

## **Review of Wayne Holmes and Kaśka Porayska-Pomsta (Eds.). (2022). The ethics of artificial intelligence in education: practices, challenges and debates**

Item Type	Other
Authors	Scott, Howard
Citation	Scott, H. (2024) Review of Wayne Holmes and Kaśka Porayska-Pomsta (Eds.). (2022). The ethics of artificial intelligence in education: practices, challenges and debates. <i>Postdigital Science and Education</i> , 6, pp. 705–710. <a href="https://doi.org/10.1007/s42438-023-00439-z">https://doi.org/10.1007/s42438-023-00439-z</a>
DOI	<a href="https://doi.org/10.1007/s42438-023-00439-z">10.1007/s42438-023-00439-z</a>
Publisher	Springer Nature
Journal	<i>Postdigital Science and Education</i>
Download date	2025-04-28 12:13:56
Link to Item	<a href="http://hdl.handle.net/2436/625332">http://hdl.handle.net/2436/625332</a>

**Review of Wayne Holmes and Kaśka Porayska-Pomsta (Eds.). (2022). *The Ethics of Artificial Intelligence in Education: Practices, Challenges and Debates*. Abingdon, UK and New York, NY: Routledge. 312 pp. ISBN 9780429329067 (E-Book)**  
Howard Scott, University of Wolverhampton, UK, [Howard.scott@wlv.ac.uk](mailto:Howard.scott@wlv.ac.uk)

**Keywords:** Artificial Intelligence, Ethics, Algorithms, Learning Analytics, Intersectionality, Postdigital

### **Introduction**

*The Ethics of Artificial Intelligence in Education: Practices, Challenges and Debates* (Holmes and Kaśka Porayska-Pomsta 2022) explores the socio-technical systems of Artificial Intelligence in Education (AIED). The book takes in subjects such as automation, exploitation, algorithmic unfairness. From the Introduction, readers are teased with the descriptor of Artificial Intelligence (AI) as a power tool for working with datasets, pattern modelling, and decision-making (Falk 2020); the kind of grandiose hype which this book can help readers navigate.

In the Conclusion, the editors draw out some questions from the book, asking: ‘What should be the role of AIED research, if any, in establishing best AIED practices and related ethics auditing processes? How should the AIED research community position its contribution with respect to the growing appetite of a money-making industry?’ (Porayska-Pomsta and Holmes 2022: 275)

### **Risks and Futures**

In Part I, Treviranus (2022) cautions around the direction of travel where education becomes shaped by AI. The author defends traditional practices, arguing that the machine will not replicate ‘the unexpected, uncertain, non-formulaic, non-linear work that requires continuous human judgement and creativity’ (37), which would include creativity, risk, discovery, learning from mistakes, precarious indigenous knowledge, and so on. However, it seems to this reader that none of those are off limits to large language models predicated on machine learning. If it is written, it seems that it can be mimicked and subsumed, however poorly.

Bartoletti (2022) highlights the dichotomy that has emerged between privacy and security, to illustrate the need regulation of AI technology. The question that isn’t raised is whether it is the place of Government to regulate what tools will evolve, for investing them with that power can have potential downsides too. Bartoletti (2022) then raises a series of key risks among the opportunities inherent to AI.

The first risk is the scaling up of poor pedagogical ideas through AIE. For instance, the promise of increased personalised learning through the use of learning analytics is tempered by the lack of empirical evidence. Another risk is ‘Automatic Nudging’, used to modify or persuade behaviours. These risks include the lack of frameworks applied by companies seeking to exploit the crossover between neuroscience and technology, though it is not explicitly stated what those frameworks might comprise.

A further risk is the ‘erosion of human agency’ (Bartoletti 2022: 82), which is a common trope in AI speculation. The view is presented that AI learning systems are already codified with outcomes so that models have been written with set categories, reference interests, and points of view for ideal future behaviour (much, we might say, as teachers design learning outcomes). This means that such systems potentially influence users to conform to in-built expectations, which are broadly political, sociological, and ethical by nature. ‘Ultimately, what this translates into is that the patterns used to shape our future choices are based on historical data.’ (Bartoletti 2022: 82) The question is then raised: ‘What does this do to human agency,

mobility, and creativity?’ (Bartoletti 2022: 82). Presumably human agency is left to monitor AI; a point raised in subsequent chapters.

### **Human Monitoring and Community Responses**

Smuha (2022) observes that current ethical frameworks are insufficient to deal with issues posed by AI and shows how in 2019, some unilateral guidelines were drawn up for ‘trustworthy’ AI by the European Commission. Those guidelines seemingly put responsibility of AIED applications onto stakeholders, including the institutions that deploy them. End-users (typically students or teachers) and all other parties affected by AIED (from parents to society at large) should be informed about these requirements and be able to request and enforce accountability (Smuha 2022: 118). As I have noted elsewhere (Scott 2023), these recommendations point to a human governance.

Smuha, a researcher in law and criminology, includes the endorsement that regulators should ensure that AIED developers implement these requirements to safeguard users. Herein is an issue for this reviewer. Globally, AI development has become something of an arms race. In many instances it is developed by companies in countries where state regulation is less enforced. This prompts questions of how we can demand timely policy before the AI advances, as well as who gets to demand ethical compliance.

Smuha’s argument mirrors the recent call in a misguided draft report by Hamilton, Wiliam, and Hattie (2023), that there need be a global regulatory framework quickly established. This is a practical call, but, to my mind, one that may only be perceived as disadvantageous to corporations wishing to steal a march on competitors. The call also seems unfair. Why should regulation, determined by western commentators, be expected to realistically extend globally?

Part II begins with Porayska-Pomsta and Holmes (2022) moving the debate to the AIED community. This is drawn from transdisciplinary members, with the editors remarking that researchers working in an ‘uncomfortable scientific space at the edge of knowledge’ must avoid descriptive accounts of what may go wrong in AI ethics. Rather, AIED ‘demands actionable approaches to addressing the existing and potential implications of AI being used in and for education’ (Porayska-Pomsta and Holmes 2022: 147).

This part of the book then covers what we may consider pressing and real concerns. How AIED systems increase pre-existing social inequalities, issues with algorithmic (un)fairness, and structural inequities encoded into research? Who current AIE systems serve? What should be learnt and how it should be learnt?

Holstein and Doroudi (2022) point to the disparities of access to technologies between groups. This divide extends beyond hardware to AIED interfaces that may only use a mainstream dialect and other inequalities that can be embedded through data that reflects historical inequalities. Their chapter also shows how the lack of demographic representation in developers of the communities they wish to serve may also amplify inequity. Similarly, Kizilcec and Lee (2022) critique notions of algorithmic fairness in systems that employ predictive models. They give examples of Doroudi and Brunskill’s (2019) study, which tested intelligent tutoring systems effects on fast and slow learners, asking if it discriminates against either group.

I see problems when researchers explore education systems that seek to commodify everything into measurable units. Highlighting the diagnosis of students as ‘fast’ or ‘slow’ raises questions about the ethics of profiling: ‘Fast or slow’ at what aspects of learning? When? In which subject and how are learning outcomes negotiated or determined? If researchers persist with this kind of study, they really end up researching, and potentially validating, mechanisms that are flawed. Kizilcec and Lee (2022: 174) similarly ask: ‘What characteristics

of algorithmic systems in education are associated with greater benefits? And what counts as a beneficial impact on these stakeholders?'

### **Entrenched Inequities**

Kizilcec and Lee (2022) proceed to explain how algorithms may discriminate based on pre-existing patterns of bias, rather than conscious choice by programmers. Data that is collected for specific variables is used to determine success across the board and algorithms generate predictions accordingly. As the authors observe, this is problematic because things like 'success' are predicated on what are widely regarded as reductive models of education (see Jandrić and Hayes 2020); characteristics which may bear little resemblance to what students and educators often perceive as qualities of their learning. This is like an inversion of Goodhart's Law (1975); when measures become targets, they cease to be good measures. We risk becoming defined, even governed, by algorithms.

Kizilcec and Lee (2022) round off with a review comparing statistical notions of fairness, before proposing that Open Source tools such as Aequitas and AI Fairness 360 can be trained to monitor algorithms and mitigate against unfairness or bias. Their final recommendation is for developers to provide better explanation of the purpose and use of their systems to stakeholders to increase the warrant of trust in these systems.

This is a chapter that presents a breakdown of the opportunities with the potential for algorithm prediction to improve education (and beyond, as they show with AI applied to decisions regarding credit score applications, medical expenditures, etc.). But it shows that these must be handled with care and treated with tenets of fairness, equity, and transparency.

Adopting a critical theory lens, Madaio, Blodgett, Mayfield, and Dixon-Román (2022) build on Kizilcec and Lee (2022). Taking a more holistic view of systematic inequity and unfairness, they explore the limitations of research that purports to measure the fairness of algorithms. Their argument is for justice-oriented approaches to research, calling for 'design justice, counter hegemony and technology refusal ... for radical visions for the design (or dismantling) of educational AI' (Madaio et al. 2022: 204).

Madaio et al. (2022) argue that continual technological development is not always the answer to algorithmic unfairness. Alternatively, they suggest that some systems might be completely rebuilt and ask whether some are even necessary. They show that technologies are developed in social contexts, which give rise to problems they are developed to solve. This then impacts on the kinds of data available and its evaluation metrics. The authors outline their case compellingly, questioning notions that have become thoughtlessly embedded and commonplace (such as what constitutes 'ability', or whether some goals are worthwhile or desired).

While I suggest that the call for dismantling is simplistic, postdigital discourse must realistically accept that technological determinism and Big Tech are synergistic forces that shape facets of society, for better or worse. To this end, the chapter (and book, generally) is effective in its refusal to treat AI as novelty (Cormier et al. 2023). It supports growing debates by continually pointing to AIs as inseparable from historical inequities and subject to transdisciplinary interrogation (Fawns et al. 2023).

### **Conclusion**

*The Ethics of Artificial Intelligence in Education: Practices, Challenges and Debates* (Holmes and Kaška Porayska-Pomsta 2022) is written principally by North American and UK authors. The book's limited geographical diversity is unfortunate, since so much of the debate, particularly regarding platforms such as ChatGPT (Rowe 2023), is about the marginalised, disenfranchised, or geopolitical dominance through western innovation. However, it is a great

credit that the book's focus is largely on concrete harms in the here and now, rather than abstract speculation.

As AI shapes a fast-moving world, policy and regulation can be slow in response to its implications and new ethical concerns. It may seem odd that a book is the technology that can capture something so dynamic as AI, but here is one constructed of thoughtful analysis that can assist with reading the ethics surrounding AI. As response to era (and future) defining discourse, the editors call for a pause to 'consider what we want and need respectively from policy, from the EdTech, and from the AIED research community' (Porayska-Pomsta and Holmes 2022: 275).

The book symbolises manifestly that pause. It contributes to a growing discourse, which helps to hold processes and the progress of AI to account. The book reminds us that while the questions, critical angles, and ethical frameworks we need to take may be new, they are situated in old inequities and hegemonic dominance. Therefore, it will be beneficial for anyone wishing to confront and understand Artificial Intelligence at an introductory level. Grounded in well-informed critical discourse that will stimulate thinking, the book will also be of critical importance to researchers looking at key angles to take in looking at a broad field that will necessarily entail new ethical questions.

## References

- Bartoletti, I. (2022). AI in education: An opportunity riddled with challenges. In W. Holmes & K. Porayska-Pomsta (Eds.), *The ethics of artificial intelligence in education: Practices, Challenges and Debates* (pp. 74-90). Abingdon, UK and New York, NY: Routledge.
- Cormier, D., Jandrić, P., Childs, M., Hall, R., White, D., Phipps, L., Truelove, I., Hayes, S., & Fawns, T. (2019). Ten Years of the Postdigital in the 52group: Reflections and Developments 2009–2019. *Postdigital Science and Education*, 1(2), 475–506. <https://doi.org/10.1007/s42438-019-00049-8>.
- Doroudi, S., & Brunskill, E. (2019). Fairer but not fair enough on the equitability of knowledge tracing. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 335-339). New York: Association for Computing Machinery. <https://doi.org/10.1145/3303772.3303838>.
- Falk, K. (2020). Artificial intelligence: HR's new power tool. Ebn, 23 April. <https://www.benefitnews.com/opinion/artificial-intelligence-hrs-new-power-tool>. Accessed 27 September 2023
- Fawns, T., Ross, J., Carbonel, H., Noteboom, J., Finnegan-Dehn, S., & Raver, M. (2023). Mapping and Tracing the Postdigital: Approaches and Parameters of Postdigital Research. *Postdigital Science and Education*, 5(3), 623-642. <https://doi.org/10.1007/s42438-023-00391-y>.
- Goodhart, C. (1975). Problems of monetary management: the UK experience in papers in monetary economics. *Monetary Economics*, 1.
- Hamilton, A., Wiliam, D., & Hattie, J. (2023). The Future of AI in Education: 13 things we can do to minimize the damage. EdArXiv, 13 August. <https://doi.org/10.35542/osf.io/372vr>.
- Holmes, W., & Porayska-Pomsta, K. (Eds.). (2022). *The Ethics of Artificial Intelligence in Education: Practices, Challenges and Debates*. Abingdon, UK and New York, NY: Routledge. <https://doi.org/10.4324/9780429329067>.
- Holstein, K., & Doroudi, S. (2022). Equity and artificial intelligence in education. In W. Holmes & K. Porayska-Pomsta (Eds.), *The ethics of artificial intelligence in education: Practices, Challenges and Debates* (pp. 151-173). Abingdon, UK and New York, NY: Routledge.
- Jandrić, P., & Hayes, S. (2020). Postdigital We-Learn. *Studies in Philosophy of Education*, 39(3), 285-297. <https://doi.org/10.1007/s11217-020-09711-2>.

- Kizilcec, R. F., & Lee, H. (2022). Algorithmic fairness in education. In W. Holmes & K. Porayska-Pomsta (Eds.), *The ethics of artificial intelligence in education: Practices, Challenges and Debates* (pp. 174-202). Abingdon, UK and New York, NY: Routledge.
- Madaio, M., Blodgett, S. L., Mayfield, E., & Dixon-Román, E. (2022). Beyond “fairness”: Structural (in) justice lenses on ai for education. In W. Holmes & K. Porayska-Pomsta (Eds.), *The ethics of artificial intelligence in education: Practices, Challenges and Debates* (pp. 203-239). Abingdon, UK and New York, NY: Routledge.
- Porayska-Pomsta, K., & Holmes, W. (2022). Conclusions: Toward ethical AI/ED. In W. Holmes & K. Porayska-Pomsta (Eds.), *The ethics of artificial intelligence in education: Practices, Challenges and Debates* (pp. 271-281). Abingdon, UK and New York, NY: Routledge.
- Rowe, N. (2023). ‘It’s destroyed me completely’: Kenyan moderators decry toll of training of AI models. *The Guardian*, 2 August. <https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai>. Accessed 5 October 2023.
- Scott, H. (2023). ‘Reject All’: Data, Drift and Digital Vigilance. In S. Hayes, M. Jopling, S. Connor, & M. Johnson (Eds.), *Human Data Interaction, Disadvantage and Skills in the Community: Enabling Cross-Sector Environments for Postdigital Inclusion* (pp. 285-298). Cham: Springer. [https://doi.org/10.1007/978-3-031-31875-7\\_15](https://doi.org/10.1007/978-3-031-31875-7_15).
- Smuha, N. A. (2022). 5 Pitfalls and pathways for Trustworthy Artificial Intelligence in education. In W. Holmes & K. Porayska-Pomsta (Eds.), *The ethics of artificial intelligence in education: Practices, Challenges and Debates* (pp. 113 – 147). Abingdon, UK and New York, NY: Routledge.
- Treviranus, J. (2022). 1 Learning to learn differently. In W. Holmes & K. Porayska-Pomsta (Eds.), *The ethics of artificial intelligence in education: Practices, Challenges and Debates* (pp. 25 – 47). Abingdon, UK and New York, NY: Routledge.