Supporting AI literacy in K-12 Science Education: Raising Critical Consciousness towards Ethical AI

Selin Akgün[1], Hee Rin Lee[1], Kahyun Choi[2], Joseph Krajcik[1]
Michigan State University[1]
Indiana University Bloomington[2]

Subject/Problem

Artificial Intelligence (AI) has brought drastic change to various fields of science, technology, engineering, and education (Druga et al., 2018; Su & Yang, 2022). Personalized or intelligent tutoring systems, automated assessment systems, and predictive analytics keep changing the equation and modes of learning and teaching by increasingly being deployed in various educational contexts (Zhai et al., 2020). These AI applications have their affordances in science and STEM education by giving students individualized, timely, and detailed feedback; reducing teachers' workload, and supporting students to find future career paths (Zhai et al., 2020).

Although AI holds considerable promise in science education context, scholarly and public discussion raises critical questions towards ethics and the related power issues of AI systems (Benjamin, 2019; Holstein & Doroudi, 2021). The power imbalances amplify ethical and societal challenges surrounding AI use in terms of privacy, autonomy, surveillance, bias, and discrimination (Lee et al., 2021). Therefore, critical discussion on how, what, and for whom these AI systems are designed are vital to consider in education (Akgun & Greenhow, 2021; Vakil et al., 2020). For example, AI systems potentially perpetuate existing systemic gender and racial bias by amplifying unfair learning experiences for students from mostly disadvantaged and marginalized groups (mostly within personalized learning and automated assessment systems). These concerns stem from how AI learns from existing data sets built upon a variety of sources without necessarily accounting for any inherent biases in them (Broussard, 2018; Crawford, 2021).

Considering the criticality of these challenges, researchers, teacher educators and practitioners of science and STEM education start engaging in critical debates on how to teach and address the affordances, as well as ethical and societal challenges of AI (e.g., Boddington, 2017; Winfield & Jirotka, 2018). The growing body of research in STEM education focuses on how to support K-12 teachers to a) teach students social, cultural, and ethical implications of AI (e.g., Han, 2021; Wilson & Roscoe, 2020) and b) how to support students' critical consciousness about the algorithmic bias in designing future and ethical AI systems (e.g., Bogina et al., 2022; Holmes et. al., 2022). However, the research in science education needs to reflect empirically more on a) what pedagogical and curricular content, tools and platforms can be used to support science teachers (and their students) to identify and critique ethical implications of AI; b) how the implementation of these resources would look like (within its challenges and affordances) in formal and informal science learning environments (Ali et al., 2019; Crompton et al., 2022). Therefore, this study aims to delve into a) what pedagogical approaches and modules can be used to teach about core concepts and macro-ethical issues of AI in K-12 STEM education, and (b) how the implementation of these modules might support students' critical AI literacy. The study will

contribute to the growing body of AI and science education research by articulating future directions and recommendations for teacher educators and practitioners of science education.

Procedure

This study reflects on a smaller part of a two-year exploratory project that started in August 2021. The project uses an asset-based approach on how to teach the core concepts, ethical aspects, and community component of AI to elementary students. The project provides a series of AI education modules for a variety of community stakeholders, such as public librarians, teacher educators, and STEM educators to run their own AI literacy programs. The project team developed an AI literacy program and tested them in three public libraries across the Midwest, East, and West regions of the US. Each literacy program was conducted through 4 sessions, over 2 weekend periods with elementary students from 4th to 6th grades. The literacy program contains two teaching modules (see Table 1) which are developed upon the existing quality curriculum documents on AI literacy in K-12 settings (e.g., AI4ALL.org, 2022; MIT RAISE, 2022) as well as the authors' previous work on AI and participatory design (e.g., Authors, 2019).

Module	Day	Themes	Learning Goals	Main Activity / Investigation
1	1 and 2 (1.5 hours each)	AI classification and generation (within bias and misinformation)	- Figuring out the core concepts of AI (e.g., classification, GANs) - Developing AI classification systems and AI- generated images - Identifying and critiquing AI bias through ethical/societal issues of AI classification and misinformation	a. Classification and bias through Google's Teachable Machine: Cat and dog classifiers b. Creating AI-generated images using DALL-E2
2	3 and 4 (1.5 hours each)	AI in local communities and workplaces	- Identifying benefits and challenges of AI for science and society - Designing future ethical AI systems that are relevant and meaningful to community needs	a. Reading and discussing articles about AI chatbots: ChatGPT b. Envisioning and designing ideal future workplace with AI: Reflecting back to community

Table 1. The content of the modules of the workshop series

We created the modules within a pedagogy built upon critical race theory that views students from socially underserved communities not as people with deficits but as people with "community cultural wealth" (Ladson-Billings, 1992; Yosso, 2005). Therefore, the modules consist of three key themes. While the first two themes (classification and generation) pertain to the foundational knowledge of AI, the third theme relates to the integration of these concepts to envision future AI systems. The specific focus is given into applying knowledge to benefit communities and develop critical thinking, rather than just conveying information about AI.

Our study investigates how the workshop series (modules) help students to engage with and figure out the core knowledge and social/ethical issues of AI. Due to the limited space, the current paper focuses on one of the workshops conducted in the Midwest, US. During the 4 sessions, we worked with 6 elementary students (5th and 6th graders). One of the PIs and research assistants of the project introduced students to the two modules that aim to support students to 1)

have access to foundational knowledge and concepts about AI, and 2) envision and play an active role in designing ethical and relevant AI technologies for their own local communities through series of activities. We used multiple tools in the workshops including a) Google's Teachable Machine to figure out classification and bias of AI (Day 1); b) DALL-E2 to generate images with AI (Day 2); c) ChatGPT to generate text with AI (Day 2); d) news articles on AI to apply the core knowledge to the daily issues (Day 3), and e) drawing tools to generate AI design ideas for their local communities (Day 4). Using participatory design methodology, we aimed for students to utilize their own assets to co-design AI technologies based on the needs of their own community. Using co-design methods, we help students to envision and design the ideal AI systems from their own perspectives. Over the sessions, students as active learners (or active actors) provide us feedback to make their learning experiences more meaningful and relevant for them. We collect observation field notes from the four workshop sessions, conduct semi-structured individual interviews with each student, and collect students' models and artifacts based on their AI-related discussions and designs (Vanover et al., 2021).

Analyses and Findings

The modules include 4 main investigations that are implemented over four sessions. These investigations are adapted and modified from various quality materials considering the project's learning goals for each workshop session. In this study, we primarily report on students' investigations and work on the foundational knowledge and bias of AI. Throughout each session, students engage with a specific driving question to figure out the core AI concepts and to design AI- based artifacts.

During module 1, students are introduced with the driving question (DQ) of "What is AI and how do we use AI in our daily lives?". As they brainstorm ideas and share their experiences with AI, they start generating their own DQs that they are curious to learn about (see Fig.1). Following, students engage with examples of recent AI systems and explore the meaning of classification as one of the two main AI features. They analyze photos and videos of classification (such as detection algorithms) to figure out how an AI system can classify things. As the main investigation of the session, students work with Google's Teachable Machine to explore and design their own AI system, as well as to develop a basis of how these systems are created by using classification and lead us to transform our biases into algorithms we design. During the investigation, students work in groups of three to make their own cat and dog classifiers. Each group worked with the provided datasets by selecting various pictures of different types of cats and dogs. More specifically, they design their training data set as they select a number of different cat and dog images based on their color, posture, and types. Then, to test their dataset, they show these images to their model and record if their classifier predicts if the image is of a dog or a cat. Finally, they rearrange their dataset by sorting different images (even with different animals) to make their training dataset of cats and dogs larger and more diverse. In that way, students had first-hand experiences on how to train their own datasets, test if their AI classifies each image correctly (within the confidence score), and if their AI was better with dogs or cats and why they think so.

Figure 2 shows how one of the groups interpret the two core AI concepts—classification and generation. In their model, they represent their thinking about the two main functions of AI, the importance of data to train AI systems, and how AI classification may surface bias. During the

investigation, this group selected more cat pictures than dogs to train their dataset. Therefore, they have better confidence scores and diversity in their dataset to classify and predict different types, postures, and colors of the cats. When sharing their models with others, the group claimed that when they tested their AI system with the pictures of cats, dogs, and pigs, they got very low prediction scores with dogs and pigs and the machine identified pigs as a one type of a cat. Further, the group represented how AI classification systems can generate bias while predicting the images and how sufficient and appropriate data is the critical and the secret ingredient of the better working AI system.

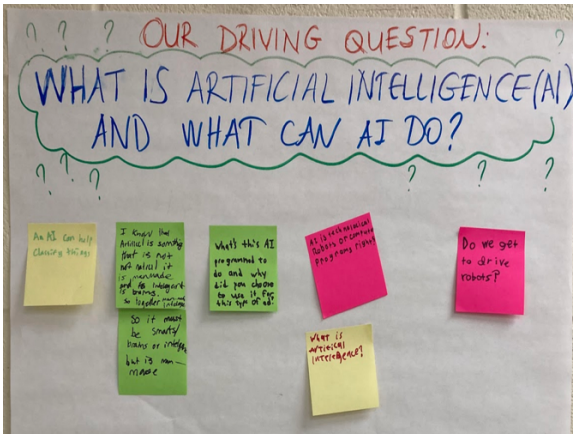

Fig. 1 Driving Question Board

Fig. 2 Group model to reflect on core concepts and bias of AI

At the end of the session, one of the students from the group shared her experiences about working with the cat and dog classifiers and her takeaways about the AI bias: *"I think that it was really cool to work with that cat and dog dataset. We had picked 30 pics of dogs but a couple pictures of a cat. The dog pictures were a lot, but then it was only some pictures of a cat, and the machine said all the cats look like these couple pictures we picked, but that's not true. I feel like the more dataset they have, the better they run... So maybe we should have a better dataset of humans. I feel like people have their own biases in their brain as they are growing up. For example, if you think about a professor, even when I think about them, I think of Albert Einstein as professor or a scientist."* Her insights suggest that implementation of the investigation (classifiers in this case) may help students to figure out a) the more sufficient and appropriate data (with the consideration of multiple variables) in training datasets can lead to a better and more ethical AI systems, and b) datasets can carry and perpetuate the stereotypical biases of the society. Engaging with these critical ideas in the first module helped students to design their own AI systems at the end of the last session by addressing the concerns of surveillance, privacy, autonomy, and bias.

During module 2, students engage with how AI can be used in different fields of science and technology, such as in healthcare systems. Students critically reflect on a) the importance of teamwork and collaboration, b) the asymmetric power dynamics among team members (such as those between doctors and nurses), and c) problems in teamwork in healthcare and how AI can be a potential solution to these problems. Engaging with these problems and their AI-oriented solutions for various workspaces, students identify the ways in which to actively co-design and generate novel AI designs considering their own future workplace which are relevant for their communities. During the investigation, the groups' future AI designs show how students start

developing awareness of power structures (such as those between doctors and nurses) leads to a focus on supporting individuals with less power.

Conclusions and Contributions

This study delves into potential pedagogical approaches and modules (within their implementation) that educators can use to teach about core concepts and ethical issues of AI in K-12 STEM education. Considering the research on how to best support science educators in incorporating and teaching AI-related content (bridging its connections with science) is still in its infancy, we articulate future directions and recommendations for teacher educators and practitioners of science education. More specifically, this study brings insight to support researchers and teacher educators of science education by a) providing a quality pedagogical and instructional module on AI literacy and ethics and b) how implementing these modules (resources) and what affordances and challenges they might bring to educate students to develop more critical, fair, and equitable AI systems in the future. In other words, within these two modules and their implementation, we hope that K-12 students not only learn core AI concepts, but also get more actively engaged with students' local civic, ethical, and societal issues as AI co-designers.

The workshop series also encourages students to work collaboratively, think critically on activities that force them to wrestle with issues of algorithmic bias and discrimination built into AI either intentionally or not intentionally. In addition, these modules can be connected or integrated into certain science lessons since they align with several ELA and NGSS standards to support students in reading informational texts, designing AI systems, and reading about relevant local and global science and technology issues by leveraging civic participation. The modules can be also used in the context of science teacher education. Practitioners can reap the benefit of these resources as they bring community-based and environmental impacts of AI into their science and technology lessons. In that way, future citizens can have a chance to incorporate their science learning and localized environmental science content with the issues of access and developing AI.

General Interest

We argue that this study is timely, relevant, and essential for NARST members (researchers, teacher educators, practitioners) to see what quality instructional resources are available for the intersection of science, AI, and ethics in K-12 settings. This work is also essential and complementary with its empirical aspect. It investigates the affordances and challenges of implementing these resources in K-12 settings. Therefore, the study might help the science education community to explore ways in which increasing the awareness of AI literacy and algorithmic bias by creating more fair and equitable science education content in K-12 settings.

References

Akgun, S., & Greenhow, C. (2021). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 1-10.

AI4ALL. (2023). https://ai-4-all.org/open-learning/resources/

Ali, S., Payne, B. H., Williams, R., Park, H. W., & Breazeal, C. (2019, June). Constructionism,

ethics, and creativity: Developing primary and middle school artificial intelligence education. In *International workshop on education in artificial intelligence k-12 (eduai'19)* (pp. 1-4).
Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim code.
Boddington, P. (2017). *Towards a code of ethics for artificial intelligence* (pp. 27-37). Cham: Springer.
Bogina, V., Hartman, A., Kuflik, T., & Shulner-Tal, A. (2022). Educating software and AI stakeholders about algorithmic fairness, accountability, transparency, and ethics. *International Journal of Artificial Intelligence in Education*, *32*(3), 808-833.
Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT
Crawford, K. (2021). *Atlas of AI*. Yale University Press.
Crompton, H., Jones, M. V., & Burke, D. (2022). Affordances and challenges of artificial intelligence in K-12 education: A systematic review. Journal of Research on Technology in Education, 1-21.
Druga, S., Williams, R., Park, H. W., & Breazeal, C. (2018, June). How smart are the smart toys? Children and parents' agent interaction and intelligence attribution. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (pp. 231-240).
Han, X. (2021, December). How Does AI Engage in Education? A Quantitative Research on AI Curriculum and Instruction in Public Primary Schools. In *2021 4th International Conference on Education Technology Management* (pp. 15-19).
Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Boston: Center for Curriculum Redesign.
Holstein, K., & Doroudi, S. (2021). Equity and Artificial Intelligence in Education: Will "AIEd" Amplify or Alleviate Inequities in Education?
Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal,* 32 (3): 465–491.
Lee, H. R., Taylor, A., Kubota, A., & Riek, L. D. (2019). Coordinating clinical teams: Using robots to empower nurses to stop the line. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-30.
MIT RAISE. (2023). DAILy curriculum. https://raise.mit.edu/daily/index.html
Su, J., & Yang, W. (2022). Artificial intelligence in early childhood education: A scoping review. *Computers and Education: Artificial Intelligence*, *3*, 100049.
Vakil, S., Marshall, J., & Ibrahimovic, S. (2020). "That's Bogus as Hell!": Getting Under the Hood of Surveillance Technologies in an Out of School STEM Learning Environment.
Vanover, C., Mihas, P., & Saldaña, J. (Eds.). (2021). *Analyzing and interpreting qualitative research: After the interview*. Sage Publications.
Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, *58*(1), 87-125.
Winfield, A. F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180085.
Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race ethnicity and education*, *8*(1), 69-91.
Zhai, X., C Haudek, K., Shi, L., H Nehm, R., & Urban-Lurain, M. (2020). From substitution to redefinition: A framework of machine learning-based science assessment. *Journal of Research in Science Teaching*, *57*(9), 1430-1459.