frontiers | Frontiers in Education

TYPE Perspective
PUBLISHED 08 January 2024
DOI 10.3389/feduc.2023.1331607

OPEN ACCESS

EDITED BY
Tom Crick,
Swansea University, United Kingdom

REVIEWED BY
Simon Robert Stones,
Envision Pharma Group, United Kingdom
Tom Prickett,
Northumbria University, United Kingdom

*CORRESPONDENCE
Ryan Thomas Williams
✉ ryan.williams@tees.ac.uk

RECEIVED 01 November 2023
ACCEPTED 18 December 2023
PUBLISHED 08 January 2024

CITATION
Williams RT (2024) The ethical implications of using generative chatbots in higher education.
Front. Educ. 8:1331607.
doi: 10.3389/feduc.2023.1331607

The ethical implications of using generative chatbots in higher education

Ryan Thomas Williams*

Teesside University International Business School, TU Online, Teesside University, Middlesbrough, United Kingdom

Incorporating artificial intelligence (AI) into education, specifically through generative chatbots, can transform teaching and learning for education professionals in both administrative and pedagogical ways. However, the ethical implications of using generative chatbots in education must be carefully considered. Ethical concerns about advanced chatbots have yet to be explored in the education sector. This short article introduces the ethical concerns associated with introducing platforms such as ChatGPT in education. The article outlines how handling sensitive student data by chatbots presents significant privacy challenges, thus requiring adherence to data protection regulations, which may not always be possible. It highlights the risk of algorithmic bias in chatbots, which could perpetuate societal biases, which can be problematic. The article also examines the balance between fostering student autonomy in learning and the potential impact on academic self-efficacy, noting the risk of over-reliance on AI for educational purposes. Plagiarism continues to emerge as a critical ethical concern, with AI-generated content threatening academic integrity. The article advocates for comprehensive measures to address these ethical issues, including clear policies, advanced plagiarism detection techniques, and innovative assessment methods. By addressing these ethical challenges, the article argues that educators, AI developers, policymakers, and students can fully harness the potential of chatbots in education, creating a more inclusive, empowering, and ethically sound educational future.

KEYWORDS

artificial intelligence, education, chatbots, data privacy, ethics in AI

Introduction

The general population has witnessed a growing intertwinement of artificial intelligence (AI) in their daily lives, raising questions about society, the economy, and education (Hasal et al., 2021). In fact, it is difficult to imagine an industry where AI will not add value in the future (Ng, 2016). AI is categorised as a 4.0 technology, which means an increasingly decentralised yet autonomous process of efficiencies (Alenizi et al., 2023), even though it has existed since the 1950s. Questions about whether machines can think are long-standing (Turing, 1950; McCarthy et al., 2006).

Microsoft (2023) describe AI as the ability of a computer system to mimic human cognitive functions such as learning and problem-solving. However, it is important to note that the notion of language models truly mimicking human cognitive abilities is complex. Zhao et al. (2022) argue that human cognitive abilities involve understanding, reasoning, and consciousness,

which are aspects that current AI models do not possess, for instance, thus, highlighting how multifaceted defining AI is.

Nevertheless, AI's advanced computational capabilities and machine learning (ML) algorithms have prompted scholars and educators to rethink traditional educational paradigms, as the technology promises a more personalised, efficient, and globally accessible education system. The benefits of AI in education have been well-publicised in the literature (Zhai, 2022), and this has led to a rise of a new area of 'best practice' guides for using AI in classrooms (Mollick and Mollick, 2022; Lieberman, 2023).

Among the burgeoning AI tools entering this domain, such as speech-to-text, text-to-speech, automatic image capturing, and personalised tutoring systems, there appears to be a particular focus on chatbots, such as Open AI's Chat Generative Pre-Trained Transformer (ChatGPT) (Hill-Yardin et al., 2023; Stokel-Walker and Van Noorden, 2023).

Chatbots can leverage natural language processing (NLP), an AI subfield that enables machines to understand, respond to, and generate human language. Previously, chatbots' primary function was simply to mimic human conversation, whereas platforms such as ChatGPT have abilities that far extend that. Tlili et al. (2023, p. 2) argue that due to the larger training set, up to 175 billion parameters, and fine-tuning, chatbots can now create new things from 'poems, stories, and novels to just about anything'. They have the potential to provide students with a quasi-human interaction experience, capable of offering explanations and answering complex queries to support their learning journey in ways that were previously inconceivable (Hill-Yardin et al., 2023).

The benefits of AI tools, especially in complementing traditional educational methods, are indeed appealing, with the potential for increased accessibility and individualised learning experiences. More specifically, universities have embraced chatbots to provide prospective students with fast and accurate answers to their questions, including queries about financial aid, available modules or assessment information for their programme (Meyer von Wolff et al., 2020). Chatbots can also save staff time by reducing the number of times they have to answer the same questions. University applicants value 24/7 availability as an important characteristic of chatbots (see Meyer von Wolff et al., 2020). Students have also become familiar with communicating with chatbots, using them on commercial apps such as retail and banking.

However, the value of chatbots extends beyond saving time on administrative burdens; rather, they can additionally transform pedagogy (Watermeyer et al., 2023). For instance, an educator may use chatbots to generate case studies for a seminar or provide best practices relating to academic skills.

As educators and policymakers consider implementing these innovative AI technologies into education, reflecting upon the ethical implications becomes a useful exercise. Integrating chatbots into education environments is not straightforward as first imagined; it resides within a complex nexus of ethical considerations that warrant scrutiny (Lund and Wang, 2023). Furthermore, concerns about advanced chatbots such as ChatGPT have yet to be explored in the education sector. Therefore, it remains unclear how AI, particularly chatbots, will overcome ethical concerns such as algorithmic bias and plagiarism.

Subsequently, this may lead to an overprotective reaction to a potential opportunity, such as New York City's schools' banning of ChatGPT from educational networks due to the risk of using it to cheat on assignments (Shen-Berro, 2023). Contrastingly, there may be a naïve acceptance of AI in education as 'the one' technology that will fundamentally change education provision and practice, overlooking and repeating high-profile failures of TEL in the past (Oppenheimer, 1997).

It is, therefore, important to investigate the concerns of using chatbots in education to ensure safe and ethical use. This article briefly introduces the ethical implications of using platforms such as ChatGPT in education. The author focuses on data privacy, algorithmic bias, autonomy in learning, and the issue of plagiarism.

Data privacy

Technology has been supporting universities in their efforts to connect with students and staff in transformative ways for a long time, such as through social media. Covid-19 and the unique period of remote learning was the catalyst that severely reduced on-campus interactions across the sector (Williams, 2022). University administrators had to develop and implement effective ways to communicate with others in the education community in ways that not only meet the needs of students but can expand and enhance relationships (Hill-Yardin et al., 2023). Williams (2023) describe this as technology being used in transformative ways.

A chatbot is AI software that can simulate a conversation with a user in natural language through messaging applications, websites, mobile apps or telephone. More specifically, when a human is interacting with a chatbot, it analyses the user request and extracts data, then returns a statistically plausible response, which can be in the form of predefined text, text selected from a knowledge base with different answers, contextualised information from data provided by a user, a clarifying question or data stored in enterprise systems (Castillo et al., 2023).

In commercial applications, chatbots can improve customer experience and provide smooth interactions, making it easier for customers to engage with an organisation and providing lower-cost customer service than live agents (Williams, 2023). However, the enhanced personalised experience is only possible because of the gathering of 'big data', such as tracking behaviour, habits, and patterns, and analysing them against historical customer activity.

Similarly, in an educational setting, the deployment of chatbots may include collecting, analysing, and storing student data. The data may be used to make helpful predictions about which students are at risk of falling behind in their work, and this is likely to be administered in a way that is beyond a simple Excel formula. This could allow academic tutors to develop early interventions and target support, which can be incorporated into programme planning (Hill-Yardin et al., 2023). As such, the data for these activities extend beyond academic performance metrics, often delving into sensitive personal information (Biswas, 2023). The functionality depends on the chatbots' ability to understand and respond to student learning habits, strengths and weaknesses. Subsequently, a vast digital footprint of each student is created and stored. There is a lack of education literature that explores how education providers should respond and handle the gathering of big data.

Furthermore, chatbots may be administered in a non-pedagogical sense to support student recruitment activities, such as UK HE admissions

processes. For example, in 2019, Leeds Beckett University launched a chatbot to allow prospective students to have an online conversation and assess their suitability for undergraduate courses available during university Clearing. The bot uses keywords that the prospective students type in the chat line to assess which courses they are interested in studying, then requests details of their qualifications and exam results before making them a provisional offer. This was administered via Messenger, so it was limited by its functionality. However, it helps illustrate how universities may view chatbots as a natural evolution in order to engage with prospective students. Similarly, in order to make a student an offer for a place at a university, a vast amount of personal data about the individual is required, thus, the importance of understanding the ethical implications associated with this.

The handling of such sensitive student information immediately raises significant privacy concerns. While the ability of generative chatbots to provide personalised experiences is certainly beneficial, the risk of misuse or unauthorised access to this data poses considerable threats to student or applicant privacy. The issue becomes more pronounced, particularly for young students, who may need to be made aware of the implications of their digital footprints and the need for digital privacy. Using the Clearing example, it is reasonable to assume that several individuals under 18 would provide information through this technology.

Perhaps most worrying is that current UK data privacy regulations allow individuals to request that their data be deleted from an organisation after a certain period. Whilst this may be possible using generative chatbots, the underlying algorithms of the technology will have already learned from the inputted data; thus true deletion of data may not be possible. In other words, the right to be forgotten is complicated with chatbots.

Hasal et al. (2021) states that if a chatbot can access the personal data of a user, the chatbot must have the GDPR mandates and regulations in place. Universities and educational institutions must establish clear and robust data collection, storage, and usage guidelines, strictly aligning with legislations such as the General Data Protection Regulation (GDPR) in the EU and the Children's Online Privacy Protection Act (COPPA) in the US. However, this is far more complex in practice.

Furthermore, another concern is that GDPR and the UK Data Protection Act 2018 (DPA) provide individuals with the 'right to be informed' about how their data is processed; however, overall algorithmic transparency is low (Meyer von Wolff et al., 2020). Whilst chatbots' algorithmic construction is known, there are few details on how it is implemented and its knowledge bases. Wolf et al. (2017) argue that this will 'never' be revealed by companies, which challenges data protection legislation.

Chatbots used in education are no longer rule-based models, as they employ NLP and ML techniques (Hasal et al., 2021). These techniques learn from a conversation that may contain personal information. Due to the nature of ML, it cannot learn from encrypted data, which presents additional challenges for policymakers that previous technologies did not. The author argues that there ought to be an open discussion to highlight the complexities of data storage obtained from users/students when using chatbots.

Simply put, data protection measures should ensure that data is only used for educational purposes, is stored securely, and is anonymised or deleted once it is no longer needed (Zeide, 2017).

Furthermore, students, parents, educators, and relevant stakeholders should be aware of the data protection measures. Transparency about data handling practices can alleviate privacy concerns and ensure the trust of all stakeholders in a particular adopted chatbot. The challenge of balancing the benefits of personalised education with the privacy rights of students will continue to be a critical issue in the application of chatbots and similar ML in education. As scholars further delve into this AI-driven educational paradigm, prioritising data privacy will be important to creating a sustainable and ethical AI-integrated educational environment.

Algorithmic bias

One of the defining attributes of chatbots such as ChatGPT is their ability to learn from diverse data sources (Qadir, 2022). Generally, enabling chatbots to deliver a wider range of responses and more nuanced interactions is considered an administrative and pedagogical advantage for education professionals. However, this also presents challenges such as algorithmic bias, a significant ethical concern arising when societal biases become encoded in our AI systems.

For several years, academics have warned about possible uneven effectiveness and lack of generalizability across populations in educational algorithms (Bridgeman et al., 2009; Ocumpaugh et al., 2014). Baker and Hawn (2021) argue that this concern became very salient to the general public in the 2020 UK GCSE and A-Level grading controversy, where the national qualifications regulator developed a set of formulas to assign predicted examination grades based on teacher predictions. The algorithm assigned poorer grades to students in state-funded schools and better grades (even better than teacher prediction) to students in smaller independent schools.

Like all AI systems, chatbots learn from large amounts of data gathered from the internet, which unavoidably represents societal biases. If the data used to train these models contains biased attitudes, the AI system will likely assimilate and reproduce these biases, even unintentionally (Bolukbasi et al., 2016). This could manifest as gender, racial, or other biases, significantly impacting a student's learning experience and worldview when surfaced in an educational context. For instance, if a university tutor used a chatbot to develop a scenario or case study, the technology may adopt gendered pronouns in certain contexts (for example, consistently referring to nurses as 'she' and engineers as 'he'), and this can perpetuate stereotypes and become problematic. It is important to note that even without AI, policies and regulations, such as GDPR, also risk reproducing societal bias and prejudices (Baker and Hawn, 2021). Nevertheless, this is an ethical dilemma that transfers the responsibility of ensuring unbiases from policymakers to educators.

Educators, policymakers, and AI developers must recognise these potential biases and take proactive steps to mitigate them. Firstly, the datasets used to train these AI systems should be diverse and representative to avoid amplifying societal biases. Nazer et al. (2023) argue that the issue stems from chatbots using data from either a single or narrow source, thus, propose that to ensure the data is truly representative, educational institutes should partner to share data. This appears to be a reasonable strategy as publicly available datasets are mostly underrepresented for many minority groups and, thus, lack diversity.

Furthermore, regular audits of the AI system's responses should be conducted to identify and rectify biases. This strategy is already taking place in the healthcare sector with the development of comprehensive frameworks and checklists to identify bias in diagnosis and medication (see Reddy et al., 2021; Nazer et al., 2023).

Moreover, there is a need for transparency about these biases and an ongoing dialogue about their implications. Students should be aware that chatbots may occasionally display biases, which could be critically evaluated rather than accepted as objective truths. For example, if a university has an AI chatbot that helps guide students through the process of selecting and applying for courses, the AI should state that it was trained on historical data, including data on which students applied for courses in the past. If, historically, female students were less likely to apply for Computer Science courses, the chatbot might have learned from this data. It could unintentionally discourage female students from applying to these courses. More specifically, when asked about the best courses for them, the chatbot might recommend humanities courses over computer science courses to female students based on past trends.

The issue of algorithmic bias highlights the importance of taking a deliberate and critical approach to developing and implementing AI in education. If these challenges are met, chatbots may be able to contribute positively to the educational landscape without perpetuating societal biases. Whilst the problem of bias in education (testing) has been documented since the 1960s and anticipated many aspects of future bias and fairness (see Hutchinson and Mitchell, 2019), scholars are now beginning to research societal bias, population bias, representative bias, aggregation bias, feedback bias, and reuse bias related to the machine learning lifecycle (Mehrabi et al., 2019; Silva and Kenney, 2019; Hellström et al., 2020; Barocas et al., 2023). However, little research exists on how education professionals and policymakers can practically mitigate dataset biases.

The author would like to re-emphasise that AI itself is not biased; AI systems learn from human-generated data, which can contain bias. The author argues that this is an important distinction in debates around debiasing platforms.

Student self-efficacy

AI chatbots can foster a learning environment where students can direct their educational journey to a significant extent by offering on-demand access to educational resources, providing explanations tailored to individual student needs, and providing a safe space to ask questions without fear of judgment. This level of autonomy is generally encouraged through contemporary educational strategies that promote self-directed learning, a method shown to increase student motivation, engagement, and learning outcomes (Wiliam, 2010). In other words, it allows learners to use software to learn individually, without the need for a class or a teacher (Shawar and Atwell, 2007). Learners benefit from immediate responses to questions and being guided through complex topics at their own pace.

However, while autonomy in learning is generally viewed positively, excessive autonomy has prompted concerns about the impact of AI on potentially lowering academic self-efficacy. For instance, whilst students get immediate responses, this may encourage them to rely solely on a chatbot for their learning.

Fryer et al. (2020) indicate that students becoming dependent on chatbots can lead to a lack of engagement and authentic learning experience, for instance. Furthermore, students may be discouraged from attending seminars, conducting the recommended reading, or participating in collaborative discussions. While this is not a new phenomenon, as educators have been grappling with concerns about technology's impact on academic self-efficacy for a while (Hilton, 2016), individuals' self-efficacy in using AI may differ significantly from their self-efficacy regarding using information technologies such as computers. Much of the literature in this area argues that the ability to make decisions is what differentiates AI technologies/products from traditional computer programs (Wang and Chuang, 2023). In other words, there is an absence of cognitive abilities in computer programmes, whereas AI attempts to reproduce this, which has contemporary implications for student self-efficacy. Computer self-efficacy has received much attention in prior studies (Compeau and Higgins, 1995; Teo and Koh, 2010), but few studies have researched AI self-efficacy. This paper highlights the significant difference between the two.

However, it must be noted that the importance of AI's growing flexibility and creativity in students has been acknowledged widely in the literature (Popenici and Kerr, 2017). With the boost of inclusive digital tools, students improve their creativity, technology applications, and other comprehensive skills supported by AI (Crittenden et al., 2018). Artificial intelligence pedagogy positively affects the development of students' information literacy, thereby enhancing students' self-efficacy (Loftus and Madden, 2020),

Furthermore, while chatbots are accredited for providing facts and explanations, the real-time nature of chat can encourage fast, reactive responses rather than thoughtful, reflective consideration. This might not always stimulate critical thinking, particularly if students are prioritising speed over depth of thought. In other words, chatbot technologies often promote brief, condensed forms of communication, which can sometimes limit the depth of discussion and critical thinking (Wang and Chuang, 2023). These skills are often fostered through more guided and interactive forms of instruction involving peer discussions, teacher-led debates, and collaborative projects. Thus, while chatbots can provide valuable support, educators can offer more than the full range of educational experiences in a more traditional learning environment.

In the literature, technology-centred perceptions, such as self-efficacy, have often been cited as crucial factors in the adoption of (new) information technology (IT) or information systems (IS) (Agarwal and Karahanna, 2000; Hsu and Chiu, 2004; Chen et al., 2011). Previous research on technology adoption has demonstrated that the higher a person's perceived self-efficacy regarding a particular application, the higher the perceived usefulness of that application (Igbaria, 1995; John, 2013; Lee and Ryu, 2013). However, since perceived self-efficacy is a highly domain-specific construct, more than perceptions of general self-efficacy measures may be required to cover the scope of AI adoption (Bandura, 2006). Chatbots' ability to promote autonomy in learning holds substantial promise for personalised, student-centred education. However, this autonomy must be carefully balanced with appropriate guidance and oversight from human educators in an increasingly multidimensional transformative way.

Plagiarism

In the context of integrating AI technologies into education, the issue of plagiarism emerges as a critical ethical concern (Teel et al., 2023). The facility of AI-powered tools such as ChatGPT may encourage students to misrepresent AI-generated outputs as their

own, thereby compromising the integrity of their academic work. This issue is particularly paramount in educational ecosystems that emphasise outcomes or end goals, such as grades or qualifications, over the learning process. For example, all phases of the UK's education systems have traditionally emphasised these quantifiable measures of academic success (Mansell, 2007).

The proficiency of chatbots generating sophisticated textual responses, solving intricate problems, and composing entire essays could create an environment conducive to academic dishonesty (Tlili et al., 2023). Given the emphasis on achieving high grades and qualifications, students may use AI-generated work to meet these goals, neglecting the importance of the learning journey itself (Els, 2022).

Interestingly, students may unintentionally breach academic integrity without realising it. For instance, a student might use a chatbot to assist them in a burdensome administrative task like filling out an ethics form. Unsurprisingly, AI could potentially identify more ethical risks for a research project related to data protection, confidentiality, and anonymity in a research project than a student might. Using AI to support a risk assessment may be useful, but there is certainly value in the student being able to identify and manage the ethical risks themselves. Additionally, for non-English speakers, inadequate command of the English language may present an obstacle for programmes that rely on essay writing assessments and may use chatbots as a powerful tool to improve several aspects of language such as vocabulary, spelling, punctuation, grammar, and syntax (Moya, 2023).

Addressing this issue requires a comprehensive and nuanced approach. Foremost, it is necessary to introduce an understanding of the importance of academic honesty and the negative implications of plagiarism on students' moral development and the integrity of their learning experience. This involves having clear policies and consequences surrounding academic dishonesty. The author has experience working with students unaware of what is and is not academic misconduct. This is particularly pronounced with international students who may be more familiar with academic best practices and ethical codes of conduct from their home country.

Additionally, deploying advanced plagiarism detection software capable of identifying AI-generated text is a practical step that can be implemented. However, as AI technologies evolve, so must our detection methodologies, necessitating continuous advancements in this field. Software such as Turnitin cannot detect essays written by AI because the text is originally generated and not copied. The author remains doubtful that development's plagiarism detection software will ever be one step ahead of AI technologies and be free of reporting false-positives.

Re-evaluating assessment strategies is another attractive measure. In systems that heavily emphasise outcomes, designing assessments that evaluate students' understanding and encourage original thinking, creativity, and skills currently beyond AI's reach becomes essential. King (2023, p. 3) encourages universities to design assignments that minimise the potential of cheating through platforms such as ChatGPT by incorporating a variety of assessment methods that go beyond traditional essay writing. For example, they could 'incorporate oral presentations, group projects, and hands-on activities that require students to demonstrate their knowledge and skills in a more interactive and engaging way'.

The author argues that oral presentations, such as viva voices and group projects, could be an effective assessment method to discourage plagiarism and promote learning outcomes. In other words, oral presentations must solely be done by a human, whereas the benefits of AI can still be realised to aid student preparation. Nevertheless, this approach may be considered a short-term solution to the constantly evolving AI technology, especially in the realms of online presentations and interviews. De Vries (2020) argues that deep fakes can blur the lines between what is fact and fiction by generating fake video footage, pictures and sounds. Similarly, AI-powered platforms such as AI Apply can quickly transcribe real-time questions posed during online presentations, formulate a rapid answer, and then vocalise it as if it were the student (Fitria, 2023). However, the author argues that this is a challenge that the wider society will likewise have to grapple with, as there will be implications for political deception, identity scams, and extortion (De Vries, 2020).

Transparency in education about the use of AI in the learning process is essential. While AI tools can be extremely beneficial in facilitating learning, these tools must be used responsibly. This includes acknowledging the assistance received from AI when students utilise its outputs. The author argues that this is similar to how MS Word spellchecker and Grammarly have become accepted in academia. While the risk of plagiarism associated with AI tools like ChatGPT is undoubtedly a concern, especially in higher education, where there is an increased focus on outcomes, it can be managed through effective education policies, advanced plagiarism detection techniques, innovative assessment strategies, and responsible AI use. The use of AI in education is not about eliminating AI altogether but rather about creating an educational environment where AI tools are used responsibly. This means using AI to augment the learning journey rather than compromising it.

Hallucinations

Hallucination or artificial hallucinations is a response generated by an AI, such as a language model which contains false or misleading information presented as fact (Ji et al., 2022). For example, when asked to generate ten examples of positivist education dissertation titles, a hallucinating chatbot might falsely state that interpretive studies were positivist.

The concept of AI hallucination gained widespread attention around 2022, coinciding with the emergence of large language models (LLMs) such as ChatGPT. Users noticed that these chatbots frequently inserted random falsehoods into their responses, seemingly without regard for their relevance or accuracy (Ji et al., 2022).

The term *AI hallucination* has been criticised for its anthropomorphic nature, as it draws an analogy between human perception and the behaviour of language models (Maynez et al., 2020). Thus, alternative terms such as *faithfulness* and *factuality* have been proposed to more accurately assess the accuracy and adherence to external knowledge sources of AI-generated content (Dong et al., 2020).

AI hallucinations can occur due to various reasons, including data discrepancies in large datasets, training errors during encoding and decoding, and a biased sequence (Ji et al., 2022). This poses a significant challenge for educators and students using generative chatbots. While this paper does not discuss the specifics of hallucination in natural language generation, the author acknowledges that it is important for educators and students to be aware of the particularly problematic issue of AI hallucinations.

For students, this can result in the development of misconceptions, which can have a long-term impact on self-efficacy, potentially affecting their understanding of key concepts or leading to different career choices (Emsley, 2023). Frequent encounters with AI hallucinations can decrease students' trust in AI as a reliable educational tool, and this distrust can extend to other digital learning resources and databases.

Educators face a unique ethical challenge when employing AI-generated content as classroom resources, as they hold the responsibility of ensuring the accuracy of the information presented. Emsley (2023) cautions educators and researchers about the falsifications that can be generated on a chatbot. In a study, investigating the authenticity and accuracy of references in case studies generated by ChatGPT, Emsley (2023) found that of 115 references that were generated, 47% were fabricated, 46% were authentic but inaccurate, and only 7% were authentic and accurate. However, Scharaschkin (2023) argues that while AI holds the potential to reduce teacher workload and improve grading reliability, it still requires close human supervision to safeguard against potential inaccuracies that can occur during AI hallucinations. This human involvement, nevertheless, introduces an additional layer of verification, which can in itself become a time-consuming and administrative burden, which may discourage its use. A similar phenomenon occurred when iPads were first introduced to the classroom, as many teachers argued that the additional workload outweighed any benefits they brought to teaching and learning (Williams, 2022).

Interestingly, Jennings (2023) argues that vetting AI-powered tools will play an important role for educators in the future but that it is only temporary. Jennings (2023, p.2) compares AI hallucinations to the early days of Wikipedia, where pages contained information full of un-cited opinions, 'The idea of citing a Wikipedia article as a source was laughable. Fast forward a few years, and it is it's become one of the most credible sources on the Internet. AI hallucinations will likely not be as big a problem in the future'. However, Ji et al. (2022) state that discrepancies between input and output are likely to continue, and that there are challenges ahead in first identifying and then mitigating hallucinations in NLG as research is preliminary in this area.

Conclusion

The integration of chatbots into education holds remarkable potential to revolutionise teaching and learning processes (Lund and Wang, 2023), such as providing personalised learning experiences to enhance student engagement. However, it is also interesting that their deployment brings increasing ethical considerations that must be navigated with thought.

The ethical landscape of AI in education contains complexities that require attention, evaluation, and adjustment. Similar to other transformative technologies, such as social media in the classroom, using AI comes with striking a reasonable balance of the benefits and shortfalls. As Williams (2022) argues in their study of social media and pedagogy, AI has the potential to both enhance and disrupt learning. Therefore, it is important to use AI in a way that maximises its benefits for practitioners and students, while minimising its risks relating to ethics and safeguarding. This will likely involve setting firm ethical boundaries to safeguard the interests of students, educators, and the broader educational community.

In conclusion, privacy considerations, although challenging, are manageable through policy and legislation. It is recommended that universities and educational institutions establish clear and robust data collection, storage, and usage guidelines, strictly aligning with legislations such as the General Data Protection Regulation (GDPR), however, there are challenges in relation to data deletion and the right to be informed. In fact, true deletion of data may not be possible. Thus, future research to understand the long-term ethical implications of data collected through AI in education would add significant value to this area.

Meanwhile, the issue of algorithmic bias demands thought about using datasets that are mostly underrepresented for many minority groups and, thus, lack diversity.

An overreliance on chatbots can lead to a lack of engagement and authentic learning experiences for students (Fryer et al., 2020), therefore, educators using AI are encouraged to foster autonomy without compromising student self-efficacy.

Plagiarism is a significant ethical concern that has been a common theme at universities for a while. Chatbots may encourage students to misrepresent AI-generated outputs as their own, thereby compromising the integrity of their academic work. Universities should continue to foster an environment that values academic integrity, using advanced plagiarism detection software, and rethinking assessment methods to discourage unethical practices (Teel et al., 2023). Finally, AI hallucination presents ethical challenges in terms of validating and verifying the accuracy of data generated by chatbots. Therefore, educators should hold some caution about the falsifications that can be generated on a chatbot.

Furthermore, it is the collective responsibility of all stakeholders in AI education, including developers, educators, policymakers, and students, to ensure that AI is used in a way that respects privacy, minimises bias, supports balanced learning autonomy, and upholds the vital role of human teachers.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

RW: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Agarwal, R., and Karahanna, E. (2000). Time flies when You're having fun: cognitive absorption and beliefs about information technology usage. *Manag. Inf. Syst. Q.* 24:665. doi: 10.2307/3250951

Alenizi, F. A., Abbasi, S., Hussein Mohammed, A., and Masoud Rahmani, A. (2023). The artificial intelligence Technologies in Industry 4.0: a taxonomy, approaches, and future directions. *Comput. Ind. Eng.* 185:109662. doi: 10.1016/j.cie.2023.109662

Baker, R. S., and Hawn, A. (2021). Algorithmic bias in education. *Int. J. Artif. Intell. Educ.* 32, 1052–1092. doi: 10.1007/s40593-021-00285-9

Bandura, A. (2006). "Guide for constructing self-efficacy scales," in *Self-efficacy beliefs of adolescents*. eds. T. Urdan and F. Pajares (Greenwich, CT: Information Age Publishing), 5, 307–337.

Barocas, S., Hardt, M., and Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. Cambridge, Massachusetts: MIT Press.

Biswas, S. (2023). Role of chat GPT in public health. *Ann. Biomed. Eng.* 51, 868–869. doi: 10.1007/s10439-023-03172-7

Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., and Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Adv. Neural. Inf. Process. Syst.* 29.

Bridgeman, B., Trapani, C., and Attali, Y. (2009). Considering fairness and validity in evaluating automated scoring. *National Council Meas. Educ.* Available at: https://www.academia.edu/20863443/Considering_Fairness_and_Validity_in_Evaluating_Automated_Scoring

Castillo, A. G. R., Silva, G. J. S., Arocutipa, J. P. F., Berrios, H. Q., Rodriguez, M. A. M., Reyes, G. Y., et al. (2023). Effect of chat GPT on the digitized learning process of university students. *J. Namib. Stud.* 33, 1–15. doi: 10.59670/jns.v33i.411

Chen, K., Chen, V., and Yen, D. C. (2011). Dimensions of self-efficacy in the study of smart phone acceptance. *Comput. Stand. Interfaces* 33, 422–431. doi: 10.1016/j.csi.2011.01.003

Compeau, D., and Higgins, C. (1995). Computer self-efficacy: development of a measure and initial test. *Manag. Inf. Syst. Q.* 19:189. doi: 10.2307/249688

Crittenden, W. F., Biel, I. K., and Lovely, W. A. (2018). Embracing digitalization: student learning and new technologies. *J. Mark. Educ.* 41, 5–14. doi: 10.1177/0273475318820895

De Vries, K. (2020). You never fake alone. Creative AI in action. *Inf. Commun. Soc.* 23, 2110–2127. doi: 10.1080/1369118x.2020.1754877

Dong, Y., Wang, S., Gan, Z., Cheng, Y., Cheung, J. C. K., and Liu, J. (2020) 'Multi-fact correction in abstractive text summarization,' *arXiv*. [Epub ahead of preprint] doi: 10.48550/arxiv.2010.02443

Els, G. (2022). "'A hybrid model in tourism postgraduate education – a learning journey" in *Team Academy in Diverse Settings*. eds. B. Urzelai and E. Vettraino (Routledge), 19–29.

Emsley, R. (2023). ChatGPT: these are not hallucinations – they're fabrications and falsifications. *Schizophrenia* 9:52. doi: 10.1038/s41537-023-00379-4

Fitria, T. N. (2023). "Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay," in *ELT Forum: Journal of English Language Teaching*. 12, 44–58.

Fryer, L. K., Thompson, A., Nakao, K., Howarth, M., and Gallacher, A. (2020). Supporting self-efficacy beliefs and interest as educational inputs and outcomes: framing AI and human partnered task experiences. *Learn. Individ. Differ.* 80:101850. doi: 10.1016/j.lindif.2020.101850

Hasal, M., Nowaková, J., Ahmed Saghair, K., Abdulla, H., Snášel, V., and Ogiela, L. (2021). Chatbots: security, privacy, data protection, and social aspects. *Concurr. Comput.* 33:e6426. doi: 10.1002/cpe.6426

Hellström, T., Dignum, V., and Bensch, S. (2020) 'Bias in machine learning - what is it good for?,' European Conference on Artificial Intelligence, pp. 3–10.

Hill-Yardin, E. L., Hutchinson, M. R., Laycock, R., and Spencer, S. J. (2023). A chat(GPT) about the future of scientific publishing. *Brain Behav. Immun.* 110, 152–154. doi: 10.1016/j.bbi.2023.02.022

Hilton, J. (2016). Open educational resources and college textbook choices: A review of research on efficacy and perceptions. *Educ. Technol. Res. Dev.* 64, 573–590. doi: 10.1007/s11423-016-9434-9

Hsu, M., and Chiu, C. (2004). Internet self-efficacy and electronic service acceptance. *Decis. Support. Syst.* 38, 369–381. doi: 10.1016/j.dss.2003.08.001

Hutchinson, B., and Mitchell, M. (2019). 50 years of test (un)fairness. *Decis. Support. Syst.* doi: 10.1145/3287560.3287600

Igbaria, M. (1995). The effects of self-efficacy on computer usage. *Omega* 23, 587–605. doi: 10.1016/0305-0483(95)00035-6

Jennings, J. (2023) *AI in Education: The Problem with Hallucinations*. Available at: https://www.esparklearning.com/blog/ai-in-education-the-problem-with-hallucinations/ (Accessed December 6, 2023).

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., et al. (2022) 'Survey of hallucination in natural language generation,' *arXiv*. [Epub ahead of preprint] doi: 10.48550/arxiv.2202.03629

John, S.P. (ed.) (2013) Antecedents and effects of computer self-efficacy on social networking adoption among Asian online users. Americas Conference on Information Systems.

King, M. R. (2023). A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cell. Mol. Bioeng.* 16, 1–2. doi: 10.1007/s12195-022-00754-8

Lee, D. Y., and Ryu, H. (2013). Learner acceptance of a multimedia-based learning system. *Int. J. Hum. Comput. Int.* 29, 419–437. doi: 10.1080/10447318.2012.715278

Lieberman, M. (2023) *What Is ChatGPT and How Is It Used in Education?*. Available at: https://www.edweek.org/technology/what-is-chatgpt-and-how-is-it-used-in-education/2023/01 (Accessed: November 7, 2023).

Loftus, M., and Madden, M. G. (2020). A pedagogy of data and artificial intelligence for student subjectification. *Teach. High. Educ.* 25, 456–475. doi: 10.1080/13562517.2020.1748593

Lund, B., and Wang, T. (2023). Chatting about ChatGPT: how may AI and GPT impact academia and libraries? *Soc. Sci. Res. Netw.* doi: 10.2139/ssrn.4333415

Mansell, R. (Ed.). (2007). The Oxford handbook of information and communication technologies. Oxford Handbooks Online. doi: 10.1093/oxfordhb/9780199548798.001.0001

Maynez, J., Narayan, S., Bohnet, B., and McDonald, R. (2020) 'On faithfulness and factuality in abstractive summarization,' *arXiv*. [Epub ahead of preprint] doi: 10.48550/arxiv.2005.00661

McCarthy, J., Minsky, M. L., Rochester, N., and Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, august 31, 1955. *AI Mag.* 27:12. doi: 10.1609/aimag.v27i4.1904

Mehrabi, N., et al. (2019) 'A survey on Bias and fairness in machine learning,' *arXiv*. [Epub ahead of preprint] doi: 10.48550/arxiv.1908.09635

Meyer von Wolff, R., Nörtemann, J., Hobert, S., and Schumann, M. (2020). Chatbots for the information acquisition at universities – a student's view on the application area. *Lect. Notes Comput. Sci.* 11970, 231–244. doi: 10.1007/978-3-030-39540-7_16

Microsoft (2023) *What is Microsoft's Approach to AI?* Available at: https://news.microsoft.com/source/features/ai/microsoft-approach-to-ai/ (Accessed April 1, 2012).

Mollick, E., and Mollick, L. (2022). New modes of learning enabled by AI chatbots: three methods and assignments. *Soc. Sci. Res. Netw.* 1–10. doi: 10.2139/ssrn.4300783

Moya, B. A. (2023). How can we use artificial intelligence Chatbots safely? *Academic Integrity Lessons*:71.

Nazer, L., Zatarah, R., Waldrip, S., Ke, J. X. C., Moukheiber, M., Khanna, A. K., et al. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digit. Health* 2:e0000278. doi: 10.1371/journal.pdig.0000278

Ng, A. (2016) 'What artificial intelligence can and Can't do right now,' *Harv. Bus. Rev.*, Available at: https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now (Accessed March 1, 2023).

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., and Heffernan, C. (2014). Population validity for educational data mining models: a case study in affect detection. *Br. J. Educ. Technol.* 45, 487–501. doi: 10.1111/bjet.12156

Oppenheimer, T. (1997) 'The computer delusion,' *Atl. Mon.*, Available at: https://www.theatlantic.com/magazine/archive/1997/07/the-computer-delusion/376899/ (Accessed May 9, 2022).

Popenici, S., and Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Res. Pract. Technol. Enhanc. Learn.* 12:22. doi: 10.1186/s41039-017-0062-8

Qadir, J. (2022) 'Engineering education in the era of ChatGPT: promise and pitfalls of generative AI for education,' *TechRxiv*. [Epub ahead of preprint] doi: 10.36227/techrxiv.21789434.v1

Reddy, S., Rogers, W., Makinen, V. P., Coiera, E., Brown, P., Wenzel, M., et al. (2021). Evaluation framework to guide implementation of AI systems into healthcare settings. *BMJ Health Care Inform.* 28:e100444. doi: 10.1136/bmjhci-2021-100444

Scharaschkin, A. (2023) 'Hallucinations do not limit AI's power to transform education,' *AQA*. Available at: https://www.aqa.org.uk/news/hallucinations-do-not-limit-ais-power-to-transform-education (Accessed December 6, 2023).

Shawar, B., and Atwell, A. (2007) *Fostering language learner autonomy through adaptive conversation tutors.* Available at: https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2007/51paper.pdf

Shen-Berro, J. (2023) *New York City schools blocked ChatGPT. Here's what other large districts are doing*. Available at: https://www.chalkbeat.org/2023/1/6/23543039/chatgpt-school-districts-ban-block-artificial-intelligence-open-ai (Accessed December 7, 2023).

Silva, S., and Kenney, M. (2019). Algorithms, platforms, and ethnic bias. *Commun. ACM* 62, 37–39. doi: 10.1145/3318157

Stokel-Walker, C., and Van Noorden, R. (2023). What ChatGPT and generative AI mean for science. *Nature* 614, 214–216. doi: 10.1038/d41586-023-00340-6

Teel, Z., Wang, T., and Lund, B. (2023). ChatGPT conundrums: probing plagiarism and parroting problems in higher education practices. *Coll. Res. Libr. News* 84:205. doi: 10.5860/crln.84.6.205

Teo, T., and Koh, J. H. L. (2010). Assessing the dimensionality of computer self-efficacy among pre-service teachers in Singapore: a structural equation modeling approach. *Int. J. Educ. Dev. Using ICT* 6, 7–18. Available at: http://files.eric.ed.gov/fulltext/EJ1085024.pdf

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., et al. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learn. Environ.* 10:15. doi: 10.1186/s40561-023-00237-x

Turing, A. (1950). I.—computing machinery and intelligence. *Mind* LIX, 433–460. doi: 10.1093/mind/lix.236.433

Wang, Y., and Chuang, Y.-W. (2023). Artificial intelligence self-efficacy: scale development and validation. *Educ. Inf. Technol.*, 1–24. doi: 10.1007/s10639-023-12015-w

Watermeyer, R., Phipps, L., Lanclos, D., and Knight, C. (2023). Generative AI and the automating of academia. *Postdigit. Sci. Educ.*, 1–21. doi: 10.1007/s42438-023-00440-6

Wiliam, D. (2010). The role of formative assessment in effective learning environments. *Educ. Res. Innov.*, 135–159. doi: 10.1787/9789264086487-8-en

Williams, R. (2022). *An exploration into the pedagogical benefits of using social media: Can educators incorporate social media into pedagogy successfully?* (Doctoral dissertation), Teesside University Repository.

Williams, R. T. (2023). Think piece: ethics for the virtual researcher. *Practice* 5, 41–47. doi: 10.1080/25783858.2023.2179893

Wolf, M. J., Miller, K., and Grodzinsky, F. S. (2017). Why we should have seen that coming. *ORBIT J.* 1, 1–12. doi: 10.29297/orbit.v1i2.49

Zeide, E. (2017). The structural consequences of big data-driven education. *Big Data* 5, 164–172. doi: 10.1089/big.2016.0061

Zhai, X. (2022). ChatGPT user experience: implications for education. *Soc. Sci. Res. Netw.* doi: 10.2139/ssrn.4312418

Zhao, J., Wu, M., Zhou, L., Wang, X., and Jia, J. (2022). Cognitive psychology-based artificial intelligence review. *Front. Neurosci.* 16:1024316. doi: 10.3389/fnins.2022.1024316