Black Feminist Thought as a Guide for Ethical Integration of Artificial Intelligence in Mathematics Classroom

Eunhye Flavin and Matthew T. Flavin (Georgia Institute of Technology)

Introduction

The rapid advancement of artificial intelligence (AI) and the advent of generative AI tools sound a clarion call for the ethical integration of AI in the mathematics classroom (National Council of Teachers of Mathematics [NCTM], 2024). The capabilities of AI tools come with risks, including the hallucination effect, lack of transparency, and inherent bias in their algorithms (Buolamwini, 2023). U.S. mathematics classrooms have been exposed to a long history of systemic racism and patriarchy, and mathematics teacher educators bear responsibility for understanding these oppressive harms in the context of these new tools (Davis & Jett, 2019; Leyva, 2021). To achieve this goal, we believe discussions should go beyond focusing solely on the inherent limitations of AI itself. Rather, it requires explicitly identifying the types of power at play. Drawing upon Black feminist scholarship (Collins & Bilge, 2016), in this study, we provide discussion-ready questions that mathematics teacher educators can utilize to prevent foreseeable harms.

Theoretical Framework: The Matrix of Domination

Patricia Hill Collins and Sirma Bilge's book, *Intersectionality* (2016), specifies the four types of power dynamics that operate in society, which may arise in the integration of AI into mathematics classroom. The *interpersonal* domain relates to the microcosm of social interactions between humans and the AI agents. The *disciplinary* domain examines how rules are applied differently to people. Within the *cultural* domain, attention is directed towards the fabrication of ideas and messages. The *structural* domain entails the creation and perpetuation of institutions and organizations that systemically favor certain individuals over others. By adapting the framework from Williams (2024) framework, we created two guiding questions to analyze the power in each domain more concretely:

Question 1. How can we characterize the status quo of AI tool usage in engaging with typical power dynamics in this domain, taking into account social and historical factors?

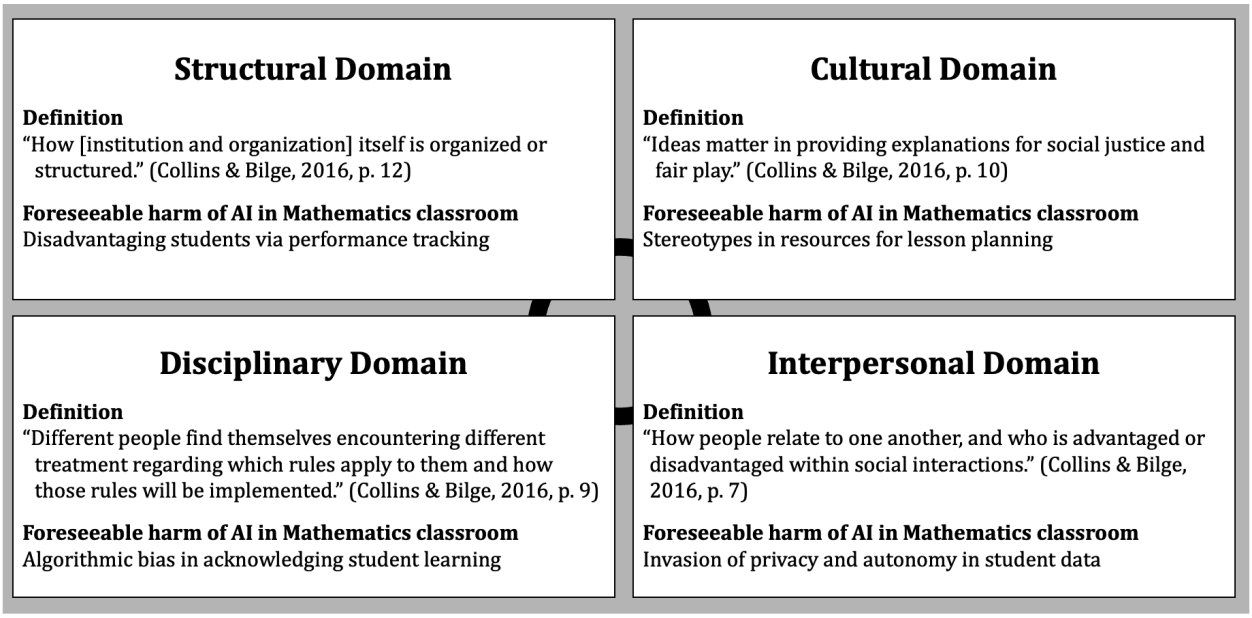
Question 2. How can we avoid harms through approaches that tackle the status quo?

Analysis of AI Power in Mathematics Classrooms

Figure 1 shows the domain of power, its definition, and related ethical concerns. We present one ethical concern for each domain. We also intentionally begin our illustration with structural power to emphasize that the power exerted by AI extends beyond human-tool interaction.

Figure 1

The Matrix of Domination Related to the Integration of AI Tools in the Mathematics Classroom



Structural Power: Disadvantaging Students via Performance Tracking

Question 1. How Can We Characterize the Status Quo of AI Tool Usage in Engaging with Typical Power Dynamics in This Domain, Taking into Account Social and Historical Factors?

School mathematics practices have structurally excluded black students through standardized test scores and tracking system (Spencer & Hand, 2015). Important applications of AI algorithms and machine learning models include the monitoring of student behaviors and the prediction of mathematics proficiency (Akgun & Greenhow, 2021). Personalized learning systems (intelligent tutoring systems) and automated assessment systems are prominent examples of AI applied in education (Akgun & Greenhow, 2021). Another example is the predictive analytics algorithm system, which was developed to predict student mathematics course completion (e.g., Gkontzis et al., 2022).

Despite the potential benefits of AI, harms stemming from algorithmic bias can potentially be introduced. The 2020 United Kingdom General Certificate of Secondary Education (GCSE) and Advanced Level qualification (A-levels) grading controversy is a notable example. During the Covid 19 pandemic, UK national qualification regulators assigned grades to students using algorithmic calculations instead of the usual exams (Coughlan, 2020). The algorithm awarded lower grades to students in state-funded, low-achieving schools than they did to students in independent schools (Smith, 2020). This approach may have structurally marginalized students who already faced

constraints in their educational settings, impacting their future opportunities for education and employment.

Question 2. How Can We Avoid Harms through Approaches That Tackle the Status Quo?

The UK GCSE and A-levels example shows how seemingly neutral measures like algorithms and gaps in grades reflect educational injustices in society. Also, it suggests how macro-level adoption of AI algorithms (e.g., national, state-level, school districts) bears a high-stakes influence on the futures of students. To address the impact of AI in a structural domain, a comprehensive approach is needed. As Baker and Hawn suggest (2022), this entails the creation of standard, education-specific guidelines for the conduction of bias audits and the calculation of bias metrics.

In addition, the use of AI should reflect the voices of communities impacted by it throughout the entire process of its development and implementation. Algorithms applied for evaluation of mathematics grades and future prospects should solicit input from all stakeholders, including students, parents, and teachers. Furthermore, sufficient information should be readily accessible to these stakeholders in an interpretable form. Overall, mathematics teachers need to push the boundaries beyond simple utilization of AI to a competent understanding of algorithmic interventions and relevant ethics (NCTM, 2024).

Cultural power: Stereotypes in resources for lesson planning

Question 1. How Can We Characterize the Status Quo of AI Tool Usage in Engaging with Typical Power Dynamics in This Domain, Taking into Account Social and Historical Factors?

UNESCO (2023) identified the role of ChatGPT as a co-designer of lessons alongside a teacher. Despite the potentials of Generative artificial intelligence, such as ChatGPT and DALLE 2, these tools are entrenched with stereotypes and biases (Sun et al., 2024). Mathematics lesson plans created by these tools are therefore vulnerable to such harms when mathematics teachers fail to critically unpack the meanings of generated media. AI models have also been known to censor words related to minority identities (e.g., queer) that uplift their representation (Agnew, 2023). Thus, even when the teachers ask the AI model to devise a lesson that uplifts minority representations, the generated outcomes can be inherently limited by this censorship. This is concerning, especially when mathematics teachers are already found to harbor implicit biases against the mathematical abilities of girls and students of color (Corpur-Gencturk et al., 2023). Without careful attention to the biases and stereotypes held by both AI models and teachers, mathematics lesson planning with AI models may suffer from these limitations.

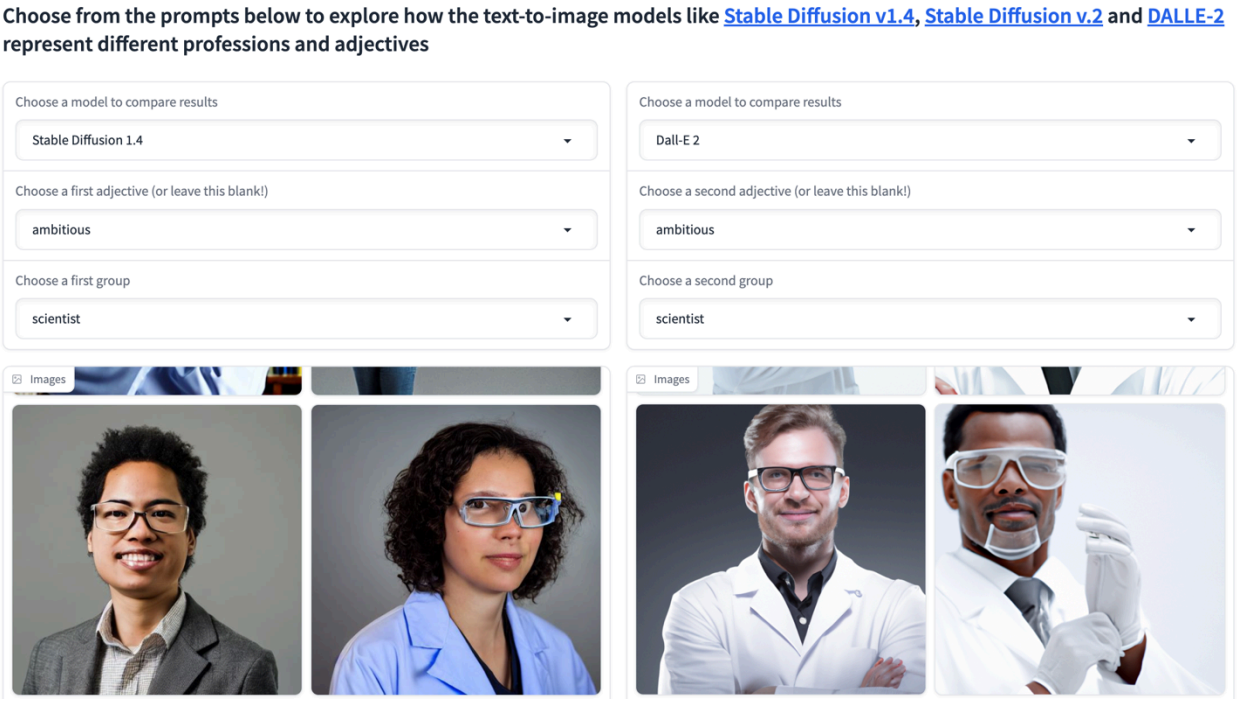
Question 2. How Can We Avoid Harms through Approaches That Tackle the Status Quo?

StableBias enables users to link textual references of identity characteristics with corresponding visual elements in their outputs (see Figure 2). Mathematics teachers can utilize this website to compare images to be used in their lesson. When using other AI models, they can also keep in mind a judicial choice of images considering implicit bias of One way to address this problem is to utilize a system that compares social representations from different AI models. StableBias, for example, leverages three AI models (i.e., Stable Diffusion v.1.4, Stable Diffusion v. 2, and DALLE-2) to quantify social biases in text-to-image systems to lower the risk of discriminatory outcomes (Luccioni et al., 202the AI models.

Mathematics teacher education programs can also offer implicit bias trainings to enact culturally responsive practices in harnessing AI tools. For example, mathematics teacher candidates can be taught how to generate a prompt for AI chatbots (e.g., ChatGPT, Gemini) that creates an activity localized to the cultural backgrounds and experiences of the students. Eldridge (n.d.) demonstrates an example where teachers create a task related to a historically significant event, tailored for a particular grade level and academic subject. They instruct to include the following warning in the prompt to mitigate bias in the generated output: “Do not use cultural stereotypes in your answer and if you are not sure about a student’s culture, do not create a response for that culture.” Mathematics teacher educators can have teacher candidates assess the outputs to look for a lack of cultural understanding to facilitate the critical, ethical use of the AI models.

Figure 2.

Images Generated by StableBias When Asked to Portray an Ambitious Scientist (Created on July 7th, 2024)



Disciplinary power: Algorithmic bias in acknowledging student learning

Question 1. How Can We Characterize the Status Quo of AI Tool Usage in Engaging with Typical Power Dynamics in This Domain, Taking into Account Social and Historical Factors?

The mathematics classroom establishes a set of moral and social norms. The issue arises in whose norms we adhere to and the differential treatments arising under these norms. In her book, *Unmasking AI*, Joy Buolamwini identifies inherent bias in AI algorithms stemming from oversampling white males in training data and questions the norm in AI tools.

Educational AIs that can recognize facial expressions or/and behaviors have been increasingly applied for supporting student learning and teaching practices (Akgun & Greenhow, 2021; Foster et al., 2024). An exciting example is the application of computer vision to support the development of mathematics teacher noticing. Foster (2024) introduces a research project where his team creates deep neural networks for computer vision to classify the activities in the videos of elementary mathematics and English language arts classrooms. The team also developed a teacher-facing analytic dashboard where teachers can examine their instructional efforts captured in the videos. To create these neural networks, the video annotators label learning moments in the videos and develop algorithms to classify which learning moments are mathematically meaningful. This value-ladened work can incorporate further analysis on whose learning behaviors is likely determined as valuable or marginalized.

Question 2. How Can We Avoid Harms through Approaches That Tackle the Status Quo?

Bias in facial analysis algorithms raises concern that AI tools reflect white gaze and, similarly, male gaze, as the default ways of seeing the world. This finding, reported in Buolamwini (2023), reveals that this bias arises in data collection and labelling along with AI model building and implementation. As the educational AI increasingly gives attention to facial expressions and behaviors (e.g., Foster, 2024), mathematics educators who are involved in the development of those AI tools can reduce model bias by being transparent about data used for AI models (Jindal, 2023). In addition, obtaining external validations by independent sources before the implementation of AI models can be another approach to reduce algorithmic bias (Jindal, 2023). Mathematics teacher educators can also organize interdisciplinary trainings for gender and racial bias and hegemonic norms encoded in AI technologies.

Interpersonal Power: Invasion of privacy and autonomy in student data

Question 1. How Can We Characterize the Status Quo of AI Tool Usage in Engaging with Typical Power Dynamics in This Domain, Taking into Account Social and Historical Factors?

AI tools are designed to easily access and gather user data. They can influence and persuade user behavior. Educational AI tools are utilized for extremely personalized assessment and learning through vast data collection (Popenici, 2023). This data can be used to invade privacy and lessen autonomy of users by manipulating their behaviors. Historically marginalized groups are at higher stakes when exploitation of personal data was used against them (New America, no date).

Question 2. How Can We Avoid Harms through Approaches That Tackle the Status Quo?

There are various AI platforms that teachers can utilize, including *Lessonplans.ai* and *MagicSchool.ai* for lesson planning, *Curipod in AI* for instruction, and *Kahoot* and *FormativeAI* for generating quizzes. At a minimum, mathematics teachers need to ensure that personalized student data is not being shared without their consent when utilizing AI platforms. The ethical concerns of AI should be taken seriously, and we believe that Black feminist scholarship can facilitate discussions of the role of mathematics teachers.

References

Agnew, W. (2023). *AI Ethics and Critique for Robotics* [Doctoral dissertation, University of Washington].

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, *2*(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7

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

Buolamwini, J. (2023). *Unmasking AI: My mission to protect what is human in a world of machines.* Random House.

Collins, P. H. & Bilge, S. (2016). *Intersectionality*. Polity.

Copur-Gencturk, Y., Thacker, I., & Cimpian, J. R. (2023). Teachers' race and gender biases and the moderating effects of their beliefs and dispositions. *International Journal of STEM Education*, *10*(31), 1–25. https://doi.org/10.1186/s40594-023-00420-z

Coughlan, S. (2020, August 14). Why did the A-level algorithm say no? *BBC News*. https://www.bbc.co.uk/news/education-53787203

Davis, J., & Jett, C. (Eds.). (2019). *Critical race theory in mathematics education*. Routledge.

Eldridge, B. (n.d.). *Create culturally contextualized activities using an AI chatbot.* AI for Education. https://www.aiforeducation.io/prompts/culturally-contextualized-activities

Foster, J. (2024). Bridging AI and mathematics teacher education: A teacher educator's journey. *Connections, Summer 2024*. https://amte.net/sites/amte.net/files/Connections%28Foster%29.pdf

Gkontzis, A. F., Kotsiantis, S., Panagiotakopoulos, C. T., & Verykios, V. S. (2022). A predictive analytics framework as a countermeasure for attrition of students. *Interactive Learning Environments*, *30*(6), 1028–1043. https://doi.org/10.1080/10494820.2019.1709209

Jindal, A. (2023). Misguided artificial intelligence: How racial bias is built into clinical models. *Brown Hospital Medicine*, *2*(1). https://doi.org/10.56305/001c.38021

Leyva, L. A. (2021). Black women's counter-stories of resilience and within-group tensions in the white, patriarchal space of mathematics education. *Journal for Research in Mathematics Education, 52*(2), 117–151. https://doi.org/10.5951/jresematheduc-2020-0027

Luccioni, A. S., Akiki, C., Mitchell, M., & Jernite, Y. (2023). Stable bias: Analyzing societal representations in diffusion models. *arXiv preprint arXiv:2303.11408*.

National Council of Teachers of Mathematics [NCTM] (2024). *Artificial intelligence and mathematics teaching*. https://www.nctm.org/standards-and-positions/Position-Statements/Artificial-Intelligence-and-Mathematics-Teaching/

New America (n.d.). *For marginalized communities, the stakes are high*. https://www.newamerica.org/oti/reports/centering-civil-rights-privacy-debate/for-marginalized-communities-the-stakes-are-high/

Popenici, S. (2023). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning and Teaching*, *6*(2). 1–7. https://doi.org/10.37074/jalt.2023.6.2.4

Spencer, J., & Hand, V. (2015). The racialization of mathematics education. In L. D. Drakeford (Eds.), *The race controversy in American education* (pp. 237–258). Praeger.

Smith, H. (2020). Algorithmic bias: should students pay the price?. *AI & Society, 35*, 1077–1078. https://doi.org/10.1007/s00146-020-01054-3

Sun, L., Wei, M., Sun, Y., Suh, Y. J., Shen, L., & Yang, S. (2024). Smiling women pitching down: auditing representational and presentational gender biases in image-generative AI. *Journal of Computer-Mediated Communication*, *29*(1), zmad045. https://doi.org/10.1093/jcmc/zmad045

United Nations Educational, Scientific, and Cultural Organization [UNESCO] (2023). ChatGPT and artificial intelligence in higher education. https://www.iesalc.unesco.org/wp-content/uploads/2023/04/ChatGPT-and-Artificial-Intelligence-in-higher-education-Quick-Start-guide_EN_FINAL.pdf

Williams, T. (2024). *Understanding roboticists' power through matrix guided technology power analysis*. https://doi.org/ 10.1145/3610978.3640766