Received: 30 August 2023 | Revised: 27 September 2023 | Accepted: 3 October 2023

DOI: 10.1002/tea.21913

COMMENT

JRST | WILEY

Bias, bias everywhere: A response to Li et al. and Zhai and Nehm

Christina Krist (Stina)[1] | **Marcus Kubsch**[2]

[1]Department of Curriculum and Instruction, University of Illinois Urbana-Champaign, Champaign, Illinois, USA

[2]Department of Physics, Freie Universität Berlin, Berlin, Germany

Correspondence

Christina Krist, Department of Curriculum and Instruction, University of Illinois Urbana-Champaign, Champaign, IL, USA.
Email: ckrist@illinois.edu

Marcus Kubsch, Department of Physics, Freie Universität Berlin, Berlin, Germany.
Email: m.kubsch@fu-berlin.de

1 | INTRODUCTION

In response to Li et al.'s (2023) and Zhai and Nehm's (2023) commentaries on Zhai et al.'s 2022 paper, *Applying Machine Learning to Automatically Assess Scientific Models*, we offer the perspective that these commentaries are talking past each other around several key issues related to artificial intelligence (AI) in science education assessment. In part, this "talking past" stems from the fact that each set of authors is approaching the conversation from a distinct perspective: Li et al. address AI through a sociopolitical lens, while Zhai and Nehm address it from a technical lens. These perspectives are not explicitly recognized by either set of authors; and as a result, while they use common terminology, there is a mismatch of (unarticulated) definitions between these two commentaries. Specifically for this commentary, we will focus on the conflation of multiple definitions of *bias*, which we also find to be a common conflation across the field.

We ultimately view this mismatch as a missed opportunity and a barrier to generative ethical conversations about the role of AI in education. We emphasize here how and why *both* perspectives are valuable, and argue that they are most valuable when in critical but productive conversation with each other.

2 | UNPACKING *BIAS* IN AI

To explain where we see Li et al. and Zhai and Nehm talk past each other around *bias*, we unpack the term by first considering the lifecycle of AI in science education. We conceptualize

Christina Krist and Marcus Kubsch contributed equally to this manuscript.

this lifecycle in three phases (Figure 1). Typically, the *inception point* of the AI lifecycle is the definition of a task that the AI is supposed to solve. Next, during *development*, a computer model is set up and model parameters are trained based on data to enable the AI to solve the task. Finally, when the model solves the task sufficiently well (as judged by whoever defined the task), the AI is *deployed*.

Figure 1 depicts the ways that multiple forms of bias can creep in at each stage in the AI lifecycle. During task definition, sources of bias can include stakeholder standpoint biases, that is, whose voices are being heard and listened to in characterizing the task (Harper & Kayumova, 2023). During the development of an AI, sources of bias can be distinguished into bias in the data (e.g., under- and overrepresentation of certain groups due to sampling bias; Baker, 2019), and bias in the actual model learning (e.g., modeling choices that conflate individual and aggregate measures; O'Neil, 2016). Lastly, the deployment of AI may cause bias via feedback loops, such as the high number of incarcerated Black men in the United States leading to implementation outcomes that result in decisions that further increase this number (Benjamin, 2019a; Crawford, 2021), effectively perpetuating systemic inequalities.

Considering these different forms of bias reveals a range of phenomena with different roots. Building on Mitchell et al. (2021), it is helpful to distinguish between *modeling and data bias* and *societal bias*. Modeling and data bias refers to measurement errors or problematic modeling choices. In contrast, societal bias reflects historical and social injustices in the data and the potential perpetuation of these injustices as a consequence of deployment choices. Importantly, an AI system can have no modeling and data bias—making perfect predictions about final grades—and still cause societal bias through the way in which it is deployed, for example, when it is used to reallocate resources away from an already underserved population because of lower grades (Benjamin, 2019a; Crawford, 2021). Similarly, an AI system can demonstrate little societal bias, that is, imperfect outcomes causing little to no harm (e.g., showing hints to students who do not need them) and still exhibit profound model and data bias by producing bad predictions. In real life, the bias exhibited by an AI system will probably include both types. Productive conversations weighing the benefits and risks of a specific AI tool require a differentiated understanding of how bias arises.

When Li et al. bring up "concerns about equity in the use of AI," they draw on work that emphasizes the societal aspects of bias, such as highlighting racism in AI (Cheuk, 2021) and that privilege is encoded in coding rubrics that underlie AI systems (Noble et al., 2012).

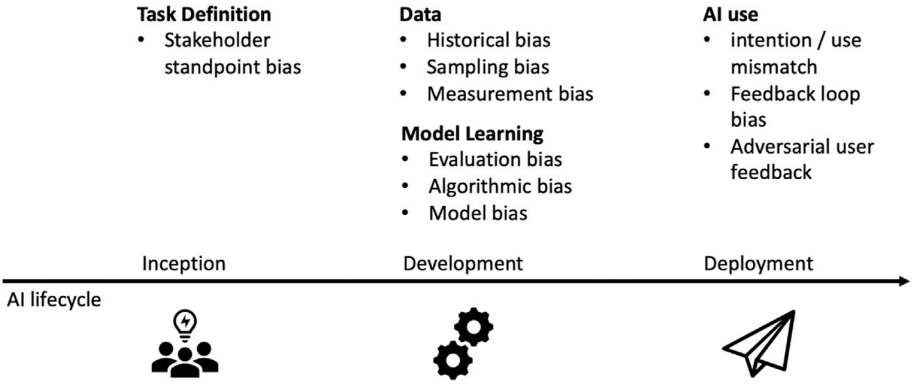

FIGURE 1 Bias in the artificial intelligence (AI) lifecycle. Adapted from Baker and Hawn (2022).

In contrast, in their response, Zhai and Nehm emphasize modeling and data bias, discussing the "unbalanced data problem" and citing research that found little to no statistical bias in AI systems (Ha & Nehm, 2016). Zhai and Nehm's emphasis on modeling and data bias is probably most prevalent in their heading "AI SCORING BIAS ≠ INJUSTICE IN EDUCATION." While this statement is certainly true, it also ignores how societal bias in AI can contribute to and perpetuate injustice in education.

3 | CONCLUSION

When conversations talk past each other by focusing on different aspects of bias in AI, there is a missed opportunity for productive discussion. Given the huge potential that AI offers the field (as Zhai and Nehm rightly argue) and the similarly huge risks that come with it (as Li et al. rightly call attention to), we—as a field and community of researchers—should not miss these opportunities. In fact, it is our professional responsibility to *not* miss them. Practically, this means that research teams conducting work on AI in science education need members with a deep technical knowledge base that allows for epistemic transparency (Héder, 2020) of algorithms and what they are doing. A key part of this knowledge is the ability to translate between the mathematical representations of the world in the models and algorithms used and the substantive understanding of the world. For example, clustering techniques often rely on some kind of distance measure. Making a decision for and implementing a distance measure that is appropriate in the research context requires an understanding of the phenomenon at hand in combination with technical knowledge about distance measures.

At the same time, research teams also need members with a deep knowledge of, and lenses for viewing, how power and oppression are at play in the socio-technical-epistemic mangle of knowledge production and generation (Lizárraga, 2023; Tanksley, 2022). This need not necessarily be a different person than someone with deep technical expertise; understanding how datasets were collected from a socio-technical systems perspective requires recognizing how methods of data collection impact the nature and quality of the data—including social biases that might be reflected in the dataset or what might be missing (D'Ignazio & Klein, 2023). Additionally, recognizing how algorithmic processes might amplify and perpetuate these biases (e.g., Benjamin, 2019b; Noble, 2018) and critically considering the implications for the learning contexts of interest requires taking a broader view than simply attempting to look inside the "black box" of how an algorithm is functioning.

When these knowledges and perspectives can interface with one another, the question about AI moves away from either a simple binary "should we?" or a surrender to the inevitable "train" and instead opens space for messy questions about what data are actually telling us, who might benefit, who might be harmed, and in what ways. We provide one simple example of how these technical and critical conversations might come together in the context of a research project to develop a teacher dashboard that shows students' progress in learning about energy. First, developing the dashboard required technical expertise to ensure that the deep learning architecture used for assessing students' progress based on their answers in a digital textbook actually provided scores that meaningfully captured students' understanding of energy (Gombert et al., 2022). Secondly, because we know that the training data came from schools with students from predominantly privileged backgrounds and most students were learning in their native language, we are very careful with making recommendations for the use of the system in other demographics. The risk is that the linguistic variation associated with different demographics, including use of everyday forms of sensemaking (e.g., Rosebery et al., 2010) and broad linguistic repertories such as onomatopoeia (Suárez & Otero, 2023), may unintentionally obfuscate sophisticated but non-normatively voiced understandings of energy, and in doing so, perpetuate the same challenges that teachers have in recognizing and valuing students' diverse sensemaking repertoires (Rosebery et al., 2016). We are currently undertaking efforts to study whether this is the case and if so, how it can be accommodated, before recommending use of the system beyond the original study context.

In sum, just because "the train has left the station does not mean the engineering of it should carry on full steam ahead, without regular check-ins, questions, and challenges from its interactions with the people and lands it crosses. With every stage of progress comes loss and harm, we hope that within science education at least, we can approach that progress with both creativity and critical care for the human side of AI.

ORCID

Christina Krist https://orcid.org/0000-0002-9738-4308
Marcus Kubsch https://orcid.org/0000-0001-5497-8336

REFERENCES

Baker, R. S. (2019). Challenges for the Future of Educational Data Mining: The Baker Learning Analytics Prizes. https://doi.org/10.5281/ZENODO.3554745

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, *32*(4), 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Benjamin, R. (2019a). *Race after technology: Abolitionist tools for the new Jim code*. Polity.

Benjamin, R. (2019b). Assessing risk, automating racism. *Science*, *366*(6464), 421–422.

Cheuk, T. (2021). Can AI be racist? Color-evasiveness in the application of machine learning to science assessments. *Science Education*, *105*, 825–836. https://doi.org/10.1002/sce.21671

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

D'Ignazio, C., & Klein, L. F. (2023). *Data feminism*. MIT Press.

Gombert, S., Di Mitri, D., Karademir, O., Kubsch, M., Kolbe, H., Tautz, S., Grimm, A., Bohm, I., Neumann, K., & Drachsler, H. (2022). Coding energy knowledge in constructed responses with explainable NLP models. *Journal of Computer Assisted Learning*, *37*, 767–786. https://doi.org/10.1111/jcal.12767

Ha, M., & Nehm, R. H. (2016). The impact of misspelled words on automated computer scoring: A case study of scientific explanations. *Journal of Science Education and Technology*, *25*(3), 358–374. https://doi.org/10.1007/s10956-015-9598-9

Harper, A., & Kayumova, S. (2023). Invisible multilingual Black and Brown girls: Raciolinguistic narratives of identity in science education. *Journal of Research in Science Teaching*, *60*(5), 1092–1124. https://doi.org/10.1002/tea.21826

Héder, M. (2020). The epistemic opacity of autonomous systems and the ethical consequences. *AI & Society.*, *38*, 1819–1827. https://doi.org/10.1007/s00146-020-01024-9

Li, T., Reigh, E., He, P., & Adah Miller, E. (2023). Can we and should we use artificial intelligence for formative assessment in science? *Journal of Research in Science Teaching*, *60*(6), 1385–1389. https://doi.org/10.1002/tea.21867

Lizárraga, J. R. (2023). Cyborg sociopolitical reconfigurations: Designing for speculative fabulation in learning. *Journal of the Learning Sciences*, *32*(1), 21–44. https://doi.org/10.1080/10508406.2022.2154159

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, *8*(1), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

Noble, S. U. (2018). *Algorithms of oppression*. New York University Press.

Noble, T., Suarez, C., Rosebery, A., O'Connor, M. C., Warren, B., & Hudicourt-Barnes, J. (2012). “I never thought of it as freezing”: How students answer questions on large-scale science tests and what they know about science. *Journal of Research in Science Teaching*, *49*(6), 778–803. https://doi.org/10.1002/tea.21026

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (1st ed.). Crown.

Rosebery, A. S., Ogonowski, M., DiSchino, M., & Warren, B. (2010). “The coat traps all your body heat”: Heterogeneity as fundamental to learning. *The Journal of the Learning Sciences*, *19*(3), 322–357.

Rosebery, A. S., Warren, B., & Tucker-Raymond, E. (2016). Developing interpretive power in science teaching. *Journal of Research in Science Teaching*, *53*(10), 1571–1600.

Suárez, E., & Otero, V. (2023). Ting, tang, tong: Emergent bilingual students investigating and constructing evidence-based explanations about sound production. *Journal of Research in Science Teaching*. https://doi.org/10.1002/tea.21868

Tanksley, T. (2022). Race, education and #BlackLivesMatter: How online transformational resistance shapes the offline experiences of Black college-age women. *Urban Education*, 004208592210929. https://doi.org/10.1177/00420859221092970

Zhai, X., & Nehm, R. H. (2023). AI and formative assessment: The train has left the station. *Journal of Research in Science Teaching*, *60*(6), 1390–1398. https://doi.org/10.1002/tea.21885

How to cite this article: Krist, C., & Kubsch, M. (2023). Bias, bias everywhere: A response to Li et al. and Zhai and Nehm. *Journal of Research in Science Teaching*, *60*(10), 2395–2399. https://doi.org/10.1002/tea.21913