Digital Society (2026) 5:10
https://doi.org/10.1007/s44206-026-00247-5

REVIEW PAPER

The Future of Education in Digital Environments: Editorial for the Topical Collection *AI in Education: Ethical and Epistemic Perspectives*

Aleksandra Vučković[1] · **Vlasta Sikimić**[2] · **Marina Budić**[3]

Received: 29 November 2025 / Accepted: 21 January 2026

Abstract
The topical collection *AI in Education: Ethical and Epistemic Perspectives* examines the transformative impact of advanced digital solutions on education, with a focus on large language models (LLMs) and predictive models. As AI tools increasingly shape teaching, learning, and assessment practices, a need for philosophical inquiries, especially those concerning responsibility, intellectual virtues, fairness, and inclusion, emerges. The papers in this collection examine the potential of AI tools to both enhance and undermine learning processes, pressing questions about cognitive offloading, deskilling, and the digital divide, along with the impact of these tools on human agency. AI tools, as presented in this collection, can also improve accessibility, personalization, and inclusion in educational environments. The collection argues that responsible implementation of AI in education demands sustained ethical, epistemic, and pedagogical reflection. The main conclusion is that the future of education depends not on whether advanced digital solutions are adopted, but on how critically and responsibly they are integrated.

Keywords AI · Education · Learning · Epistemology of AI · AI ethics

✉ Aleksandra Vučković
aleksandra.vuchkovic@gmail.com

Vlasta Sikimić
v.sikimic@tue.nl

Marina Budić
marina.budic.93@gmail.com

[1] Department of Philosophy, Faculty of Philosophy, Institute for Philosophy, University of Belgrade, Belgrade, Serbia

[2] Department Industrial Engineering & Innovation Sciences, Philosophy & Ethics, Eindhoven University of Technology, Eindhoven, The Netherlands

[3] Centre for Philosophy, Institute of Social Sciences, Belgrade, Serbia

Published online: 12 March 2026

Advanced digital tools, such as intelligent tutoring systems and Large Language Models (LLMs), have already made a significant impact on education. Teachers are asked to rethink assessment methods, while students increasingly use advanced digital technologies for learning. The AI boom in education opens urgent philosophical questions, as recommendations for the responsible use of AI arise from recognizing epistemic and ethical concerns about the reliability and sustainability of these tools (Milano et al., 2023; Fütterer et al., 2025). A crucial epistemic dimension concerns the impact of digital technologies on critical thinking and the question of whether AI tools can foster intellectual virtues. As AI systems enable cognitive offloading, it is essential to distinguish cases in which they *extend* learners' abilities (e.g., Pritchard et al., 2021) from those in which they *impede* them, for instance, by diminishing independent thinking.

On ethical grounds, it is important to mitigate algorithmic bias, as they misrepresent real-world data and, in some circumstances, reflect harmful prejudices (Baker & Hawn, 2022; Austin et al., 2023; Khan, 2023; Chinta et al., 2024; Boateng & Boateng, 2025; Suchithra & Arya, 2025). To ensure an unbiased approach and avoid the perpetuation of social injustice, AI tools for education must be transparent, safe, and trustworthy (Vincent-Lancrin & Van der Vlies, 2020; Li et al., 2023; García Peñalvo et al., 2024). Students should be evaluated on academic merit and independently of their prior performance, social status, or personal background. For this reason, both educators and learners need to know *how* AI systems make predictions about future performance and *why* they recommend particular assignments. Moreover, they have to be prepared to attune and override biased and ungrounded algorithmic decisions.

The development and use of AI tools must strike a balance between gathering sufficient information to support students and avoiding excessive knowledge that could infringe on their privacy. This requires continuous oversight at all levels, from developers to teachers, and, finally, students and their caretakers. However, responsible AI development and implementation are not limited to minimizing the potential damages. When properly trained, AI tools can contribute to the personalization of education, thus potentially increasing inclusion and promoting equity (Mohammed & 'Nell'Watson, 2019; Salas-Pilco et al., 2022; Xia et al., 2022; Sikimić, 2023; Julien, 2024; Pagliara et al., 2024; Viberg et al., 2024; Gupta & Kaul, 2024). Consequently, the pursuit of bias elimination and greater inclusion in education is both an epistemic and an ethical challenge.

A promising research area is the global impact of the evolving interplay between advanced digital solutions and teaching practices. For instance, the widespread availability of AI tools creates new opportunities for language learning and translation, thereby expanding educational prospects and providing greater access to students from the Global South (Vučković & Sikimić, 2025). On the other hand, data samples for AI training may not be sufficiently diverse, which contributes to the underrepresentation of minorities. To ensure equity, we need thorough considerations of the responsibilities and boundaries of AI use in classrooms. These are not merely intellectual exercises, but have practical consequences, as they inform future AI and education policies. Strong epistemic and ethical foundations promote fair and reliable use of emerging technologies and can shape the ongoing digitalization in ways that benefit all students.

To address this need, we set out for an ambitious task of building a community around philosophical reflections on the responsible use of AI in education. As a first step, in March 2024, we organized a workshop on the ethical and epistemic aspects of the application of LLMs, predictive models, and other advanced digital solutions in education. The workshop was held at the Eindhoven University of Technology, with participants joining both in-person and online. Over the course of two days (6–7 March 2024), 18 presentations were delivered, accompanied by lively and thought-provoking discussions.[1] The workshop's primary objective was to foster critical dialogue among scholars regarding the transformative potential and inherent risks these technologies pose to educational systems worldwide. The remarkable depth of engagement and scholarly exchange that emerged from these discussions catalyzed our decision to expand this discourse into a written collection, enabling sustained exploration of these issues through academic inquiry.

As the second step of our initiative, we issued an open call for papers for the topical collection in *Digital Society*, titled *AI in Education: Ethical and Epistemic Perspectives*. We chose this title because it was sufficiently broad to encompass the range of insights about various AI technologies, yet focused enough to inspire the authors to consider the specific effects these tools have had on the field of education. All the papers went through the double-blind peer review process, ensuring the highest quality of the publications assembled for this collection. While some submissions came from participants in our workshop, many were authored by researchers with whom we had not previously collaborated, highlighting the broad scholarly interest in these issues.

The collection now includes seven research papers on topics such as the epistemic and ethical consequences of AI implementation in classrooms, intellectual virtues, benefits and dangers of technologically induced schooling transformations, and so on. In the rest of this article, we will briefly present each of these contributions.

The rapid development of advanced digital technologies, coinciding with the COVID-19 pandemic and the proliferation of misinformation, has generated a sense of uncertainty among many. In the paper "AI-Informed Pedagogy for a Post-Truth Era" (2025), David Nally explores the complex interaction between AI development and social challenges in the post-truth era, as well as the implications of this relationship for educational systems. To preserve the epistemic goals of education and avoid issues such as improper cognitive offloading of curricula to AI or perpetuation of social bias, Nally argues that we need to rethink the roles in schooling. Through the exploration of six possible human-AI relationships in education (ranging from the full delegation of tasks to AI to its complete exclusion), Nally advocates for an AI-informed pedagogy where humans and artificial agents each perform the tasks to which they are best suited.

Still, not all forms of uncertainty are negative. In the article "Harnessing ambiguity. An epistemic-ethical lens for analyzing deeper student engagement with LLMs in challenge-based learning" (2025), Gunter Bombaerts examines the dual nature of the notion of ambiguity: as a source of uncertainty and a catalyst for (further) explora-

[1] More information about the event, including the full program and the book of abstracts, is available at the following link: https://ephil.ai/event/workshop-ai-education/.

tion. In the era of LLMs, overt aversion to ambiguity could lead students to use these tools uncritically, as they may settle for simplified and superficial solutions. Conversely, if properly utilised, ambiguity can serve as fuel for inquiry and creativity, as in the case of challenge-based learning, where students are encouraged to tackle real-world problems given by the relevant stakeholders. By analyzing epistemic and ethical aspects of ambiguity, Bombaerts proposes a pedagogical approach that uses ambiguity to mitigate the risks of new technologies while developing students' problem-solving skills.

One of the social breakthroughs that AI-enhanced education can contribute to is the increase in inclusion and better representation of diversity and minorities. In the article "Pluralism, Disability and Inclusive Artificial Intelligence" (2025), Giovanni Galli analyzes the notions of disability and inclusion through a Wittgensteinian framework. Disability is understood as an open and pluralistic concept that derives its meaning(s) through its use in various contexts. Similarly, inclusion is a dynamic process whose full meaning is shaped by different language games and social practices. This pluralistic approach enables the development of context-sensitive AI systems that can recognize various forms of disability and adapt to users' diverse living experiences, leading to greater inclusion. When employed in education, these tools should not only passively avoid preexisting biases but also actively contribute to shaping a fairer society.

Educational practices have already changed to equip students with the skills required by new technologies. However, further adaptations will be necessary. Michael Quinn and William Gopal, in their article "Engaging with Film to Teach Students About Virtuous Outsourcing of Cognitive Tasks to LLMs" (2025), reflect on the Use-First Critical Thinking approach, which encourages students to engage with LLMs while critically evaluating them. The authors criticize this method for its lack of guidelines on when and for what purposes LLMs should be used, which may result in cognitive deskilling. Instead, they propose an alternative approach, inspired by the film *Memento*, where students analyze examples of inappropriate cognitive offloading. The idea is that epistemic agents should first learn to recognize when it is reasonable to delegate intellectual labor to LLMs, rather than utilizing these tools without such insight.

The risk of cognitive deskilling is a pressing issue in contemporary education, as uncritical reliance on generative AI may hinder the development of problem-solving abilities. In the article "When to Drive and When to Walk: LLMs, Higher Education, and Understanding" (2025), Jacob Rump argues that to prevent cognitive deskilling, education should promote the skill of understanding, as it is non-propositional and goes beyond pure linguistic competencies that LLMs imitate. While LLMs can be reliably used for tasks focused on mastering content, they are far less suitable for tasks that demand genuine understanding. To foster intellectual virtues and good learning practices, educational systems should be neither wholly technophilic (which Rump compares to always driving and never walking) nor technophobic (compared to always walking and never driving). By balancing the use of AI with traditional methods, educators can avoid the pitfalls of both technophilia and technophobia.

The growing exposure to AI technologies is transforming the way epistemic and moral agents view themselves, their relationships with others, and their position

in society. In the paper "'Where is Agency Moving to?': Exploring the Interplay between AI Technologies in Education and Human Agency" (2025), Ana Mouta, Ana María Pinto-Llorente, and Eva María Torrecilla-Sánchez present conclusions of their qualitative research on how teachers evaluate the influence of AI systems on the subjective, intersubjective, and collective human agency. At the subjective level, the main concerns were about the effects of AI on students' individuation and decision-making. At the intersubjective level, participants focused on the importance of role models that cannot be replaced by AI technologies. At the collective level, they viewed AI primarily as a short-term solution for the organization of educational facilities, while emphasizing the importance of human cooperation in the long run. Finally, they observed that AI systems affected authorship, diversity, and moral development at all three levels of human agency.

Whether students can benefit from advanced digital solutions in education is, in part, determined by their competencies to use these tools adequately and responsibly. Hans-Olav Hodøl, Per Ivar Kjærgård, and Øyvind Økland, in the article "Artificial intelligence and new digital divides among students in higher education" (2025), present their findings from the study on the opinions of the Norwegian higher education employees on new digital divides. In this context, digital divides are understood as disparities both between student groups and across academic disciplines. The participants anticipated that students who are already epistemically resourceful, analytically skilled, and independent would benefit the most from AI tools, raising concerns that such technologies may reinforce existing advantages. On a more positive note, the majority argued that these tools could, in general, provide a certain level of support for students.

Taken together, the contributions in this topical collection underscore that AI in education is neither an unqualified promise nor an unmanageable threat. It constitutes a rapidly evolving domain that demands continuous ethical, epistemic, and pedagogical reflection. These studies highlight that the future of education depends not on *whether* we adopt AI, but on *how* we engage with it. As AI continues its rapid evolution and integration into educational ecosystems, the philosophical discourse surrounding its ethical and epistemic dimensions necessarily intensifies. While definitive conclusions regarding AI's ultimate educational impact remain elusive, the contributions assembled in this collection provide important theoretical and empirical foundations for ongoing scholarly engagement.

This collection deepens scholarly understanding of AI in education and fosters a broader community of researchers, educators, and philosophers committed to exploring these important questions. They collectively argue that the future of education depends on engaging with AI critically, ethically, and with a focus on humanistic principles. We anticipate the frameworks and findings presented here will encourage further research and inform policy as educational institutions adapt to these changes. The interdisciplinary dialogue spanning education, technology studies, but with the focus on philosophy, establishes essential foundations for responsible use of AI in future educational paradigms.

Data Availability Not applicable.

Declarations

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Austin, T., Rawal, B. S., Diehl, A., & Cosme, J. (2023). AI for equity: Unpacking potential human bias in decision making in higher education. *AI Computer Science and Robotics Technology*, *13*. https://doi.org/10.5772/acrt.20

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *Int J Artif Intell Educ*, *32*, 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Boateng, O., & Boateng, B. (2025). Algorithmic bias in educational systems: Examining the impact of AI-driven decision making in modern education. *World Journal of Advanced Research and Reviews*, *25*(1), 2012–2017. https://doi.org/10.30574/wjarr.2025.25.1.0253

Bombaerts, G. (2025). Harnessing ambiguity in Challenge-Based learning as a practice of deep engagement with large Language models. *Digital Society*, *4*, 80. https://doi.org/10.1007/s44206-025-00233-3

Chinta, S. V., Wang, Z., Yin, Z., Hoang, N., Gonzalez, M., Quy, T. L., & Zhang, W. (2024). FairAIED: Navigating fairness, bias, and ethics in educational AI applications. *ArXiv Preprint arXiv:2407 18745*. https://doi.org/10.48550/arXiv.2407.18745

Fütterer, T., Goldberg, P., Bühler, B., Sikimić, V., Trautwein, U., Gerjets, P., & Kasneci, E. (2025). Artificial intelligence in classroom management: A systematic review on educational purposes, technical implementations, and ethical considerations. *Computers and Education: Artificial Intelligence*, 100483. https://doi.org/10.1016/j.caeai.2025.100483

Galli, G. (2025). Pluralism, disability and inclusive artificial intelligence. *Digital Society*, *4*, 68. https://doi.org/10.1007/s44206-025-00231-5

García Peñalvo, F. J., Alier, M., Pereira, J., & Casany, M. J. (2024). Safe, Transparent, and ethical artificial intelligence: Keys to quality sustainable education (SDG4). *IJERI: International Journal of Educational Research and Innovation*, (22), 1–21. https://doi.org/10.46661/ijeri.11036

Gopal, W., & Quinn, M. (2025). Engaging with film to teach students about virtuous outsourcing of cognitive tasks to LLMs. *Digital Society*, *4*, 79. https://doi.org/10.1007/s44206-025-00234-2

Gupta, M., & Kaul, S. (2024). AI in inclusive education: A systematic review of opportunities and challenges in the Indian context. *MIER Journal of Educational Studies Trends and Practices*, 429–461. https://doi.org/10.52634/mier/2024/v14/i2/2702

Hodøl, H. O., Kjærgård, P. I., & Økland, Ø. (2025). From access to ability: Educator's perspectives on AI and new digital divides in higher education. *Digital Society*, *4*, 78. https://doi.org/10.1007/s44206-025-00237-z

Julien, G. (2024). How artificial intelligence (AI) impacts inclusive education. *Educational Research and Reviews*, *19*(6), 95–103. https://doi.org/10.5897/ERR2024.4404

Khan, S. (2023). The ethical imperative: Addressing bias and discrimination in AI-driven education. *Social Sciences Spectrum*, *2*(1), 89–96.

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., & Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, *55*(9), 1–46. https://doi.org/10.1145/3555803

Milano, S., McGrane, J. A., & Leonelli, S. (2023). Large Language models challenge the future of higher education. *Nature Machine Intelligence*, *5*(4), 333–334. https://doi.org/10.1038/s42256-023-00644-2

Mohammed, P. S., & 'Nell'Watson, E. (2019). Towards inclusive education in the age of artificial intelligence: Perspectives, challenges, and opportunities. *Artificial Intelligence and Inclusive Education: Speculative futures and emerging practices*, 17–37. https://doi.org/10.1007/978-981-13-8161-4_2

Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2025). Where is agency moving to? Exploring the interplay between AI technologies in education and human agency. *Digital Society*, *4*, 49. https://doi.org/10.1007/s44206-025-00203-9

Nally, D. (2025). AI-Informed pedagogy for a Post-Truth era. *Digital Society*, *4*, 76. https://doi.org/10.1007/s44206-025-00230-6

Pagliara, S. M., Bonavolonta, G., Pia, M., Falchi, S., Zurru, A. L., Fenu, G., & Mura, A. (2024). The integration of artificial intelligence in inclusive education: A scoping review. *Information*, *15*(12), 774. https://doi.org/10.3390/info15120774

Pritchard, D., English, A., & Ravenscroft, J. (2021). Extended cognition, assistive technology and education. *Synthese*, *199*, 8355–8377. https://doi.org/10.1007/s11229-021-03166-9

Rump, J. (2025). AI, Education, and understanding: when to drive and when to walk. *Digital Society*, *4*, 77. https://doi.org/10.1007/s44206-025-00236-0

Salas-Pilco, S. Z., Xiao, K., & Oshima, J. (2022). Artificial intelligence and new technologies in inclusive education for minority students: A systematic review. *Sustainability*, *14*(20), 13572. https://doi.org/10.3390/su142013572

Sikimić, V. (2023). Epistemic inclusion as the key to benefiting from cognitive diversity in science. *Social Epistemology*, *37*(6), 753–765. https://doi.org/10.1080/02691728.2023.2258831

Suchithra, V., & Arya, C. (2025). The study on ethics and biases in AI-Powered education. *European Journal of Contemporary Education and E-Learning*, *3*(2), 37–43. https://doi.org/10.59324/ejceel.2025.3(2).04

Viberg, O., Kizilcec, R. F., Wise, A. F., Jivet, I., & Nixon, N. (2024). Advancing equity and inclusion in educational practices with AI-powered educational decision support systems (AI-EDSS). *British Journal of Educational Technology*, *55*(5), 1974–1981. https://doi.org/10.1111/bjet.13507

Vincent-Lancrin, S. (2020). R. (Ed.), Trustworthy artificial intelligence (AI) in education: Promises and challenges. *OECD Education Working Papers* 218 01–17 https://doi.org/10.1787/a6c90fa9-en.

Vučković, A., & Sikimić, V. (2025). Global justice and the use of AI in education: Ethical and epistemic aspects. *AI & Society*, *40*, 3087–3104. https://doi.org/10.1007/s00146-024-02076-x

Xia, Q., Chiu, T. K., Lee, M., Sanusi, I. T., Dai, Y., & Chai, C. S. (2022). A self-determination theory (SDT) design approach for inclusive and diverse artificial intelligence (AI) education. *Computers & Education*, *189*, 104582. https://doi.org/10.1016/j.compedu.2022.104582