students' perceptions and preferences of generative artificial intelligence feedback for programming. In *Proc. 38th AAAI Conference on Artificial Intelligence* 23250–23258 (AAAI, 2024).
41. Liang, Z., Sha, L., Tsai, Y.-S., Gašević, D. & Chen, G. Towards the automated generation of readily applicable personalised feedback in education. In *Proc. 25th International Conference on Artificial Intelligence in Education* 75–88 (Springer, 2024).

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. *Comput. Educ. Artif. Intell.* **6**, 100228 (2024).
43. Yan, L. et al. VizChat: enhancing learning analytics dashboards with contextualised explanations using multimodal generative AI chatbots. In *Proc. 25th International Conference on Artificial Intelligence in Education* 180–193 (Springer, 2024).
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 Trans. Learn. Technol.* **13**, 226–245 (2019).
45. Yang, M. & Carless, D. The feedback triangle and the enhancement of dialogic feedback processes. *Teach. High. Educ.* **18**, 285–297 (2013).
46. Dawson, P. et al. in *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy* 695–739 (Springer, 2023).
47. Wang, T. et al. RODIN: a generative model for sculpting 3D digital avatars using diffusion. In *Proc. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 4563–4573 (IEEE, 2023).
48. Le, M. et al. Voicebox: text-guided multilingual universal speech generation at scale. In *Advances in Neural Information Processing Systems* (eds Oh, A. et al.) 14005–14034 (Curran Associates, 2023).
49. McCarthy, J. Evaluating written, audio and video feedback in higher education summative assessment tasks. *Issues Educ. Res.* **25**, 153–169 (2015).
50. Orlando, J. A comparison of text, voice, and screencasting feedback to online students. *Am. J. Distance Educ.* **30**, 156–166 (2016).
51. Henderson, M. & Phillips, M. Video-based feedback on student assessment: scarily personal. *Austral. J. Educ. Technol.* **31**, 51–66 (2015).
52. Swiecki, Z. et al. Assessment in the age of artificial intelligence. *Comput. Educ. Artif. Intell.* **3**, 100075 (2022).
53. Wu, Q. et al. AutoGen: enabling next-gen LLM applications via multi-agent conversation. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2308.08155 (2023).
54. Park, J. S. et al. Generative agents: interactive simulacra of human behavior. In *Proc. 36th Annual ACM Symposium on User Interface Software and Technology* 1–22 (ACM, 2023).
55. Fan, Y. et al. Towards investigating the validity of measurement of self-regulated learning based on trace data. *Metacogn. Learn.* **17**, 949–987 (2022).
56. Allen, L. K., Creer, S. C. & Öncel, P. in *The Handbook of Learning Analytics* 2nd edn (eds Lang, C et al.) 46–53 (Society for Learning Analytics Research, 2022).
57. Gašević, D., Greiff, S. & Shaffer, D. W. Towards strengthening links between learning analytics and assessment: challenges and potentials of a promising new bond. *Comput. Hum. Behav.* **134**, 107304 (2022).
58. Lagakis, P. & Demetriadis, S. EvaAI: a multi-agent framework leveraging large language models for enhanced automated grading. In *Proc. 20th International Conference on Intelligent Tutoring Systems* 378–385 (Springer, 2024).
59. Shahzad, R. et al. Multi-agent system for students cognitive assessment in e-learning environment. *IEEE Access* **12**, 15458–15467 (2024).
60. Yang, K. et al. Content knowledge identification with multi-agent large language models (LLMs). In *Proc. 25th International Conference on Artificial Intelligence in Education* 284–292 (Springer, 2024).
61. Song, W. et al. An intelligent virtual standard patient for medical students training based on oral knowledge graph. *IEEE Trans. Multimedia* **25**, 6132–6145 (2022).
62. Ji, S., Pan, S., Cambria, E., Marttinen, P. & Philip, S. Y. A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **33**, 494–514 (2021).
63. Rehm, J., Reshodko, I., Børresen, S. Z. & Gundersen, O. E. The virtual driving instructor: multi-agent system collaborating via knowledge graph for scalable driver education. In *Proc. 38th AAAI Conference on Artificial Intelligence* 22806–22814 (2024).
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 *Proc. 2024 CHI Conference on Human Factors in Computing Systems* 1–28 (ACM, 2024).
65. Yang, Q.-F., Lian, L.-W. & Zhao, J.-H. Developing a gamified artificial intelligence educational robot to promote learning effectiveness and behavior in laboratory safety courses for undergraduate students. *Int. J. Educ. Technol. High. Educ.* **20**, 18 (2023).
66. Thanh, B. N. et al. Race with the machines: assessing the capability of generative AI in solving authentic assessments. *Australas. J. Educ. Technol.* **39**, 59–81 (2023).
67. Chesler, N. C. et al. A novel paradigm for engineering education: virtual internships with individualized mentoring and assessment of engineering thinking. *J. Biomech. Eng.* **137**, 024701 (2015).
68. Cant, R. P. & Cooper, S. J. Simulation-based learning in nurse education: systematic review. *J. Adv. Nurs.* **66**, 3–15 (2010).
69. Maynez, J., Narayan, S., Bohnet, B. & McDonald, R. On faithfulness and factuality in abstractive summarization. In *Proc. 58th Annual Meeting of the Association for Computational Linguistics* 1906–1919 (Association for Computational Linguistics, 2020).
70. Ji, Z. et al. Survey of hallucination in natural language generation. *ACM Comput. Surv.* **55**, 1–38 (2023).
71. Carlini, N. et al. Extracting training data from large language models. In *Proc. 30th USENIX Security Symposium* 2633–2650 (USENIX, 2021).
72. Borji, A. A categorical archive of ChatGPT failures. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2302.03494 (2023).
73. Chelli, M. et al. Hallucination rates and reference accuracy of ChatGPT and bard for systematic reviews: comparative analysis. *J. Med. Internet Res.* **26**, e53164 (2024).
74. Sahoo, N. R. et al. Addressing bias and hallucination in large language models. In *Proc. 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* 73–79 (ELRA Language Resource Association, 2024).
75. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W. & Qiao, M. S. Conceptualizing AI literacy: an exploratory review. *Comput. Educ. Artif. Intell.* **2**, 100041 (2021).
76. Leiser, F. et al. From ChatGPT to FactGPT: a participatory design study to mitigate the effects of large language model hallucinations on users. In *Proc. Mensch Und Computer 2023* 81–90 (Association for Computing Machinery, 2023).
77. Schneider, J., Richner, R. & Riser, M. Towards trustworthy autograding of short, multi-lingual, multi-type answers. *Int. J. Artif. Intell. Educ.* **33**, 88–118 (2023).
78. Khosravi, H. et al. Explainable artificial intelligence in education. *Comput. Educ. Artif. Intell.* **3**, 100074 (2022).
79. Yang, S. J., Ogata, H., Matsui, T. & Chen, N.-S. Human-centered artificial intelligence in education: seeing the invisible through the visible. *Comput. Educ. Artif. Intell.* **2**, 100008 (2021).
80. Short, H. A critical evaluation of the contribution of trust to effective technology enhanced learning in the workplace: a literature review. *Br. J. Educ. Technol.* **45**, 1014–1022 (2014).

81. Mutimukwe, C., Viberg, O., Oberg, L.-M. & Cerratto-Pargman, T. Students' privacy concerns in learning analytics: model development. *Br. J. Educ. Technol.* **53**, 932–951 (2022).
82. Brown, H., Lee, K., Mireshghallah, F., Shokri, R. & Tramèr, F. What does it mean for a language model to preserve privacy? In *Proc. 2022 ACM Conference on Fairness, Accountability, and Transparency* 2280–2292 (ACM, 2022).
83. Nasr, M. et al. Scalable extraction of training data from (production) language models. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2311.17035 (2023).
84. Winograd, A. Loose-lipped large language models spill your secrets: the privacy implications of large language models. *Harvard J. Law Technol.* **36**, 616–656 (2023).
85. Yao, Y. et al. A survey on large language model (LLM) security and privacy: the good, the bad, and the ugly. *High Confid. Comput.* **4**, 100211 (2024).
86. Pugh, S. L. et al. Say what? Automatic modeling of collaborative problem solving skills from student speech in the wild. *Proc. 14th International Conference on Educational Data Mining* 55–67 (International Educational Data Mining Society, 2021).
87. Sha, L. et al. Assessing algorithmic fairness in automatic classifiers of educational forum posts. In *Proc. 22nd International Conference on Artificial Intelligence in Education* 381–394 (Springer, 2021).
88. Merine, R. & Purkayastha, S. Risks and benefits of AI-generated text summarization for expert level content in graduate health informatics. In *Proc. 10th International Conference on Healthcare Informatics* 567–574 (IEEE, 2022).
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 Trans. Learn. Technol.* **15**, 481–492 (2022).
90. Sha, L., Li, Y., Gasevic, D. & Chen, G. Bigger data or fairer data? Augmenting BERT via active sampling for educational text classification. In *Proc. 29th International Conference on Computational Linguistics* 1275–1285 (International Committee on Computational Linguistics, 2022).
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. *Wirel. Commun. Mob. Comput.* https://doi.org/10.1155/2022/5378694 (2022).
92. Tlili, A. et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learn. Environ.* **10**, 15 (2023).
93. EU AI act: first regulation on artificial intelligence. *European Parliament* https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (2023).
94. Mao, J., Chen, B. & Liu, J. C. Generative artificial intelligence in education and its implications for assessment. *TechTrends* **68**, 58–66 (2023).
95. Yang, Z. et al. AppAgent: multimodal agents as smartphone users. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2312.13771 (2023).
96. Viberg, O., Hatakka, M., Bälter, O. & Mavroudi, A. The current landscape of learning analytics in higher education. *Comput. Hum. Behav.* **89**, 98–110 (2018).
97. Siemens, G. et al. Human and artificial cognition. *Comput. Educ. Artif. Intell.* **3**, 100107 (2022).
98. Järvelä, S. et al. Hybrid intelligence—human–AI co-evolution and learning in multirealities (HI). In *Proc. 2nd International Conference on Hybrid Human–Artificial Intelligence* 392–394 (IOS Press, 2023).
99. Long, D. & Magerko, B. What is AI literacy? Competencies and design considerations. In *Proc. 2020 CHI Conference on Human Factors in Computing Systems* 1–16 (ACM, 2020).
100. Weiser, B. Here's what happens when your lawyer uses ChatGPT. *The New York Times* (28 May 2023).
101. Kabir, S., Udo-Imeh, D. N., Kou, B. & Zhang, T. Is stack overflow obsolete? an empirical study of the characteristics of chatgpt answers to stack overflow questions. In *Proc. 2024 CHI Conference on Human Factors in Computing Systems* 1–17 (ACM, 2024).
102. Bjork, R. A., Dunlosky, J. & Kornell, N. Self-regulated learning: beliefs, techniques, and illusions. *Annu. Rev. Psychol.* **64**, 417–444 (2013).
103. Kabir, S., Udo-Imeh, D. N., Kou, B. & Zhang, T. Is stack overflow obsolete? an empirical study of the characteristics of chatgpt answers to stack overflow questions. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2308.02312 (2023).
104. Rafner, J., Beaty, R. E., Kaufman, J. C., Lubart, T. & Sherson, J. Creativity in the age of generative AI. *Nat. Hum. Behav.* **7**, 1836–1838 (2023).
105. Shneiderman, B. Human-centered artificial intelligence: reliable, safe & trustworthy. *Int. J. Hum. Comput. Interact.* **36**, 495–504 (2020).
106. Giannini, S. Generative artificial intelligence in education: think piece by Stefania Giannini. *unesco.org* https://www.unesco.org/en/articles/generative-artificial-intelligence-education-what-are-opportunities-and-challenges (UNESCO, 2023).
107. Kung, T. H. et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digit. Health* **2**, e0000198 (2023).
108. Choi, J. H., Hickman, K. E., Monahan, A. B. & Schwarcz, D. ChatGPT goes to law school. *J. Leg. Educ.* **71**, 387 (2021).
109. Terwiesch, C. *Would Chat GPT3 Get a Wharton MBA? A Prediction Based on its Performance in the Operations Management Course* (Wharton University of Pennsylvania, 2023).
110. Zhang, S. J. et al. Exploring the MIT Mathematics and EECS curriculum using large language models. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2306.08997 (2023).
111. Chowdhuri, R., Deshmukh, N. & Koplow, D. No, GPT4 can't ace MIT. *Raunak Does Dev* https://bit.ly/No-GPT4-can-t-ace-MIT (2023).
112. Lorenz, P., Perset, K. & Berryhill, J. *Initial Policy Considerations for Generative Artificial Intelligence* (OECD, 2023).

Acknowledgements

This study was supported by grants from the Australian Research Council (grant agreement numbers DP220101209 and DP240100069 to D.G.). L.Y.'s work is fully funded by the Digital Health Cooperative Research Centre (DHCRC). D.G.'s work was supported in part by the DHCRC and Defense Advanced Research Projects Agency (DARPA) through the Knowledge Management at Speed and Scale (KMASS) programme (HR0011-22-2-0047). The DHCRC is established and supported under the Australian Government's Cooperative Research Centres Program. The US Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the US Government. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Samuel Greiff or Dragan Gašević.

Peer review information *Nature Human Behaviour* thanks René Kizilcec and Stephen Aguilar for their contribution to the peer review of this work.