IJSES
Volume 9, Issue 10, pp. 46-50, 2025.
International Journal of Scientific Engineering and Science
ISSN (Online): 2456-7361

Algorithmic Bias and Education: Artificial Intelligence as a Factor Exacerbating Educational Inequities

Yahya Haoumi[1], Younes Wadiai[2], Fatima Ezahra El Kamouny[3]
[1]Public Administration and Information Sciences, School of Business, Long Island University, New York, USA.
[2]Laboratory of Innovative Systems Engineering, National School of Applied Sciences of Tetouan, Abdelmalek Essaâdi University, Tetouan, Morocco.
[3]Laboratory LAROSERI, Department of Computer Science, Faculty of Science, Chouaib Doukkali University, El Jadida, Morocco.
Email address: yahya.haoumi@gmail.com, y.wadiai@uae.ac.ma, f.elkamouny@gmail.com

Abstract— *The integration of artificial intelligence (AI) into education is rapidly transforming teaching, learning, and assessment pra ctices. Tools such as adaptive learning platforms, predictive analytics, and automated grading systems are increasingly adopted wit h the promise of improving academic outcomes and supporting personalized learning. However, these systems often rely on datasets that are incomplete, skewed, or demographically unrepresentative, which can inadvertently reproduce and even amplify existing s ocial and educational inequities. This paper explores the phenomenon of algorithmic bias in educational AI, situating it within broader societal and structural contexts, and reviews empirical evidence from prior studies that reveal disproportionate impacts on marginalized student groups. To complement this review, w e present an empirical investigation into the influence of ChatGPT on student performance. Using Support Vector Machines (SVM) and Random Forest (RF) models to analyze grade data before and after AI use, the study demonstrates measurable improve-ments in student achievement, with SVM showing superior predictive ac-curacy and RF offering greater robustness to variability. While these results highlight AI's potential to enhance academic performance, they also underscore persistent challenges, such as class imbalance and calibration issues, which refle ct deeper structural biases. The findings suggest that the benefits of AI in education can only be realized through careful, transparent design, inclusive data practices, and sustained attention to equity. Ul-timately, the study argues that AI in education must move beyond tech-nical efficiency to actively promote fairness, inclusivity, and social justice.*

Keywords— *Artificial Intelligence, Algorithmic Bias, Educational Equity, ChatGPT, Machine Learning, Support Vector Machines, Random Forest, Student Achievement, Inclusive Design, Ethical AI.*

I. Introduction

It is undeniably true that integrating various adaptive learning systems, predictive analytical tools, automated grading or scoring tools, and artificial intelligence into education offers many benefits aimed at improving students' academic performance. Nonetheless, these technological affordances are accompanied by profound risks and ethical intricacies. These issues arise because AI systems are often engineered using datasets that fail to provide comprehensive and demographically representative coverage of all students. As such they inherit and, sometimes, even propagate latent biases, stereotypes, and structural inequities that are embedded within the training data. This is because these tools rely on collected datasets that may contain prejudice and other stereotypes. In other words, they do not undergo training on data that is responsibly gathered and supplied to ensure equity and inclusivity for all. Perhaps this is because the discipline is still developing and much more work is needed before the complex mosaic of our modern human society can be represented objectively to this degree. As a result, there is a significant chance that social and educational disparities will be copied and even exacerbated. This is more of an ethical and structural issue than a purely technological one. This study examines how bias in AI appears in educational technologies and how it can intensify the differences between disadvantaged and privileged student groups. We consider the historical context, practical applications, and professional advice for more equitable AI development.

II. Literature Review

Prior studies have demonstrated that algorithmic bias in criminal justice, healthcare, and finance can result in unequal and different results according to gender, race, and socioeconomic position [1, 2]. The main explanation for this is because algorithms are trained using the data that is supplied to them. As a result, the system will learn to reflect and reinforce any biases included in this data. AI has been incorporated into education without a careful consideration of the ethical ramifications. Predictive methods that evaluate "student risk," according to a number of experts, rely on datasets that are frequently skewed or lacking [3]. Students from minority or underprivileged backgrounds are often underrepresented as a result of these restrictions, which reduces the effectiveness of these programs for these groups [4]. Additionally, cultural and linguistic varieties are marginalized by language models employed for tutoring or feedback, which frequently reflect Western and affluent norms [5]. Recent research offers specific examples of these problems. The requirement for context-aware techniques is highlighted by Švábenský et al. [6], who demonstrated that geographical origin affects the fairness of predicted academic performance models in the Philippines.

According to Gándara et al. [7], tested mitigation strategies are still insufficient, and algorithms used to forecast minority students' success at universities are less accurate. The issue was expanded to include learning management systems, admissions, and assessments by Boateng & Boateng [8], who came to the conclusion that algorithmic biases aid in the perpetuation of educational disparities. Weissburg et al. [9] emphasized the detrimental effects of demographic biases in large language models, including those pertaining to handicap, wealth, and other characteristics, on the learning process. Lastly, Vartiainen et al. [10] concentrated on using generative AI in Finnish schools to teach students about algorithmic biases in order to foster critical thinking abilities. Lastly, Vartiainen et al. [10] concentrated on using generative AI in Finnish schools to teach students about algorithmic biases in order to foster critical thinking abilities. This discussion is enhanced by further scholarly contributions. Few research examine biases in particular contexts or fields, as demonstrated by Li et al.'s [11] investigation of social biases in ChatGPT in higher education. With their FairAIED paradigm, Chinta et al. [12] offer a typology of biases pertaining to data, algorithms, and user interactions along with remedial techniques. Madaio et al. [13] support a method that incorporates a viewpoint on historical and structural inequalities in addition to the straightforward idea of "algorithmic fairness." From an empirical standpoint, Gándara et al. [7] attest to the fact that existing corrective methods are still restricted and that prediction models of university success are less successful for colored students. The OECD [14] gave an overview of algorithmic bias in European education at the institutional level, citing specific instances (voice recognition, dropout prediction, automated grading), and developing policy recommendations to reduce these dangers. While pointing out obstacles such digital gaps and inadequate teacher preparation, a systematic evaluation by MDPI [15] also emphasized AI's inclusive potential, especially for kids with impairments. Technically speaking, a study that was published in the International Journal of Artificial Intelligence in Education [16] examined debiasing techniques like resampling and examined sensitive characteristics that are taken into account in equity research, such as gender, native language, and socioeconomic position. There are also more and more warnings. Emerald Insight [17] highlights the danger of escalating inequality if tools are badly constructed and calls for coordinating the use of AI in education with Sustainable Development Goal 4. Exam surveillance software and ethical biases in academic integrity are highlighted by IJRISS [18], which unjustly penalizes marginalized students, particularly Black women. Lastly, Edutopia [19-20] highlights efforts to lessen prejudices by creating inclusive tools like Latimer.ai and increasing diversity in design teams. In general, research indicates that educational AI is not neutral; rather, it is a reflection of prevailing societal structures and disparities. In addition to reproducing, algorithmic biases have the ability to magnify educational disparities. Although there are various technical mitigating techniques, they are still only partially effective. However, AI still has a lot of room to be inclusive if its use is supported by transparent educational practices, explicit policies, and design that is truly sensitive to the diversity of learners.

III. METHODOLOGY

The procedures in the suggested methodology for examining how ChatGPT use affects students' academic achievement are exact. Academic results from students enrolled in the Bachelor of Computer Science program at Chouaïb Doukkali University of El Jadida's Faculty of Sciences for the 2020 academic year are gathered at the start of the procedure. To guarantee their quality and uniformity, the raw grades are preprocessed both before and after using ChatGPT. This preparation entails handling missing values, eliminating duplicates, anonymizing students by providing unique identities, and confirming the authenticity of grades (e.g., $0 \leq$ grade ≤ 20). The challenge is changed to a binary classification assignment to make the analysis easier: predicting if a student would use ChatGPT to raise their grade by two points or more (Success = Yes / Failure = No). Two prediction models are created in light of this:

- To capture linear or non-linear correlations between variables, use Support Vector Machines(SVMs).
- Random Forest, a supplementary method to SVM, was selected due to its capacity to manage heterogeneous data and minimize overfitting.

These models are then trained using the preprocessed data, and their performance and stability are evaluated using cross-validation. To find the optimal configurations for each model, grid search is used for hyperparameter optimization. The capacity of the improved models to accurately predict student success or failure is evaluated by retraining and testing on unseen data. Metrics including accuracy, precision, recall, F1-score, and AUC that are appropriate for binary classification are used to assess performance. In order to demonstrate the models' capacity to forecast ChatGPT's influence on academic achievement, the findings are finally displayed using confusion matrices and comparing charts. This methodology guarantees a thorough, repeatable, and practical way to assess the AI tool's impact in an educational setting.

A. Dataset Description and Preprocessing

We gathered the grades of 200 students enrolled in the Bachelor of Computer Science program for the 2020 academic year in order to assess ChatGPT's effect on academic success. Every student has two grades: one before and one after using the AI tool. This creates a paired sample that makes it possible to examine each tool's unique impact. The grade distribution demonstrates an overall improvement following ChatGPT use, as seen in Figure 1. Prior to AI, the majority of scores fell between 10 and 14/20, with a high around 12 or 13/20. The distribution changes in favor of higher scores after AI, peaking between 14 and 15/20 and rising over 16/20. These findings imply that including ChatGPT leads to a discernible gain in academic performance, while they also highlight individual differences in the extent of improvement.

An intensive preprocessing phase was used to make sure the data met our analytical needs. This stage is crucial because it enables us to prepare and polish the raw data, turning it into a format that works for our modeling and analytic procedures. A number of essential preprocessing steps are included in order to enhance the dataset's quality and suitability for sophisticated

modeling methods. In order to facilitate machine learning model learning and guarantee that all values are on a comparable scale, the data were meticulously cleaned and transformed.

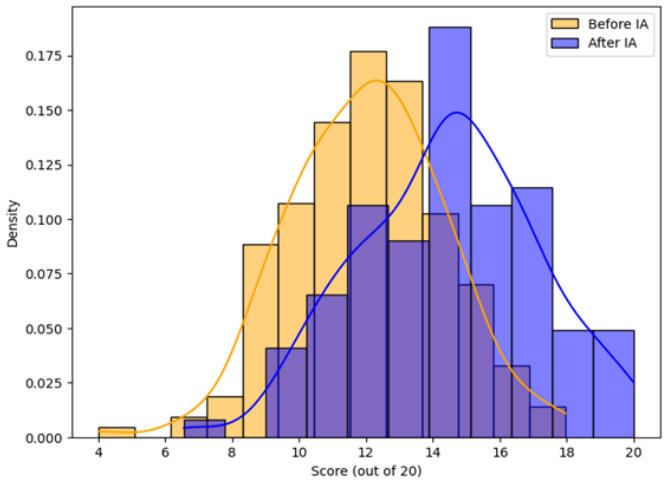

Fig. 1. Distribution of Students' Grades Before and After Using AI

Duplicates were eliminated, all grades were confirmed to fall within the 0–20 range, missing values were handled through imputation or exclusion when needed, and grades were normalized. A particular coding was used to turn the problem into a binary classification task: a student was rated as Successful = 1 if their grade improved by two points or more after utilizing ChatGPT in comparison to their initial grade; if not, they were labeled as Failure = 0. By turning the prediction of continuous grades into a classification problem, this coding streamlines the study and enables models to calculate the likelihood that a student will achieve a notable improvement in performance through artificial intelligence. To further improve the accuracy and robustness of the model, more preprocessing activities like resolving class imbalances between Success and Failure or developing explanatory variables may be taken into consideration. For reliable and accurate predictions, they guarantee that machine learning algorithms receive data that is clean, standardized, and appropriately scaled.

B. *Support Vector Machine*

The Support Vector Machine (SVM) [21] is one supervised learning technique that is widely recognized for its robustness and ability to handle classification and regression problems. This method may employ a linear or non-linear strategy by utilizing kernel functions, allowing the modeling of complex data interactions [22]. determining the best The fundamental concept underlying Support Vector Machines (SVM), which improves the model's ability to generalize, is the hyperplane that maximizes the margin between points from different classes. These hyperplanes facilitate the precise classification of new observations by acting as decision boundaries [23. Although SVM's main function is in classification [24], it may also be used to regression problems, where it seeks to match an approximation function by reducing errors. The model's regularization is governed by the parameter C, which seeks to balance the model's complexity with its ability to tolerate categorization errors. A low C value allows for more errors, but it also allows for a wider margin, whereas a high C value reduces mistakes at the cost of overfitting [25]. The Radial Basis Function (RBF) kernel is frequently employed for dealing with nonlinear data utilized. With the use of this kernel, the data can be projected onto a higher-dimensional space, enabling a linear separation. Its purpose is outlined by:

$$k(x, x') = exp\ exp\ (-\gamma(x - x')^2) \qquad (1)$$

where γ gamma (gamma) is a hyperparameter that regulates the local influence of a training vector, and $||x - x'||$ is the squared Euclidean distance between two locations x and x'. While very low values of γ gamma may underestimate the complexity of the data, high values might cause overfitting by making the decision boundary too sensitive to individual data points [26]. For the classifier to perform as well as feasible, it is essential to simultaneously adjust the parameters C and γ gamma in order to achieve an ideal trade-off between bias and variance [28].

C. *Random Forest*

The Random Forest (RF) approach is a machine learning technique that integrates several decision trees under supervision. Leo Breiman first introduced it in 2001 [29]. It combines two essential methods: bagging and the random selection of variable subsets for each split. Even when working with noisy or extremely connected data, this combination strategy lowers the model's variance, increases its ability to generalize, and lessens overfitting [30]. The fundamental idea behind RF is to construct a set of separate decision trees, each of which is based on a bootstrapped sample of the learning game. During the development of each tree, a sub-ensemble aléatoire of variables is selected at each noeud to achieve the optimum separation. As a result of this process, there was structural variety among the trees, which greatly improved their overall resilience. The last prediction is made by a majority vote if it's a classification or by an average vote after the fort is built. in the event of a regression. Additionally, because of the bootstrap's nature, around one layer of data, referred to as out-of-bag or OOB, is not used to build any one tree. As a result, this information may be utilized to evaluate the model's performance without requiring an explicit cross-check [31]. The RF provides a number of advantages, such as the ability to handle big data sets, naturally handle complex variable interactions, and provide measures of the significance of attributes. and continue to be somewhat immune to outlier values. However, its effectiveness may be reduced in situations of extremely imbalanced classes if no corrective actions are implemented [32], and it may be less intelligible than a single decision tree.

IV. Result and Discussion

The impact of ChatGPT on improving academic performance was assessed using two supervised learning models: Random Forest (RF) and Support Vector Machine (SVM). An overview of their performances is given in Table 2, and a comparison between each model's predictions and actual values is displayed in Figure 2-3.

With cross-validation and testing accuracy of 0.970 and 0.986, respectively, the SVM model demonstrated a strong

capacity for generalization. The model successfully identifies the majority of students who saw an improvement in their grades after utilizing AI, as seen by the recall of 0.980. The model's dependability is demonstrated by the tight correspondence between the forecasts and actual values, as seen in Figure 2's left section. The accuracy (3.8446) and F-measure (1.689) metrics, however, show some discrepancies, suggesting a calibration problem associated with the imbalance between the success and failure classes.

TABLE 1. Performance Results of SVM,and Random Forest, in Cross-Validation and Testing

Model	Phase	Accuracy	Error Rate	Precision	Recall	F-Measure
SVM	Cross-Validation	0.97	0.012	3 .8446	0.98	1.689
	Testing	0.986	0.014			-
RF	Cross-Validation	0.960	0.040	1.5700	0.960	1.191
	Testing	0.910	0.050			

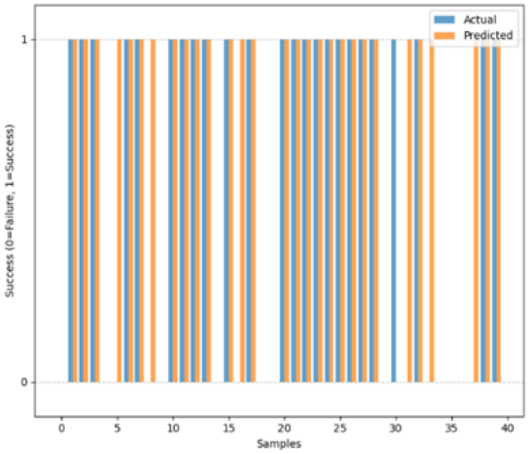

Fig. 2. SVM: Actual vs. Predicted

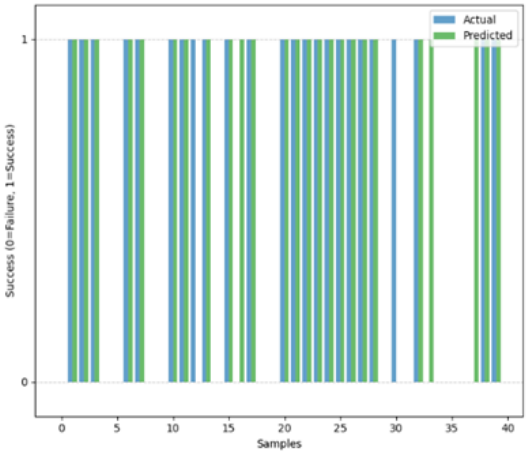

Fig.3. Random Forest: Actual vs. Predicted

With cross-validation accuracy of 0.960 and testing accuracy of 0.910, the Random Forest model performed somewhat worse than SVM. Despite having a recall of 0.960 and an F-measure of 1.191, it is nevertheless able to identify pupils who have made improvement. Although there are still some differences, its forecasts are largely in line with the actual values, as seen in the right portion of Figure 3. Decision trees' ensemble structure, which makes the model more resilient to inter-individual variability but marginally less accurate than SVM, explains this pattern. The results emphasize complimentary traits overall. Despite its outstanding accuracy and remarkable generalization capacity, SVM is nonetheless prone to calibration issues. However, because Random Forest is more resilient and more tolerant of data heterogeneity, it is more suited for real-world situations. These results show that employing ChatGPT enhanced academic performance, with SVM being more accurate and RF being more stable when data variability was present.

REFERENCES

1. R. S. Baker and A. Hawn, "Algorithmic bias in education," *Int. J. Artif. Intell. Educ.*, vol. 32, no. 4, pp. 1052–1092, 2022.
2. R. S. Baker, A. Hawn, and S. Lee, "Algorithmic bias: The state of the situation and policy recommendations," *Unpublished manuscript*, 2023.
3. R. S. Baker, M. Z. N. L. Saavedra, and A. Shimada, "Evaluating algorithmic bias in models for predicting academic performance of Filipino students," *arXiv preprint*, arXiv:2405.09821, 2024.
4. L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
5. J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Conf. Fairness, Accountability and Transparency*, PMLR, pp. 77–91, Jan. 2018.
6. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.
7. S. V. Chinta, Z. Wang, Z. Yin, N. Hoang, M. Gonzalez, T. L. Quy, and W. Zhang, "FairAIED: Navigating fairness, bias, and ethics in educational AI applications," *arXiv preprint*, arXiv:2407.18745, 2024.
8. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.
9. M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3133–3181, 2014.
10. S. A. Friedler, C. Scheidegger, and S. Venkatasubramanian, "The (im)possibility of fairness: Different value systems require different mechanisms for fair decision making," *Commun. ACM*, vol. 64, no. 4, pp. 136–143, 2021.
11. D. Gándara, H. Anahideh, M. P. Ison, and L. Picchiarini, "Inside the black box: Detecting and mitigating algorithmic bias across racialized groups in college student-success prediction," *Area Open*, vol. 10, p. 23328584241258741, 2024.
12. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Cambridge, MA: MIT Press, 2016. [Reviewed in: J. Heaton, "Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning," *Genet. Program. Evolvable Mach.*, vol. 19, no. 1, pp. 305–307, 2018.]
13. T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., New York: Springer, 2009.
14. C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," *Tech. Rep.*, National Taiwan University, 2003.
15. R. F. Kizilcec and H. Lee, "Algorithmic fairness in education," in *The Ethics of Artificial Intelligence in Education*, Routledge, pp. 174–202, 2022.
16. H. Lebovits, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, 2019.
17. M. Li, A. Enkhtur, B. A. Yamamoto, F. Cheng, and L. Chen, "Potential societal biases of ChatGPT in higher education: A scoping review," *Open Praxis*, vol. 17, no. 1, pp. 79–94, 2025.
18. A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
19. B. D. Lund, T. H. Lee, N. R. Mannuru, and N. Arutla, "AI and academic integrity: Exploring student perceptions and implications for higher education," *J. Acad. Ethics*, pp. 1–21, 2025.

20. M. Madaio, S. L. Blodgett, E. Mayfield, and E. Dixon-Román, "Beyond 'fairness': Structural (in)justice lenses on AI for education," in *The Ethics of Artificial Intelligence in Education*, Routledge, pp. 203–239, 2022.
21. N. Madnani, A. Loukina, A. Von Davier, J. Burstein, and A. Cahill, "Building better open-source tools to support fairness in automated scoring," in *Proc. 1st ACL Workshop on Ethics in NLP*, pp. 41–52, Apr. 2017.
22. V. A. Melo-López, A. Basantes-Andrade, C. B. Gudiño-Mejía, and E. Hernández-Martínez, "The impact of artificial intelligence on inclusive education: A systematic review," *Educ. Sci.*, vol. 15, no. 5, p. 539, 2025.
23. S. U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York: NYU Press, 2018.
24. B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, Cambridge, MA: MIT Press, 2002.
25. A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, pp. 199–222, 2004.
26. V. Švábenský, M. Verger, M. M. T. Rodrigo, and C. J. G. Monterozo, "Evaluating algorithmic bias in models for predicting academic performance of Filipino students," *arXiv preprint*, arXiv:2405.09821, 2024.
27. S. Wang, J. Zhang, Y. Fu, and Y. Li, "ACM transactions on intelligent systems and technology," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 4, 2011.
28. I. Žliobaitė, "Measuring discrimination in algorithmic decision making," *Data Min. Knowl. Discov.*, vol. 31, no. 4, pp. 1060–1089, 2017.