World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

Algorithmic bias in educational systems: Examining the impact of AI-driven decision making in modern education

Obed Boateng * and Bright Boateng

Independent Researcher, USA.

World Journal of Advanced Research and Reviews, 2025, 25(01), 2012-2017

Publication history: Received on 15 December 2024; revised on 25 January 2025; accepted on 28 January 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.25.1.0253

Abstract

The increasing integration of artificial intelligence and algorithmic systems in educational settings has raised critical concerns about their impact on educational equity. This paper examines the manifestation and implications of algorithmic bias across various educational domains, including admissions processes, assessment systems, and learning management platforms. Through analysis of current research and studies, we investigate how these biases can perpetuate or exacerbate existing educational disparities, particularly affecting students from marginalized communities. The study reveals that algorithmic bias in education operates through multiple channels, from data collection and algorithm design to implementation practices and institutional policies. Our findings indicate that biased algorithms can significantly impact students' educational trajectories, creating new forms of systemic barriers in education. We propose a comprehensive framework for addressing these challenges, combining technical solutions with policy reforms and institutional guidelines. This research contributes to the growing discourse on ethical AI in education and provides practical strategies for creating more equitable educational systems in an increasingly digitized world.

Keywords: Algorithmic Bias; Education; Artificial Intelligence; Education Equity

1. Introduction

Algorithms are playing an increasingly significant role in education, particularly in higher education and online learning environments. They are being used for various purposes, including predicting student behavior, automating decision-making processes, and personalizing learning experiences (Knox, 2018; McConvey & Guha, 2024). While algorithms promise efficiency and cost-savings, their implementation raises concerns about surveillance, fairness, and the rise in existing inequities (McConvey & Guha, 2024; McConvey et al., 2023). Despite the trend towards more complex algorithms and increased use of personal data, there is a lack of human-centered approaches in their development, leading to challenges in interpretability and explainability (McConvey et al., 2023). As algorithms become more entangled with educational processes, there is a need for critical discourse and resistance to ensure ethical implementation (Knox, 2018; Thibeault, 2014).

Algorithmic bias in education is a growing concern, with potential impacts on various demographic groups including race, gender, nationality, socioeconomic status, and disability (Baker & Hawn, 2021). The bias can originate from multiple sources within the machine learning pipeline, affecting measurement, model learning, and action stages (Kizilcec & Lee, 2020). To address these issues, researchers have proposed various strategies such as adjusting sample weights, bias attenuation methods, and adversarial learning (Idowu, 2024). Additionally, speculative fiction and liberatory design approaches have been suggested to engage underrepresented communities in identifying and solving challenges (Gaskins, 2022). Fairness assessment metrics like ABROCA and disparity metrics are commonly used to evaluate algorithmic fairness (Idowu, 2024). Moving forward, it is crucial to evaluate data and feature fairness before

* Corresponding author: Obed Boateng.

addressing algorithmic fairness, expand studies beyond gender and race, and assess the impact of fair algorithms on end users (Idowu, 2024).

Educational equity remains a critical concern in modern education systems, with research highlighting persistent disparities based on gender, geographic location, and socioeconomic status (Appiah Takyi et al., 2021). Studies have shown that within-school differences have a greater impact on educational outcomes than between-school variations, emphasizing the importance of examining instructional organization and practices (Gamoran, 1989). The allocation of educational resources, family backgrounds, and educational policies significantly influence access to higher education opportunities (Li, 2024). Recent developments in generative AI technologies present both opportunities and challenges for educational equity. While AI-assisted tutoring and personalized learning experiences can enhance accessibility and support, particularly in resource-constrained environments, concerns about digital divides and potential bias perpetuation remain (Gabriel, 2024). Addressing these issues requires thoughtful integration of AI technologies, balanced with human instruction, to create more equitable learning environments and expand access to quality education for all students.

2. Literature Review

2.1. Understanding Algorithmic Bias in Educational Context

Algorithmic bias in education is concern as data-driven models become more prevalent in supporting students, instructors, and administrators (Kizilcec & Lee, 2020). These biases can manifest in various ways, affecting groups based on race, ethnicity, gender, nationality, socioeconomic status, and disability (Baker & Hawn, 2021). Addressing these issues requires a multifaceted approach, incorporating statistical, similarity-based, and causal notions of fairness (Kizilcec & Lee, 2020). Some researchers propose using speculative fiction and liberatory design thinking to better understand the contexts of underrepresented groups (Gaskins, 2022). However, achieving total fairness may be impossible, as different definitions of fairness often require different solutions, particularly in high-stakes educational applications like language proficiency assessments (Loukina et al., 2019). Researchers have identified multiple types of algorithmic discrimination, including bias by algorithmic agents, feature selection bias, proxy discrimination, disparate impact, and targeted advertising (Wang et al., 2024).

These biases can disproportionately affect marginalized groups based on race, ethnicity, gender, nationality, socioeconomic status, and disability (Baker & Hawn, 2021). The roots of algorithmic injustice in education are traced to colonial continuities and historical biases embedded in data and system design (Karumbaiah & Brooks, 2021). Sources of bias include theoretical, methodological, interpretative, decontextualization, and data training issues, which are distributed across various elements of the educational system (Ferrero & Gewerc Barujel, 2019). Addressing these challenges requires interdisciplinary research, proactive policy development, public awareness, and ongoing monitoring to promote fairness and accountability in algorithmic decision-making within educational contexts (Wang et al., 2024).

Recent research highlights the challenges of algorithmic bias in educational AI systems and proposes strategies to address them. Studies have identified various forms of bias, including data-related, algorithmic, and user-interaction biases, which can perpetuate prejudice against specific demographics (Chinta et al., 2024). Common mitigation strategies include adjusting sample weights, bias attenuation methods, and adversarial learning (Idowu, 2024). While fairness metrics like ABROCA and disparity metrics are used to assess bias (Idowu, 2024), some argue that focusing solely on performance disparities is inadequate for addressing systemic inequities (Madaio et al., 2021). Researchers emphasize the need for context-specific approaches, evaluating data and feature fairness before algorithmic fairness, and expanding studies beyond gender and race (Idowu, 2024). To achieve equitable outcomes, a collaborative, interdisciplinary approach is recommended, considering ethical and legal frameworks in shaping fair educational environments (Chinta et al., 2024; Baker & Hawn, 2021).

2.2. Major Areas of Impact

2.2.1. Admissions and Enrollment

Recent research highlights significant concerns regarding algorithmic bias in college admissions processes. Studies show that removing race data from applicant ranking algorithms reduces diversity without improving academic merit (Lee et al., 2024). Machine learning models used in admissions can perpetuate societal injustices, including racism, by relying on historical data (G'andara et al., 2023). The ban on race-conscious admissions has forced changes to these models, potentially increasing arbitrariness in outcomes for most applicants (Lee et al., 2024). To address these issues, researchers are developing tools to minimize both human and algorithmic bias in the selection process (Martinez Neda et al., 2021). However, challenges persist, as predictive algorithms for student success continue to produce racially biased results despite efforts to mitigate unfairness (G'andara et al., 2023). These findings underscore the complex interplay between algorithmic decision-making and diversity in higher education admissions.

2.2.2. Assessment and Grading

Algorithmic bias in educational assessment and grading systems poses significant challenges to fairness and equity. Automated Essay Scoring (AES) systems have shown biases related to students' gender, race, and socioeconomic status (Litman et al., 2021). Rater bias can also affect AES models, as they are typically designed to emulate human raters (Amorim et al., 2018). However, some studies have found no evidence of bias in certain automated writing quality scores, suggesting that bias may stem from other assessment components like multiple-choice questions (Matta et al., 2023). The impact of algorithmic bias extends beyond AES to various educational technologies, affecting students based on race/ethnicity, gender, nationality, socioeconomic status, disability, and military-connected status (Baker & Hawn, 2021). Addressing these issues requires a comprehensive approach, including improved bias detection methods, fairness metrics, and efforts to move from fairness to equity in educational algorithms (Baker & Hawn, 2021).

2.2.3. Learning Management and Support

As data-driven technologies become more prevalent, the need to address algorithmic bias in education will continue to persist. Studies on Bayesian Knowledge Tracing and carelessness detectors have shown promising results, with performance being relatively equal across demographic groups (Zambrano et al., 2024). However, traditional bias metrics may not be suitable for educational settings due to hierarchical dependencies in classrooms, necessitating adapted measurements using hierarchical linear models (Belitz et al., 2024). To address these challenges, researchers recommend focusing on solidifying understanding of concrete impacts, moving from unknown to known bias, and transitioning from fairness to equity (Baker & Hawn, 2021). Policymakers and developers are advised to consider statistical, similarity-based, and causal notions of fairness when designing and implementing educational algorithms to promote fairness and mitigate potential biases (Kizilcec & Lee, 2020).

2.2.4. Documented instances of algorithmic bias in education

Studies have shown that predictive models used in higher education can reinforce racial inequities, with Black students often receiving fewer resources when algorithms are used to identify "at-risk" students (Bird et al., 2024). The magnitude of bias varies depending on how "at-risk" is defined, emphasizing the contextual nature of algorithmic bias (Bird et al., 2024). Research has also demonstrated that commonly used features in college student success prediction models can produce racially biased results (G'andara et al., 2023). To address these issues, researchers recommend ensuring that privacy requirements do not hinder the identification and mitigation of bias (Baker et al., n.d.). Additionally, collecting more comprehensive data may help reduce bias, particularly for new students and underrepresented groups (Bird et al., 2024).

2.2.5. Implications for Educational Equity

Algorithmic bias in educational systems poses significant threats to educational equity, potentially amplifying existing social and economic disparities in unprecedented ways (Baker & Hawn, 2021). When AI-driven systems perpetuate or increase biases, they can systematically disadvantage students from marginalized communities, creating a self-reinforcing cycle of educational inequality (Farahani & Ghasemi, 2024). For instance, predictive analytics tools that use historical data to forecast student success may inherit past patterns of discrimination, leading to reduced opportunities for certain demographic groups in areas like college admissions, course recommendations, or academic support services.

These biases can manifest in subtle yet pervasive ways, such as adaptive learning platforms that may not effectively accommodate diverse learning styles or cultural perspectives, or early warning systems that disproportionately flag minority students as "at-risk" (Bird et al., 2024). The long-term implications are particularly concerning, as these algorithmic decisions can impact students' educational trajectories, career opportunities, and socioeconomic mobility (Zeide, 2017). Moreover, the widespread adoption of these technologies in education, accelerated by remote learning needs, means that biased algorithms could systematically affect entire generations of students, creating new forms of digital redlining in education (Baker & Hawn, 2021). This technological stratification threatens to undermine decades of progress in educational equity, making it crucial for institutions to critically examine and address algorithmic bias in their educational technology implementations.

3. Mitigation Strategies and Solutions

The mitigation of algorithmic bias in educational systems requires a comprehensive, multi-stakeholder approach that combines technical innovation with robust policy frameworks and institutional accountability (Idowu, 2024). Researchers have identified several promising strategies, including algorithmic auditing protocols that systematically evaluate AI systems for potential biases across different demographic groups and educational contexts (Murikah et al., 2024). For instance, the implementation of fairness metrics and regular bias assessments (Pagano et al., 2023) throughout the development and deployment lifecycle can help identify discriminatory patterns before they impact student outcomes. Educational institutions must also prioritize diverse representation in technology development teams and incorporate feedback from affected communities, particularly those historically marginalized in educational settings (Smolansky et al., 2023).

Technical solutions such as debiasing techniques (Idowu, 2024), which involve adjusting model parameters and training data to ensure more equitable outcomes, should be complemented by transparent documentation of algorithmic decision-making processes and regular impact assessments. Furthermore, the establishment of clear governance frameworks and ethical guidelines for educational AI deployment, coupled with mandatory bias impact assessments (Shukla, 2024), can help institutions proactively address potential inequities. These strategies should be supported by ongoing professional development for educators and administrators to build institutional capacity for critical evaluation of algorithmic systems and their implications for educational equity. Success in mitigating algorithmic bias ultimately depends on sustained commitment to these strategies and regular evaluation of their effectiveness through rigorous empirical research.

4. Conclusion and Recommendation

The future considerations surrounding algorithmic bias in education present a complex landscape of challenges and opportunities that demand careful attention from researchers, educators, and policymakers. As artificial intelligence and machine learning technologies continue to evolve at an unprecedented pace, educational institutions face the critical task of anticipating and preparing for emerging forms of algorithmic bias. The integration of more sophisticated AI systems, including large language models and advanced predictive analytics, into educational processes raises new questions about fairness, transparency, and accountability. Of particular concern is the potential emergence of more subtle and complex forms of bias that may be harder to detect and mitigate, especially as these systems become more autonomous in decision-making processes. Additionally, the growing trend toward personalized learning environments, while promising, could inadvertently create new forms of educational segregation if not carefully monitored and regulated.

Through examining the multifaceted nature of algorithmic bias in educational settings, it becomes evident that these technologies can either bridge or widen existing educational gaps, depending on how they are developed, implemented, and monitored. The impact of biased algorithms extends far beyond immediate academic outcomes, potentially shaping students' educational trajectories, career opportunities, and socioeconomic mobility for generations. As educational institutions increasingly rely on AI-driven decision-making systems, the urgency to address algorithmic bias cannot be overstated. The intersection of technology and education must be carefully navigated to ensure that innovation serves to enhance rather than hinder educational equity.

The effective mitigation of algorithmic bias in education requires educational institutions to establish robust oversight mechanisms, including regular audits and diverse task forces, while investing in comprehensive staff training on AI literacy and bias detection. Technology developers must prioritize transparency and diverse representation in their development teams, implementing rigorous testing protocols and creating adaptable systems that can accommodate different cultural and educational contexts. Policymakers should focus on creating clear regulatory frameworks and mandatory equity impact assessments, while allocating sufficient resources for research into bias mitigation strategies and ethical AI implementation standards. Educators and administrators must maintain active oversight of algorithmic systems, participate in ongoing professional development, and foster open dialogue with students and families about the impact of educational technologies, ensuring that innovation serves to enhance rather than hinder educational equity.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Amorim, E., Cançado, M., & Veloso, A. (2018). Automated Essay Scoring in the Presence of Biased Ratings. North American Chapter of the Association for Computational Linguistics.

[2] Baker, R., & Hawn, A. (2021). Algorithmic Bias in Education. International Journal of Artificial Intelligence in Education, 32, 1052 - 1092.

[3] Belitz, C., Lee, H., Nasiar, N., Fancsali, S.E., Ritter, S., Almoubayyed, H., Baker, R.S., Ocumpaugh, J.L., & Bosch, N. (2024). Hierarchical Dependencies in Classroom Settings Influence Algorithmic Bias Metrics. Proceedings of the 14th Learning Analytics and Knowledge Conference.

[4] Bird, K.A., Castleman, B.L., & Song, Y. (2024). Are algorithms biased in education? Exploring racial bias in predicting community college student success. Journal of Policy Analysis and Management.

[5] Chinta, S.V., Wang, Z., Yin, Z., Hoang, N., Gonzalez, M., Le Quy, T., & Zhang, W. (2024). FairAIED: Navigating Fairness, Bias, and Ethics in Educational AI Applications. ArXiv, abs/2407.18745.

[6] Fu, R., Huang, Y., & Singh, P.V. (2020). AI and Algorithmic Bias: Source, Detection, Mitigation and Implications. InfoSciRN: Machine Learning (Sub-Topic).

[7] Gabriel, S. (2024). Generative AI and Educational (In)Equity. International Conference on AI Research.

[8] Gamoran, A. (1989). Instructional Organization and Educational Equity.

[9] Gaskins, N. (2022). Interrogating Algorithmic Bias: From Speculative Fiction to Liberatory Design. Techtrends, 67, 417 - 425.

[10] Gándara, D., Anahideh, H., Ison, M.P., & Tayal, A. (2023). Inside the Black Box: Detecting and Mitigating Algorithmic Bias across Racialized Groups in College Student-Success Prediction. ArXiv, abs/2301.03784.

[11] Idowu, J.A. (2024). Debiasing Education Algorithms. International Journal of Artificial Intelligence in Education, 1-31.

[12] Karumbaiah, S., & Brooks, J. (2021). How Colonial Continuities Underlie Algorithmic Injustices in Education. 2021 Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT), 1-6.

[13] Kizilcec, R.F., & Lee, H. (2020). Algorithmic Fairness in Education. ArXiv, abs/2007.05443.

[14] Knox, J. (2018). Beyond the “c” and the “x”: Learning with algorithms in massive open online courses (MOOCs). International Review of Education, 64, 161 - 178.

[15] Lee, J., Harvey, E., Zhou, J., Garg, N., Joachims, T., & Kizilcec, R.F. (2024). Algorithms for College Admissions Decision Support: Impacts of Policy Change and Inherent Variability. ArXiv, abs/2407.11199.

[16] Li, C. (2024). Research on the Impact of Educational Equity on Access to Higher Education Opportunities. Lecture Notes in Education Psychology and Public Media.

[17] Litman, D.J., Zhang, H., Correnti, R., Matsumura, L.C., & Wang, E.L. (2021). A Fairness Evaluation of Automated Methods for Scoring Text Evidence Usage in Writing. International Conference on Artificial Intelligence in Education.

[18] Loukina, A., Madnani, N., & Zechner, K. (2019). The many dimensions of algorithmic fairness in educational applications. BEA@ACL.

[19] Madaio, M.A., Blodgett, S., Mayfield, E., & Dixon-Rom'an, E. (2021). Beyond "Fairness:" Structural (In)justice Lenses on AI for Education.

[20] Matta, M., Mercer, S.H., & Keller-Margulis, M.A. (2023). Implications of bias in automated writing quality scores for fair and equitable assessment decisions. School Psychology, 38(3), 173-181.

[21] McConvey, K., & Guha, S. (2024). "This is not a data problem": Algorithms and Power in Public Higher Education in Canada. Proceedings of the CHI Conference on Human Factors in Computing Systems.

[22] McConvey, K., Guha, S., & Kuzminykh, A. (2023). A Human-Centered Review of Algorithms in Decision-Making in Higher Education. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.

[23] Murikah, W., Nthenge, J.K., & Musyoka, F.M. (2024). Bias and Ethics of AI Systems Applied in Auditing - A Systematic Review. Scientific African.

[24] Neda, B.M., Zeng, Y., & Masagué, S.G. (2021). Using Machine Learning in Admissions: Reducing Human and Algorithmic Bias in the Selection Process. Proceedings of the 52nd ACM Technical Symposium on Computer Science Education.

[25] Pagano, T.P., Loureiro, R.B., Lisboa, F.V., Peixoto, R.M., Guimarães, G.A., Cruz, G.O., ... & Nascimento, E.G. (2023). Bias and unfairness in machine learning models: a systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. Big Data and Cognitive Computing, 7(1), 15.

[26] Shahvaroughi Farahani, M., & Ghasemi, G. (2024). Artificial Intelligence and Inequality: Challenges and Opportunities. Qeios.

[27] Shukla, S. (2024). Principles Governing Ethical Development and Deployment of AI. Journal of Artificial Intelligence Ethics, 7(1), 45-60.

[28] Smolansky, A., Nguyen, H.A., Kizilcec, R.F., & McLaren, B.M. (2023). Equity, Diversity, and Inclusion in Educational Technology Research and Development. International Conference on Artificial Intelligence in Education.

[29] Takyi, S.A., Amponsah, O., Asibey, M.O., & Ayambire, R.A. (2021). An overview of Ghana's educational system and its implication for educational equity. International Journal of Leadership in Education, 24, 157 - 182.

[30] Thibeault, M.D. (2014). Algorithms and the Future of Music Education: A Response to Shuler. Arts Education Policy Review, 115, 19 - 25.

[31] Wang, X., Wu, Y.C., Ji, X., & Fu, H. (2024). Algorithmic discrimination: Examining its types and regulatory measures with emphasis on US legal practices. Frontiers in Artificial Intelligence, 7.

[32] Zambrano, A.F., Zhang, J., & Baker, R.S. (2024). Investigating Algorithmic Bias on Bayesian Knowledge Tracing and Carelessness Detectors. Proceedings of the 14th Learning Analytics and Knowledge Conference.

[33] Zeide, E. (2017). The structural consequences of big data-driven education. Big Data, 5(2), 164-172.