Vol. 24 (2026): Proceedings of the 2nd International Conference on Linguistics, Psychology and Creative Education (ICLPCE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/iclpce.2026231236

The Algorithmic Glass Ceiling: Gender Bias and STEM Identity in AI-Supported Secondary School Education

Jianjia Ye*

Miss Hall's School, Pittsfield, MA, United States

**Corresponding author: Jianjia Ye.*

Abstract

Gender inequality in STEM education concerns more than participation rates or achievement gaps. It also concerns recognition: who is treated as a legitimate knower, problem solver, and future practitioner. As AI-driven learning systems enter secondary school STEM classrooms, existing research may underestimate algorithmic mediation, the process through which educational algorithms translate social patterns into content, feedback, and learning trajectories. This paper argues that AI-supported STEM environments can reorganize gender bias through three connected mechanisms. Representation sets expectations about who belongs in STEM. Interaction shapes the kind of support, challenge, and recognition students receive. Personalization can turn early behavioral differences into lasting learning paths. Drawing on research in educational AI, learning analytics, and STEM identity, this paper shifts the focus from detecting bias to understanding how students experience it. The main claim is that AI systems can make structural inequality feel like an individual difference. That matters because the effects reach beyond grades. They shape whether girls see themselves as people who can question, create, and contribute to STEM knowledge. The paper contributes a framework for understanding algorithmic bias as a process that forms identity over time.

Keywords

algorithmic bias, STEM identity, AI in education, gender inequality, learning analytics

1. Introduction

Gender inequality in STEM is often explained through performance, but that explanation is too narrow. Performance alone does not explain why women remain underrepresented in STEM careers, especially in computing and engineering. The central issue is recognition: who is seen as someone who belongs in STEM? Whose engagement is read as competence, and whose is read as effort or compliance? Carlone and Johnson describe STEM identity as a process built through competence, performance, and recognition [1]. Repeated signals tell students who is allowed to know, create, and claim expertise. AI-driven learning tools now send many of those signals. Intelligent tutoring systems, adaptive learning platforms, automated feedback tools, and recommendation systems are increasingly used in secondary school STEM classrooms. This makes the issue urgent. A teacher's comment may happen once. An AI system can personalize content, hints, difficulty, and feedback across hundreds of interactions. Baker and Hawn show that algorithmic bias in educational systems is well-documented and connected to gender, race, and socioeconomic inequality [2]. When systems learn from historical patterns of achievement and engagement, they may carry those patterns forward unless schools and designers address them directly. This paper examines how AI-supported STEM systems turn existing gendered patterns into learning experiences. The argument follows a sequence. Representation sets expectations about who counts as a scientist. Interaction then shapes behavior through feedback, support, and challenge. Personalization stabilizes those differences by using early patterns to guide later learning paths. Together, these mechanisms form a feedback loop. Small differences can become durable inequalities, and the system can make them harder to see. To make this argument clearer, it is important to note that these mechanisms do not operate independently. They interact continuously in real classroom systems. Representation influences interpretation of behavior; interaction reinforces that interpretation; personalization then locks it into future instruction. The paper asks two questions. At the mechanism level, how do representation, interaction, and personalization create gendered learning conditions in AI-supported STEM classrooms? At the experience level, how do those conditions affect girls' sense of belonging, epistemic agency, and STEM identity? The central claim is that AI-driven learning environments can reinforce gender inequality by making unequal treatment appear to reflect individual ability.

2. Theoretical Background

2.1 Gender Stereotypes as Structural Patterns

Gender stereotypes in STEM do not require open discrimination. They often work through quiet assumptions about what competence looks like. The link between technical ability and masculinity is especially persistent. Bian et al. found that children begin to associate brilliance with males before age seven [3]. Gunderson et al. show that gendered beliefs about mathematics can shape girls' interest and performance throughout schooling [4]. By secondary school, many girls have already absorbed a difficult question: Is this field for people like me? These patterns matter for AI because educational systems are trained on data produced inside unequal environments. Data about who continues with STEM content, who asks questions, and who scores well may look like a record of ability. It also records social conditions. Bias can therefore enter a system before any student uses it. It is built into the data, the categories, and the assumptions that guide model design. It is also important that these stereotypes not only influence explicit judgments; they also influence behavioral expression. Students adjust participation, risk-taking, and confidence based on perceived expectations. These behavioral adjustments are exactly the signals AI systems later interpret as ability.

2.2 STEM Identity and Epistemic Agency

STEM identity develops through relationships and repeated experiences. Zhao et al. find that science self-efficacy and STEM identity influence each other over time among adolescent girls [5]. Girls who see themselves as capable are more likely to seek science opportunities. Those experiences can strengthen identity. When girls receive signals that weaken their sense of ability, they may step away from the same opportunities that could have supported growth. Epistemic agency names another part of this process. It refers to whether learners experience themselves as people who can question ideas, form explanations, and build arguments. Du et al. show that AI feedback can shape this experience [6]. Autonomy-supportive feedback invites exploration and acknowledges the learner's perspective. Directive feedback pushes the learner toward a specific answer. Over time, these differences can shape whether students develop an active or passive relationship with STEM knowledge. In AI systems, epistemic agency is not only psychological but also structural, because systems decide what kinds of responses are even possible.

2.3 Representation and Symbolic Annihilation

Tuchman's concept of symbolic annihilation describes how women can be made socially invisible through absence, trivialization, and limited representation [7]. AI-curated educational content can repeat this pattern through examples, images, problem scenarios, and role-model profiles. These choices create the STEM world students meet on screen: who solves hard problems, who leads research, and whose face stands for technical expertise. A large-scale analysis of AI-generated images found that text-to-image models reinforce traditional gender roles, with women more often shown in care-oriented contexts and men more often shown in technical or physically demanding roles [8]. This is not only a visual bias; it is also a cognitive framing effect. Students repeatedly exposed to these patterns may internalize assumptions about disciplinary belonging.

2.4 Differential Interaction and Algorithmic Paternalism

Algorithmic paternalism describes a pattern in which a system appears supportive while giving some students more control, structure, and lowered challenge. AI systems make constant micro-decisions: how much scaffolding to provide, how difficult the next problem should be, and whether feedback should invite reasoning or direct compliance. These decisions are meant to respond to ability. The problem is that the system reads ability through behavior, and behavior in STEM classrooms is shaped by gender socialization. Girls who have been socialized toward caution may ask for help more often, attempt fewer speculative answers, or seek confirmation before moving forward. A system may read these behaviors as lower confidence or lower ability. It may then respond with more support and less challenge. Du et al. found a related asymmetry in large language model feedback: male-associated cues generated more autonomy-supportive feedback, while female-associated cues generated more controlling feedback [6]. This creates a loop: behavior influences system response, and system response reinforces behavior.

2.5 Personalization as Trajectory

AI personalization adjusts instruction to a learner's pace, preferences, and performance. In education, this is often presented as a clear advantage. A student receives material that matches her current level. The concern is that the system adapts to behavior that may already reflect discouragement, stereotype threat, or unequal recognition. A girl who approaches mathematics cautiously may be guided toward easier tasks, more structure, and a slower pace, even when caution reflects past experience instead of ability. The result can resemble cumulative advantage. Merton's Matthew effect describes how small early differences grow over time [9]. Lietz et al. show how cumulative advantage can widen gender gaps in computer science careers [10]. In AI-driven learning, early differences in behavior can influence the path a student receives. Less challenging material may lead to lower performance, which then confirms the system's earlier judgment. This creates what can be understood as a self-reinforcing trajectory system.

3. Literature Review

3.1 The STEM Culture That AI Inherits

AI systems are trained on data generated within existing social and educational contexts. As a result, they inevitably inherit the structural conditions embedded in those environments. STEM education, in particular, has long been shaped by cultural associations that link technical competence and intellectual authority with masculinity, Western educational backgrounds, and forms of economic privilege. These associations are not always explicit, but they are deeply embedded in participation patterns, classroom interactions, and historical achievement data. Friedman and Nissenbaum describe this phenomenon as preexisting bias, referring to bias that exists in social structures prior to the development of any computational system [11]. In the context of AI-supported STEM education, this implies that training data already reflects unequal distributions of participation and recognition. Consequently, machine learning systems trained on such data may implicitly privilege behavioral patterns that were historically more visible or rewarded. For example, behaviors such as rapid responses, frequent participation, and confident articulation are often overrepresented in datasets because they have historically been associated with higher perceived competence. However, these behaviors are also socially shaped. Students who experience stereotype threat or lower expectations may adopt more cautious interaction styles, including hesitation, verification-seeking, or reduced participation frequency. When AI systems interpret these behavioral signals as direct indicators of ability, they risk converting socially produced differences into individualized assessments of competence.

3.2 Representation Bias: Defining Competence Through Absence

Representation in AI-supported learning systems extends beyond the presentation of information; it actively shapes epistemic norms regarding who is recognized as a legitimate producer of knowledge. This includes the selection of examples, contextual narratives in problem sets, visual materials, and recommended role models embedded in educational content. Women remain systematically underrepresented in domains culturally associated with intellectual brilliance and technical expertise [3]. Steinke demonstrates that media representations of scientists play a significant role in shaping adolescent girls' STEM identity formation, particularly through contextual cues that signal belonging or exclusion [12]. When translated into AI-generated or AI-curated educational content, these representational patterns may be reproduced at scale. If AI systems repeatedly present STEM authority figures as predominantly male, or if women appear primarily in supporting or exceptional roles, students may internalize an implicit boundary around disciplinary belonging. Even when female representation is present, it may still reinforce exclusion if it is framed as rare or exceptional. Over time, this can shape whether girls perceive STEM participation as a normative pathway or as an exceptional achievement requiring extraordinary effort.

3.3 Interaction Bias: How Differential Support Produces Differential Outcomes

Interaction in AI systems refers to the continuous stream of pedagogical decisions made through feedback generation, hint provision, task sequencing, and difficulty adjustment. Du et al. provide empirical evidence that AI systems may generate systematically different feedback depending on gender-coded cues embedded in user profiles [6]. In their benchmarking study, male-associated cues were more likely to receive autonomy-supportive responses, while female-associated cues tended to receive more directive feedback emphasizing correctness and compliance. Although each individual interaction may appear minor or pedagogically reasonable, its cumulative effect becomes significant over time. Students interacting with AI systems may encounter hundreds of feedback instances, each subtly shaping how they approach learning tasks. For female students, repeated exposure to directive feedback may encourage reliance on verification strategies rather than exploratory reasoning. This may reduce opportunities for intellectual risk-taking, which is a key component of STEM identity development and advanced problem-solving. Importantly, interaction bias does not require intentional discrimination. It can emerge from patterns embedded in training data or from design choices that optimize for short-term correctness rather than long-term epistemic development.

3.4 Opacity, Personalization, and the Matthew Effect

One of the central challenges in identifying and addressing algorithmic bias in education is system opacity. Students and educators often lack visibility into how decisions are made regarding task assignment, feedback type, or difficulty progression. As a result, learners may interpret system outputs as objective reflections of their ability rather than as contingent outcomes of algorithmic processes. Baker and Hawn emphasize that algorithmic predictions can influence learner motivation and self-perception even in the absence of explicit bias [2]. This psychological effect becomes particularly important when combined with well-documented educational phenomena such as stereotype threat. Bedynska et al. show that chronic stereotype threat is associated with reduced mathematical achievement through mechanisms such as working memory depletion and intellectual helplessness [13]. When AI systems respond to hesitation or reduced performance by lowering task difficulty or increasing scaffolding, they may unintentionally reinforce this cycle. Although such adaptation is often framed as supportive personalization, it may also contribute to a Matthew effect dynamic, where initial disadvantages accumulate over time [9]. In this way, personalization can shift from being an adaptive feature to a structural mechanism that stabilizes early behavioral differences into long-term learning trajectories.

4. Discussion and Synthesis

AI-supported STEM systems do not simply reproduce gender bias; rather, they actively reorganize it through their operational logic. Each individual interaction, task selection, feedback generation, scaffolding intensity, and difficulty adjustment, may appear neutral when considered in isolation. However, when aggregated across time, these micro-decisions form a structured and directional pattern that can shape learning trajectories in systematic ways. In this sense, bias is not only embedded in system outputs, but also distributed across sequences of instructional decisions that accumulate over repeated engagement. Within this process, representation and interaction function as mutually reinforcing mechanisms. Representation shapes expectations about who is perceived as belonging in STEM disciplines, including implicit assumptions about competence, authority, and intellectual legitimacy. Interaction then operationalizes these expectations through pedagogical responses, including the type of feedback provided, the level of autonomy allowed, and the degree of cognitive challenge assigned. Over time, these two layers co-produce each other: expectations influence interactional design, while interactional experiences reinforce initial representational assumptions, producing a recursive system of reinforcement. Although technical debiasing strategies are necessary for mitigating algorithmic unfairness, they remain insufficient on their own. Baker and Hawn outline existing fairness approaches in educational AI systems [2], particularly those focused on outcome-level parity. However, such approaches often do not address a deeper issue: how systems interpret and operationalize behavioral signals that are themselves shaped by unequal social, cultural, and educational conditions. As a result, even "fair" outputs may still emerge from structurally biased input interpretations. Moreover, AI systems often present themselves as objective or data-driven, which can amplify their perceived neutrality. This perceived objectivity may lead students to interpret system-generated feedback as accurate reflections of their ability rather than as contingent outputs of algorithmic design. In such cases, students may be less likely to question, contest, or reinterpret feedback signals, even when those signals reflect structural bias rather than actual competence. Ultimately, AI systems shape not only academic performance but also learners' perceived horizons of possibility. Learning trajectories remain technically flexible, yet they become structurally constrained through repeated interactions. As a result, some students experience expanding opportunities for exploration and advancement, while others encounter progressively narrowing pathways that subtly limit their epistemic and disciplinary development.

5. Conclusion

This paper argues that AI-driven STEM learning environments can reinforce and obscure gender bias through three connected mechanisms: representation, interaction, and personalization. These mechanisms form a feedback system in which structural inequality becomes an individualized learning experience. The contribution of this paper is to frame algorithmic bias as a process of STEM identity formation. When algorithmic systems are opaque, students may internalize system outputs as evidence of ability rather than as products of design and history. Fairness in AI education therefore requires attention not only to outputs, but also to data, interaction structures, and recognition processes embedded in learning environments. Future work should investigate these mechanisms empirically in classroom contexts over time.

References

[1] Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of women of color: Science identity as an analytic lens. Journal of Research in Science Teaching, 44, 1187–1218. https://doi.org/10.1002/tea.20237

[2] 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

[3] Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. Science, 355(6323), 389–391. https://doi.org/10.1126/science.aah6524

[4] Gunderson, E. A., Ramírez, G., Beilock, S. L., & Levine, S. C. (2012). The role of parents and teachers in the development of gender-related math attitudes. Sex Roles, 66(3–4), 153–166.

[5] Zhao, M., Ozturk, E., Law, F., Joy, A., Deutsch, A. R., Marlow, C. S., Mathews, C. J., McGuire, L., Hoffman, A. J., Balkwill, F., Burns, K. P., Butler, L., Drews, M., Fields, G., Smith, H., Winterbottom, M., Mulvey, K. L., Hartstone-Rose, A., & Rutland, A. (2024). Reciprocal associations between science efficacy, STEM identity, and scientist career interest among adolescent girls within the context of informal science learning. Journal of Youth and Adolescence, 53, 322–338. https://doi.org/10.1007/s10964-023-01868-6

[6] Du, Y., Borchers, C., & Cukurova, M. (2025). Benchmarking educational LLMs with analytics: A case study on gender bias in feedback. arXiv. https://arxiv.org/abs/2511.08225

[7] Tuchman, G. (1978). The symbolic annihilation of women by the mass media. In G. Tuchman, A. K. Daniels, & J. Benét (Eds.), Hearth and home: Images of women in the mass media. Oxford University Press.

[8] Girrbach, L., Alaniz, S., Smith, G., & Akata, Z. (2025). A large-scale analysis of gender biases in text-to-image generative models. arXiv. https://arxiv.org/abs/2503.23398

[9] Merton, R. K. (1968). The Matthew effect in science. Science, 159(3810), 56–63.

[10] Lietz, H., Jadidi, M., Kostic, D., Tsvetkova, M., & Wagner, C. (2024). Individual and gender inequality in computer science: A career study of cohorts from 1970 to 2000. Quantitative Science Studies, 5(1), 128–152. https://doi.org/10.1162/qss_a_00283

[11] Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Transactions on Information Systems, 14(3), 330–347.

[12] Steinke, J. (2017). Adolescent girls' STEM identity formation and media images of STEM professionals: Considering the influence of contextual cues. Frontiers in Psychology, 8, 716. https://doi.org/10.3389/fpsyg.2017.00716

[13] Bedynska, S., Krejtz, I., & Sedek, G. (2018). Chronic stereotype threat is associated with mathematical achievement in a representative sample of secondary school girls: The role of gender identification, working memory, and intellectual helplessness. Frontiers in Psychology, 9, 428. https://doi.org/10.3389/fpsyg.2018.00428

[14] Kalim, U., Kanwar, A., Xu, L., Xu, H., Chen, X., Irfan, S., & Rasool, A. (forthcoming). Female gender bias in artificial intelligence applications for education: A systematic review of regional disparities and equity implications. AI & Society. https://doi.org/10.1007/s00146-025-02811-y

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.