

Analysis of the Possible Issues Caused by the Application of GenAI in Education and the Corresponding Solutions

Mengjie Yang

School of Foreign Language, Sichuan Normal University, Chengdu, Sichuan 610101, China

Abstract. This paper examines the multifaceted problems arising from the integration of Generative Artificial Intelligence (GenAI) in educational systems, focusing on three critical dimensions: hallucination, unfairness, and privacy. First, the study interrogates the epistemic crisis triggered by GenAI's propensity for hallucination—generating plausible but factually inaccurate content—arguing that its uncritical adoption in education risks entrenching misinformation and undermining intellectual autonomy in the post-truth era. Second, it critiques systemic unfairness rooted in global asymmetries, where GenAI entrenches data colonialism by disproportionately benefiting technologically advanced regions while marginalizing under-resourced communities. Third, the paper explores privacy violations through the lens of Foucault's digital panopticon, demonstrating how GenAI-enabled surveillance—via behavioral tracking, cognitive profiling, and data monetization—transforms classrooms into sites of pervasive control. To address these challenges, the study proposes interdisciplinary solutions, and by synthesizing critiques of epistemic reliability, global justice, and privacy erosion, this paper contributes to urgent debates on balancing GenAI's transformative potential with its risks to the field of education in an increasingly algorithm-driven world.

Keywords: GenAI; AI in Education (AIED); Hallucination; Unfairness; Privacy.

1. Introduction

Artificial General Intelligence (or AGI, Strong AI) “refers to a type of AI that can understand, learn, adapt, and implement knowledge in a wide variety of tasks at the level of a human being” [1]. The wide application of generative AI systems from November late 2022 with the launch of ChatGPT has indeed showcased the remarkable momentum driving AI development. This shift from Narrow AI (or Weak AI) which was “predominantly discriminative AI” and “primarily focuses on classifying and searching information” [2], to GenAI has significantly impacted the public sphere not only technically, but also in the field of education. GenAI is reshaping both teaching methodologies and learning dynamics, forcing educators and institutions to confront unprecedented opportunities and challenges.

Education not only enables transmitting from one generation to another the ideas and principles for the development of the personality and social integration of the individual, but it is also a powerful tool for reducing poverty and inequality, improving health and social well-being, enhancing identity, and achieving economic growth [3]. Therefore, it is urgent to hold a discussion on the issue of AI in Education (AIED) in the age when the maturity of education systems fails to keep pace with the rapid development of industry and media. In 2019, the World Economic Forum, through a crowdsourcing initiative, began outlining its vision for students education with the title of “Education 4.0 Framework”, which was designed to align with the Fourth Industrial Revolution [4]. Since then, we are on the path to enter the era of “Education 4.0”, which will take place in “the middle of the 21st century-Learning inheres in every thing; students can learn anywhere; the school is electronic and mobile and is connected to artificial intelligence.”[5] On the contrary, the dim reality in that we are still on the slowly heading path from Education 2.0 to Education 3.0, just as Hungarian researchers John Senior and Éva Gyarmathy pointed out that, “In a very few countries, personalized education that is assisted with solutions from digital technology is the norm, but in several countries, the central goal is still the mass education of version 2.0, with a single central curriculum and uniform implementation”[6], which indicates that the average development of education worldwide has lagged far behind the requirements of the times, when Industry 4.0 and Media 4.0---characterized by comprehensiveness---have reached respectively in 2014 and 2020. Therefore, it is undeniable that AI

in Education is an issue that has emerged as both a potential solution and a disruptive force in this critical juncture. While manufacturing sectors have embraced smart automation and media ecosystems have transformed through immersive technologies like AR/VR and big data analytics, education systems remain largely anchored in 20th-century pedagogical frameworks. This growing chasm between technological advancement and educational evolution raises urgent questions about safe usage, ethical governance, global equity and so on. For instance, The West Virginia Department of Education(2024), published guidelines for AI use in schools, showcasing predictable risks include over-reliance on AI technologies, challenges to independent and creative thinking, reduced social interactions, privacy and safety issues, furthering of digital divides, as well as plagiarism and cheating. These risks are not isolated cases but manifestations of a deeper systemic conflict between the rapid iteration of technology and the inertia of educational structures.

2. Hallucination: GenAI’s Epistemic Crisis in Post-Truth Era

GenAI tools are transferring the traditional approach of education through providing learners generated teaching texts instead of time-tested authorized materials. “In the future, content generated through human GenAI conversations may become one of the main sources of knowledge production.” [7] Nonetheless, as learners’ exposure to published textbooks being possibly cut down by the convenience and instance the GenAI information giver provides, the unverified resources and claims will take an overwhelming role in providing learning materials. However, since what the the GenAI system offers is with a high likeness to be fake and inaccurate on account of hallucination deeply embedded in its inner generation mechanism, The content it generates appears to be truthful will mislead the learners especially the novices in a certain field who usually failed to recognize the authentic resources from the crowd of the fabricated ones.

Hallucination is possible because the AI doesn’t actually “know” anything and nor does it “think,” [8] Presently, the majority of GenAI models function primarily as sophisticated predictive tools. They are trained to anticipate the sequences of words that humans would likely generate, relying on the patterns and data they were trained on, rather than understanding meaning or context in a comprehensive way. “In fact, generative AI simply tries its best with polite efforts at completion, like a human eager to please who embellishes what they say with unnecessary, and sometimes made-up, certainty.” [9] When prompted to generate content outside common training data patterns, some GenAI models frequently exhibit context collapse, losing track of core premises in longer interactions. Thus, Merchant (2023) noted that ethicists are concerned that large language models (LLMs), such as Open AI’s ChatGPT, could add to the misinformation tunnel on the Internet due to their human-like intelligence. [10] Additionally, “The algorithms that power these tools deliberately conceal the sources they draw from in order to make the writing they produce read like original content.” [11] and “It has long been recognized that artificial neural networks (ANNs) are usually ‘black boxes’; that is, that their inner workings are not open to inspection” [12]. As a consequence, the inner generating logic of GenAI is unexplainable, failing to make how the outputs were generated transparent and visualized. Therefore, GenAI’s self-reliant interpretation usually surpasses the honest and accurate representation of stored information online, leading to the paradoxes between fictional facts and reality. The statistical nature of their predictions means they inherently mirror and amplify biases present in training data, while remaining oblivious to the ethical implications of their outputs.

Users may credulously accept false information due to the “rationality” of AI outputs, such as legal statues and medical advice, leading to decision-making biases. This “technological simulacrum” masks the fictional nature of content by mimicking the logic of human language. In French sociologist Jean Baudrillard’s theory of simulacra, simulacra disrupt the connection with reality, absorbing reality into themselves and rendering the distinction between reality and unreality meaningless, thus achieving a state of “hyperreality.” As Jean Baudrillard pointed out that, “We live in a world where there is more and more information, and less and less meaning” [13], in this sense, the erroneous content generated by AI hallucinations, like simulacra, is detached from objective reality and

represents a distorted representation of truth. Baudrillard argued that simulacra have taken a dominant place in postmodern society, and individual's cognition is shaped by symbolic systems such as media and technology. AI hallucinations are an extreme manifestation of this trend: algorithm-generated simulacra not only replace reality but even become new cognitive benchmarks.

Governments should actively promote AI literacy courses in primary, secondary, and higher education to teach students to identify AI-generated content, detect logical flaws, and verify information sources. Additionally, it is essential to standardize the transparency of GenAI tools by introducing policies that require GenAI developers to disclose whether content is generated by AI, such as adding watermarks or metadata tags. Special funds should be allocated to develop tools capable of flagging unreliable AI-generated content (such as plagiarism detectors adapted to identify hallucinations), and partnerships with tech enterprises should be established to launch open-source solutions for schools. Schools should actively innovate teaching methods by shifting the focus of assessments from content reproduction to process-oriented evaluation. For example, students could be required to submit research logs, peer-reviewed drafts, or oral defenses to reduce their reliance on AI-generated text as a substitute for critical thinking. In this way, educators can track some of the processes by scaffolding assignments, requesting outlines and drafts, tracking changes and version histories, or they can change to a product that is less susceptible to AI use, such as oral reports or in-class exams [14], implementing self annotation and visible reflection through metacognition; they can also “encourage revision for examining and improving the ideas themselves” [15]. At the same time, professional training should be provided for teachers to enable them to recognize AI hallucinations and guide students in verifying information sources. Furthermore, clear guidelines should also be established by schools to prohibit submitting AI-generated hallucinated content as original work.

3. Unfairness: Global Asymmetries and Data Colonialism GenAI cases

While developed regions are promoting AI-powered personalized learning platforms, many low-income areas still struggle with basic internet connectivity and device accessibility. This disparity could solidify educational inequities, as students in under-resourced schools lack the digital literacy training and AI tools necessary to compete in an increasingly technology-driven society. Since apart from its iterative innovations in AI architectures and training methods, GenAI is reliant on a massive number of data and strong computing power for the whole system to develop, “which are mostly only available to the largest international technology companies and a few economies (mostly the United States, People’s Republic of China, and to a lesser extent Europe)” [16], with the swift spread of GenAI across tech-intensified countries, especially those in the north of earth, the gap between the access of digital learning resources produced by GenAI in the North and South globally will be greatly deepened, exacerbating disparity not only technically, but also economically and culturally. As a result, regions with limited data access will be trapped in “data poverty” on account of “digital divide” [17], facing a heightened possibility to be marginalized, as well as being vulnerable to progressive cultural or ideological colonization through the norms encoded in GenAI systems in the process of enrolling GenAI in educational system. In this way, the uneven spread of GenAI may further entrench data colonialism, which means northern countries, leveraging their computational power advantages and data monopolies, export digital standards embedded with their own values through the underlying architecture of GenAI systems (such as algorithmic logic and training data biases). For example, when southern countries rely on GenAI tools developed in the north for critical sectors like education and healthcare, their indigenous knowledge systems such as traditional medical practices and oral cultures may be automatically filtered out or undervalued by these systems, leading to cultural identity crises. Additionally, economic data dependency could relegate southern countries to the role of suppliers of digital raw materials (e.g., providing cheaply labeled data or computational power outsourcing) in the algorithm-driven global value chain, while northern enterprises capture core technological profits, creating a new form of technological rent.

Moreover, unfairness that GenAI's application in educational field brings about not only lies in the disparity in economic development between regions, but also is related to the inner organization of GenAI systems. Stanford computer scientists designed an experiment to find that the AI detectors such as Turnitin[18] have a strong possibility to flag international students' writings as false positives ones, but almost never made such mistakes when assessing the work of native English speakers.[19] This bias stems from the inherent design flaws of AI detection systems. Most detectors are trained on writing samples dominated by native English speakers, with algorithms implicitly treating complex vocabulary and diverse grammatical structures as markers of human authorship. For non-native speakers, however, constrained by language proficiency, their writing often relies on simpler sentence patterns and common word choices--a linguistic feature that overlaps with the "predictable expression" characteristics of AI-generated text. The dominance of English corpus in language model training may suppress cultural expressions in minority languages, while algorithmic recommendation mechanisms could subtly reinforce Western-centric narratives. Unlike traditional forms of colonialism, this "soft colonialism" achieves cognitive homogenization through technological penetration, causing southern countries to gradually lose cultural autonomy without awareness.

To address this challenge, take bias complaints seriously so AI does not replicate historic injustice, unfairness, or discrimination in data or algorithms, and the international community must establish an inclusive AI governance framework: on one hand, transferring data technologies and computational resources to southern countries through a mechanism of digital development re-balancing, such as establishing global public data pools and open-source algorithm communities; on the other hand, incorporating cultural diversity protection clauses into GenAI ethical guidelines to require system developers to respect indigenous knowledge sovereignty and avoid universalizing single cultural standards.

4. Privacy: Surveillance and Digital Panopticon from GenAI

GenAI models generally rely on large datasets—including texts, audios, images, and codes—scraped from the internet, frequently without obtaining authorization from the original content creators. As a result, many GenAI systems have been accused of violating intellectual property rights. The proliferation of AI systems in educational environments creates direct entry points into sensitive spaces where these technologies accumulate unprecedented amounts of personal data. Such information risks being harvested for algorithmic training purposes without explicit permission from educators and learners, or repurposed for commercial gain—including monetization through practices such as behavioral profiling and personalized marketing campaigns targeting vulnerable learner populations. As the scholar Priten Shan pointed out that "These companies could also share student data with third parties, including colleges, major corporations, insurance companies, or whomever they can sell the data to if the right policies and regulations are not implemented." [20] What's worse, although those companies refuse to offer personal information to other agencies, "AI systems are vulnerable to cyber-crime which may cause damage if used maliciously" [21]; "There is the hardcore security issue of chatbot jailbreaking, phishing, DoS and other cybersecurity issues". [22] "Moreover, despite calls for regulation from the AI industry itself, the drafting of legislation on the creation and use of all AI, including GenAI, often lags behind the rapid pace of development". [23]

Moreover, many artificial neural networks operate as "black boxes," which means "their inner workings are not open to inspection" [24]. Thus, the lack of transparency in how GenAI models process and utilize data further complicates the issue, making it difficult for educators, students, and policymakers to understand what data is being collected, how it is being used, or who it is being shared with. This opacity not only undermines trust in these technologies but also hinders the development of effective regulatory frameworks. Without clear visibility into data practices, it becomes nearly impossible to hold companies accountable for privacy violations or unethical data use.

In contemporary society, the rise of GenAI technology has endowed Micheal Foucault's description of "panopticon" with new forms of expression, evolving into a privacy crisis in the digital age. "Technology continually restructures the conditions of human experience. It shapes our relationships, values, landscapes, and expectations. It alters power relationships" [25]. Through big data collection and algorithmic analysis, GenAI systems construct a monitoring network far more vast and secretive than physical spaces previously. In the educational field, the "transparent classroom" under GenAI surveillance embodies the penetration of discipline from behavior to cognition by digitizing students' in-class behaviors. As Foucault noted, "Surveillance is permanent in its effects, even if it is discontinuous in its action" [26], smart cameras and voice recognition systems are implemented to analyze real-time classroom performance (such as eye gaze direction, speech frequency, sitting posture duration, etc.), transforming these into quantified data for concentration or participation and upgrading the limited physical monitoring of teachers to a multi-dimensional digital way of gaze. What's more, students' data collected by GenAI through educational platforms includes not only academic performance but also social relationships (such as students' interactions in online discussion groups), mental health (such as analyzing students' depressive tendencies through composition texts), and consumption habits (such as students' purchase records in educational apps). GenAI will also help to generate cognitive maps of students based on their answer records and browsing trajectories to precisely push customized learning content. While this appears to be teaching according to aptitude, it actually prescribes cognitive paths through algorithms, potentially suppressing students' ability to form or express personalized views out of GenAI's viewpoint, reflecting the covert manipulative power of AI-influenced personalized learning systems. Educational institutions or enterprises may use anonymized data for commercial cooperation, such as sharing student learning data with educational technology companies, leading to privacy exposure. Students may also be misled by algorithmic biases, such as being automatically categorized as low-potential students due to regional or family background labels, which may affect life trajectories such as college entrance and career choices. In short, the panoramic profile of educational data generated with GenAI's assistance actually reflects a chain from privacy to power. With the lengthy and complex privacy agreements that students check when using educational apps, students show a sense of powerlessness as "data prisoners".

To reduce the lease and misuse of students' private information in GenAI usage, governments need to restrict the scope of data collected by AI and only retains necessary academic performance data while deleting sensitive information such as biometrics and family privacy to avoid excessive surveillance for disciplinary purposes. Additionally, grant students the right to access, correct, and delete their own educational data, as well as to appeal algorithmic evaluation results, transforming their roles from passive acceptance of surveillance into active participation in management.

5. Conclusion and Prospects

This paper analyzes the inequality phenomena arising from GenAI's application in education through three lenses: hallucination, unfairness, and privacy. Hallucination-driven misinformation undermines educational integrity, while global data colonialism and algorithmic biases exacerbate digital divides. The digital panopticon of AI surveillance further infringes on student privacy and autonomy. Current solutions—including AI literacy programs, regulatory transparency, and privacy-by-design frameworks—offer pathways to mitigate these risks, but their effectiveness depends on cross-stakeholder collaboration and equitable resource allocation. Future research should prioritize longitudinal studies on GenAI's socio-cognitive impacts, particularly its role in shaping learner autonomy and epistemic trust.

References

- [1] Shan, P. *AI and the future of education*. John Wiley & Sons, Inc., 2023, p.17.

- [2] Ibid., p.16.
- [3] Hajji, S. Educational sovereignty and artificial intelligence challenges: The case of Morocco. In F. Roumate (Ed.), *Artificial intelligence in higher education and scientific research: Future development*. International Institute of Scientific Research, 2023, p.104.
- [4] World Economic Forum. Schools of the future: Defining new models of education for the fourth industrial revolution, 2020. <https://www.weforum.org/publications/schools-of-the-future-defining-new-models-of-education-for-the-fourth-industrial-revolution/>.
- [5] Senior, J., & Gyarmathy, É. *AI and developing human intelligence*. Routledge, 2022, p.12.
- [6] Ibid.
- [7] UNESCO. Guidance for generative AI in education and research. United Nations Educational, *Scientific and Cultural Organization*, 2023, p.36.
- [8] Shan, P. *AI and the future of education*. John Wiley & Sons, Inc., 2023, p.23.
- [9] Clark, D. *Artificial intelligence for learning: Using AI and generative AI to support learner development*. Kogan Page Limited, 2024, p.207.
- [10] Merchant, B. Business: Who’s helping train AI? Almost all of us. *Los Angeles Times*, 2023, August 26.
- [11] Writing Across the Curriculum. (2023, Winter). *AI-resistant assignments*. Carleton College. <https://www.carleton.edu/writing/resources-for-faculty/working-with-ai/ai-resistant-assignments/>.
- [12] UNESCO. Guidance for generative AI in education and research. United Nations Educational, *Scientific and Cultural Organization*, 2023, p.15.
- [13] Baudrillard, J. *Simulacra and Simulation* (S. F. Glaser, Trans.). University of Michigan Press, 1994, p.79.
- [14] Hernandez, M. AI-resistant assignments? Show student thinking and promote better writing with UChicago-supported tools. *Academic Technology Solutions*, University of Chicago, 2023. <https://academictech.uchicago.edu/2023/02/08/ai-resistant-assignments-show-student-thinking-and-promote-better-writing-with-uchicago-supported-tools/>.
- [15] Writing Across the Curriculum. (2023, Winter). *Ethical issues in AI*. Carleton College. <https://www.carleton.edu/writing/resources-for-faculty/working-with-ai/ethical-issues/>.
- [16] UNESCO. Guidance for generative AI in education and research. United Nations Educational, *Scientific and Cultural Organization*, 2023, p.14.
- [17] National Telecommunications and Information Administration (NTIA). Falling through the net: A survey of the “Have Nots” in rural and urban America. *The Digital Divide*. MIT Press, 1995, p.42.
- [18] Turnitin is a software used by over 16,000 academic institutions worldwide to spot plagiarized text.
- [19] Mathewson, T. G. AI detection tools falsely accuse international students of cheating. *The Markup*, 2023. <https://themarkup.org/machine-learning/2023/08/14/ai-detection-tools-falsely-accuse-international-students-of-cheating>.
- [20] Shan, P. *AI and the future of education*. John Wiley & Sons, Inc., 2023, p.206.
- [21] Bhise, A., Munshi, A., Rodrigues, A., & Sawant, V. Overview of AI in education. In P. Churi, S. Joshi, M. Elhoseny, & A. Omrane (Eds.), *Artificial intelligence in higher education: A practical approach*. CRC Press, 2023, p.55.
- [22] Clark, D. *Artificial intelligence for learning: Using AI and generative AI to support learner development*. Kogan Page Limited, 2024, p.208.
- [23] UNESCO. Guidance for generative AI in education and research. United Nations Educational, *Scientific and Cultural Organization*, 2023, p.15.
- [24] Ibid.
- [25] Sandler, R. L. (Ed.). *Ethics and emerging technologies*, 2014, p.23. <https://doi.org/10.1057/9781137349088>.
- [26] Foucault, M. *Discipline and punish: The birth of the prison* (2nd ed.). Vintage Books, 1995, p.201.