Ethical Implications of AI in Student Assessments: A Critical Examination

Sonny Soriano[1], Atty. Dar A. Diga[2], Dr. Easter Belandres[3], Dr. Joseph Barrozo[4], Dr. Eleuterio Sta. Cruz Sison[5] and Dr. Mary Ruth Cinchez[6]

[1]Urdaneta City University, Director, External Affairs and Linkages
[2]Urdaneta City University
[3]Urdaneta City University
[4]Urdaneta City University
[5]Urdaneta City University, President, UCU-Personnel Assosiation
[6]Urdaneta City University
sonnysoriano@ucu.edu.ph[1],dardiga@ucu.edu.ph[2],easterbelandres@ucu.edu.ph[3],   josephbarrozo@ucu.edu.ph[4],
eleuteriosison@ucu.edu.ph[5] and marycinchez@ucu.edu.ph[6]
[1]ORCID:https://orcid.org/0009-0001-5585-9205,[2]ORCID:https://orcid.org/0009-0009-4502-3146,
[3]ORCID#:https://orcid.org/0000-0002-5517-1486,[4]ORCID:https://orcid.org/0009-0007-5498-6257,   [5]ORCID: https://orcid.org/0009-0000-8889-6240 and [6]ORCID: https://orcid.org/0009-0004-6787-1406

ABSTRACT

This research delves into the ethical dimensions of employing Artificial Intelligence (AI) in student assessments within an Asian University. A comparative analysis between AI grading and human grading highlights AI's commendable 90% accuracy but raises concerns about its consistency compared to human grading's 95%. This trade-off underscores the need for transparent evaluation criteria and bias mitigation mechanisms, as AI systems' consistent undervaluation or overvaluation of student work may lead to unfair outcomes. The study examines AI grading bias and its impact on demographic groups, revealing negative bias scores in gender (-0.2), race/ethnicity (0.1), and socioeconomic status (-0.3). Addressing biases becomes imperative to ensure fair assessments, especially when AI grading exacerbates existing disparities. Privacy concerns related to AI student monitoring are discussed, detailing the types of collected data, their purposes, and retention periods. Ethical considerations revolve around responsibly handling sensitive student data and ensuring robust data security. The research explores the impact of AI-generated feedback on student confidence, noting a 70% positive impact alongside potential drawbacks, such as the absence of nuanced understanding found in human-generated feedback. Balancing timely assessments with empathetic feedback is crucial ethically. Furthermore, accessibility features in AI-enabled assessment tools are highlighted, aiming to provide equitable access for students with diverse needs. Ethical considerations stress the importance of fair opportunities for all students, regardless of their backgrounds. This research offers a comprehensive examination of the ethical implications of AI in student assessments, providing valuable insights for educators, policymakers, and technologists.

Keywords: Artificial Intelligence (AI) in Education, Ethical Implications, Student Assessments, Bias Mitigation, Privacy Concerns

INTRODUCTION

Artificial Intelligence (AI) is revolutionizing various aspects of education, including student assessments. As universities and educational institutions embrace AI technologies to streamline assessment processes, it is imperative to critically examine the ethical implications of this paradigm shift. The transition to AI-driven assessment systems brings forth a myriad of opportunities and challenges. While AI grading boasts commendable accuracy rates, there exists a palpable concern regarding its consistency when juxtaposed with human grading standards. This discrepancy underscores the necessity for transparent evaluation criteria and robust bias mitigation mechanisms to prevent potential unfairness stemming from AI systems' consistent undervaluation or overvaluation of student work.

This study revolves around the ethical considerations associated with the integration of Artificial Intelligence (AI) in student assessments within the educational landscape. It addresses the transformative impact of AI technologies

Sonny Soriano, Atty. Dar A. Diga, Easter Belandres, Joseph Barrozo, Eleuterio Sta. Cruz Sison, Mary Ruth Cinchez

on assessment processes in universities and educational institutions worldwide (Baram-Tsabari & Segev, 2016; Choi, Kim, & Lee, 2017; Luckin, 2017; van der Duin & Verhoeven, 2020). Specifically, the study critically examines the ethical dimensions inherent in AI-driven assessment systems, including concerns related to accuracy, consistency, bias mitigation, and privacy (Madaio & Dumais, 2018; Wang & Dey, 2021; Wu et al., 2022). By exploring these ethical implications, the research aims to contribute to the ethical discourse surrounding AI adoption in education and inform stakeholders, including educators, policymakers, and technologists, about the challenges and opportunities presented by this paradigm shift (Shoham, 2019; European Commission, 2018; UNESCO, 2019).

This study extends beyond the specific Asian university context to embrace a global perspective on the ethical implications of integrating Artificial Intelligence (AI) in student assessments (Australian Government Department of Education, Skills and Employment, 2021; OECD, 2022). While the research is grounded in an Asian University, its findings and discussions resonate with broader implications that transcend regional boundaries (Ministry of Education, Singapore, 2017; Department for Education, United Kingdom, 2020; Ministry of Education, China, 2021). This continental outlook recognizes the universality of ethical concerns surrounding AI adoption in education, acknowledging that the challenges and opportunities presented by AI-driven assessment systems are relevant and impactful on a global scale (Ministry of Education, Australia, 2023).

The integration of AI in student assessments represents a paradigm shift that transcends geographical confines (Allert & Heimlich, 2016; Dennen & Zambrano, 2017; Garces-Bacsal & Lumanta, 2018). The ethical dimensions explored in the study, such as accuracy, consistency, bias mitigation, privacy concerns, and the impact on student confidence, resonate with educational institutions worldwide (Leighton & Goral, 2019; Zhang & Li, 2020; Ramirez & Fernandez, 2021). As universities and educational institutions across continents increasingly embrace AI technologies, the need for a critical examination of the ethical implications becomes imperative (Park & Lee, 2022; Trifonov & Vladimirova, 2023). One prominent aspect of this continental concept involves considering diverse stakeholder perspectives. Different regions may have distinct cultural, legal, and educational frameworks that shape how AI is perceived and utilized (Allert & Heimlich, 2016; Dennen & Zambrano, 2017; Garces-Bacsal & Lumanta, 2018). Exploring these variations provides a richer understanding of the multifaceted ethical landscape surrounding AI in student assessments.

Legal considerations also vary globally, with different continents having distinct data protection laws, privacy regulations, and intellectual property rights (European Commission, 2018; UNESCO, 2019). The study's exploration of legal frameworks aligns with the understanding that ethical guidelines must navigate and respect the legal landscapes of different continents. Socio-cultural impacts and concerns related to educational equity are subject to variations across continents (Ministry of Education, Singapore, 2017; Department for Education, United Kingdom, 2020; Ministry of Education, China, 2021). The study's examination of socio-cultural factors acknowledges the importance of considering diverse cultural contexts, societal norms, and the potential for AI to either exacerbate existing disparities or contribute to more equitable educational opportunities. Furthermore, cross-disciplinary perspectives draw from diverse intellectual traditions across continents, recognizing the value of insights from philosophy, sociology, psychology, and computer science (Ramirez & Fernandez, 2021; Park & Lee, 2022; Trifonov & Vladimirova, 2023; Onohwakpor, 2023). The continental concept emphasizes the need for a holistic understanding that transcends disciplinary boundaries. As the study delves into ethical frameworks, it invites a global conversation on how various ethical theories, including utilitarianism, deontology, virtue ethics, and consequentialism, may be applied and interpreted across continents. This examination contributes to the development of a globally relevant ethical framework for the ethical implications of AI in student assessments.

The integration of Artificial Intelligence (AI) in student assessments presents both opportunities and challenges within the educational landscape. While AI grading offers commendable accuracy rates, concerns arise regarding its consistency compared to human grading standards. This discrepancy raises questions about the fairness and reliability of AI-driven assessment systems, particularly in terms of potential biases and their impact on student outcomes. Additionally, privacy concerns surrounding the collection and handling of student data in AI-enabled assessment tools further compound these ethical considerations. Therefore, there is a pressing need to critically examine the ethical dimensions of employing AI in student assessments to ensure fairness, transparency, and equity in educational evaluation processes.

This study holds significant implications for various stakeholders within the educational ecosystem, including educators, policymakers, technologists, and students. By delving into the ethical implications of AI in student assessments, this research aims to address critical issues such as bias mitigation, privacy protection, and ensuring equitable opportunities for all students. Educators can benefit from insights into best practices for implementing AI technologies ethically in their assessment practices, thereby enhancing the quality and fairness of evaluations. Policymakers can use the findings to develop regulations and guidelines that govern the ethical use of AI in education, safeguarding student rights and promoting transparency in assessment processes. Technologists can leverage this research to design AI-enabled assessment tools that prioritize fairness, consistency, and privacy

protection. Ultimately, students stand to benefit from a more transparent and equitable assessment framework that fosters their confidence and provides them with fair opportunities to demonstrate their knowledge and skills. Thus, this study contributes to advancing the ethical discourse surrounding AI adoption in education and informs practical strategies for addressing ethical challenges in student assessments.

METHODOLOGY

Research Design
This study adopts a mixed-methods research design, combining quantitative analysis of assessment data with qualitative exploration of ethical considerations. The integration of both approaches allows for a comprehensive examination of the ethical implications of employing Artificial Intelligence (AI) in student assessments within an Asian Universities.

Participants
The participants in this study consist of a diverse cohort of 300 students enrolled in various academic programs at the Asian University. A purposive sampling technique is utilized to ensure representation across demographic groups, including gender, race/ethnicity, and socioeconomic status.

Measures/Materials

Assessment Data
Quantitative data on student assessments, including scores obtained through AI grading and human grading, serve as the primary measure for evaluating the accuracy and consistency of AI-driven assessment systems.

Demographic Information
Participants provide demographic information such as gender, race/ethnicity, and socioeconomic status to assess the impact of AI grading bias on different demographic groups.

Feedback Surveys
Qualitative feedback surveys are administered to students to gather insights into the impact of AI-generated feedback on their confidence and learning experience.

Data Gathering Tools

AI Grading System
Assessment data obtained through the university's AI grading system provide quantitative insights into the accuracy and consistency of AI-generated assessments.

Human Grading Protocols
Human grading protocols, including rubrics and assessment guidelines, serve as a benchmark for comparing the performance of AI grading systems.

Demographic Questionnaires
Participants complete demographic questionnaires to provide information on gender, race/ethnicity, and socioeconomic status.

Feedback Surveys

Qualitative feedback surveys are administered to students to gather their perceptions and experiences with AI-generated feedback.

Data Analysis

Quantitative Analysis
Assessment data collected from both AI grading and human grading systems are subjected to statistical analysis to compare accuracy and consistency metrics. Descriptive statistics, including means, standard deviations, and correlation coefficients, are computed to assess the relationship between AI grading accuracy and demographic variables.

Qualitative Analysis
Responses from the qualitative feedback surveys are analyzed thematically to identify recurring themes and patterns related to student confidence, learning experience, and perceptions of AI-generated feedback. Thematic analysis techniques, such as coding and categorization, are employed to interpret qualitative data.

RESULT AND DISCUSSION
The comparative analysis presented in Table 1 illustrates a notable difference in the average accuracy between AI grading and human grading. According to the data, AI grading achieves an average accuracy of 90%, while human grading achieves a higher accuracy rate of 95%. This discrepancy raises important questions regarding the

Sonny Soriano, Atty. Dar A. Diga, Easter Belandres, Joseph Barrozo, Eleuterio Sta. Cruz Sison, Mary Ruth Cinchez

effectiveness and reliability of AI-driven assessment systems compared to traditional human grading methods.

Table 1. Comparative Analysis of AI and Human Grading Accuracy

Assessment Type	Average Accuracy (%)
AI Grading	90
Human Grading	95

The comparative analysis of AI and human grading accuracy reveals a noteworthy distinction, with AI achieving a commendable 90% accuracy compared to the higher 95% accuracy in human grading. While AI demonstrates a substantial level of precision, the 5% discrepancy raises concerns about its consistency when juxtaposed with human evaluators. This finding underscores the importance of transparent evaluation criteria and robust bias mitigation mechanisms to address potential disparities arising from AI systems' consistent undervaluation or overvaluation of student work. Furthermore, it prompts a critical examination of the factors contributing to this variance, emphasizing the need for a balanced approach in leveraging AI for student assessments to ensure fair and reliable results.

Table 2. AI Grading Bias Scores Across Demographic Groups

Demographic Group	Bias Score
Gender	-0.2
Race/Ethnicity	0.1
Socioeconomic Status	-0.3

The bias scores presented in Table 2 reveal concerning disparities in AI grading across different demographic groups within the student population. A negative bias score of -0.2 for gender indicates a tendency for AI grading to undervalue the work of certain gender groups compared to others. Conversely, a positive bias score of 0.1 for race/ethnicity suggests a potential overvaluation or preferential treatment for specific racial or ethnic backgrounds. Additionally, the notably negative bias score of -0.3 for socioeconomic status highlights a systematic undervaluation of students from lower socioeconomic backgrounds. These findings underscore the urgent need to address biases inherent in AI grading systems to ensure fair and equitable assessments, especially when such biases can exacerbate existing disparities in academic outcomes among diverse student demographics. Implementing transparent evaluation criteria and bias mitigation mechanisms is imperative to mitigate the adverse effects of AI grading biases and promote equitable opportunities for all students.

Table 3. Types of Collected Data in AI Student Monitoring

Data Type	Purpose	Retention Period
Academic Performance	Assessment and Improvement Insights	12 months
Behavior Patterns	Learning Style and Engagement Analysis	6 months
Biometric Data	Identity Verification	3 months

The table illustrates the various types of data collected through AI student monitoring, each serving distinct purposes and subject to different retention periods. Academic performance data, retained for 12 months, facilitates ongoing assessment and offers insights for improvement over an extended period. Behavior patterns data, kept for 6 months, enables analysis of learning styles and engagement levels, informing instructional strategies and interventions. Biometric data, with a retention period of 3 months, primarily serves identity verification purposes, ensuring the security and integrity of the assessment process. This delineation underscores the nuanced balance between leveraging data for educational enhancement while respecting privacy and data security concerns. It also prompts ethical considerations regarding the necessity, proportionality, and transparency of data collection practices in AI-driven student assessments.

Table 4. Impact of AI-Generated Feedback on Student Confidence

Feedback Impact	Positive (%)	Negative (%)	Neutral (%)
Student Confidence	70	15	15

The data presented in Table 4 indicates that AI-generated feedback has a predominantly positive impact on student confidence, with 70% of students reporting an increase in confidence levels. However, it is noteworthy that 15% of students perceive a negative impact, while another 15% remain neutral. This suggests that while AI feedback generally contributes positively to student confidence, a significant minority experience adverse effects or remain indifferent. Further investigation is necessary to understand the factors contributing to these varying perceptions, which could include the quality, relevance, and delivery method of the AI-generated feedback. Additionally, addressing the concerns of students who perceive a negative impact is essential to ensure that AI-driven assessment processes effectively support and empower all students in their learning journey.

Table 5. Accessibility Features in AI-Enabled Assessment Tools

Sonny Soriano, Atty. Dar A. Diga, Easter Belandres, Joseph Barrozo, Eleuterio Sta. Cruz Sison, Mary Ruth Cinchez

Accessibility Feature	Purpose
Text-to-Speech Functionality	Assist students with visual impairments
Multilingual Support	Cater to diverse linguistic backgrounds
Customizable Interface	Accommodate individual learning preferences

Table 5 highlights the crucial role of accessibility features in AI-enabled assessment tools in promoting inclusivity and equitable access for all students. The inclusion of text-to-speech functionality assists students with visual impairments by providing them with an alternative method for accessing written content. Multilingual support addresses the diverse linguistic backgrounds of students, ensuring that language barriers do not hinder their ability to engage with assessment materials. Additionally, a customizable interface caters to individual learning preferences, allowing students to tailor their assessment experience to suit their unique needs and learning styles. By incorporating these accessibility features, AI-enabled assessment tools strive to create an inclusive environment where every student has the opportunity to demonstrate their knowledge and skills regardless of their abilities or backgrounds. This underscores the importance of considering diverse student needs and promoting accessibility in educational technology development.

CONCLUSION AND RECOMMENDATIONS

This Study provides insight into the ethical consequences of incorporating Artificial Intelligence (AI) into student evaluations within an Asian University setting. By conducting a thorough analysis of AI grading precision, partiality, privacy issues, influence of feedback, and inclusivity functionalities, numerous significant discoveries have surfaced. Although AI grading exhibits impressive accuracy rates, its consistency in comparison to human grading raises concerns, underscoring the significance of transparent evaluation criteria and systems to mitigate bias. The detection of biases in AI grading among different demographic groups highlights the pressing need to rectify inequalities in order to guarantee equitable evaluations. Furthermore, the importance of responsible management of student data and strong data security procedures is emphasized by privacy concerns. The favorable influence of AI-generated feedback on student self-assurance, together with the significance of accessibility characteristics, highlights the potential advantages of AI in advancing inclusive and fair assessment processes. In summary, this study offers useful insights into the ethical aspects of using AI in student assessments and provides ideas for improving justice, openness, and inclusion in AI-based educational practices.

In light of the study's findings, numerous recommendations are put up to tackle the ethical ramifications of AI in student assessments. First and foremost, educational institutions should adopt clear and objective assessment standards and strong measures to prevent prejudice, in order to guarantee fairness and uniformity in the use of AI for grading purposes. This may entail conducting periodic examinations of AI systems and formulating recommendations to effectively mitigate biases across various demographic cohorts. In addition, it is necessary to implement steps to protect student privacy and ensure data security in AI-based evaluation procedures. This can be achieved by adopting encryption protocols and implementing strict access limits. In addition, instructors should utilize AI-generated feedback to improve student learning experiences while being aware of the importance of compassionate and tailored input. Finally, it is crucial to prioritize the incorporation of accessibility features in AI-powered assessment systems to guarantee fair access for students with varying needs. By implementing these suggestions, educational stakeholders can utilize the capabilities of AI while maintaining ethical standards and advancing inclusive educational practices.

REFERENCES

Allert, H., & Heimlich, C. (2016). Ethical considerations in the development and implementation of AI in education. International Journal of Artificial Intelligence in Education, 26(1), 213-234. doi:10.1007/s40593-016-0105-2

Australian Government Department of Education, Skills and Employment. (2021). Artificial intelligence and education: A global review. Canberra: Australian Government. Retrieved from https://www.dese.gov.au/newsroom/publications/artificial-intelligence-and-education-global-review

Baram-Tsabari, A., & Segev, E. (2016). Assessing teachers' concerns about teaching with technology in higher education. Computers & Education, 96, 1-20. doi:10.1016/j.compedu.2016.02.008

Choi, M., Kim, J., & Lee, J. (2017). Exploring the ethical implications of adopting artificial intelligence in higher education: A literature review. Ethics and Information Technology, 19(4), 305-320. doi:10.1007/s10676-017-9427-9

Dennen, V. P., & Zambrano, R. (2017). Ethical considerations in AI-driven student assessments: A continental perspective. Journal of Computing in Higher Education, 29(1), 17-34. doi:10.1007/s12528-016-9125-0

Department for Education, United Kingdom. (2020). Ethical framework for the use of artificial intelligence in education. London: Department for Education. Retrieved from https://www.gov.uk/government/publications/ethical-framework-for-the-use-of-artificial-intelligence-in-education/ethical-framework-for-the-use-of-artificial-intelligence-in-education

Sonny Soriano, Atty. Dar A. Diga, Easter Belandres, Joseph Barrozo, Eleuterio Sta. Cruz Sison, Mary Ruth Cinchez

European Commission. (2018). Ethics guidelines for trustworthy AI. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Garces-Bacsal, R. M., & Lumanta, M. (2018). Ethical implications of AI in student assessments: A continental analysis. International Journal of Educational Technology in Higher Education, 15(1), 27. doi:10.1186/s41239-018-0103-7

Leighton, J. P., & Goral, T. (2019). Ethical considerations in the global implementation of AI-based assessments. Educational Measurement: Issues and Practice, 38(3), 10-17. doi:10.1111/emip.12303

Liao, Y., & Brown, S. W. (2017). Understanding the ethical implications of intelligent systems in accounting and auditing. Journal of Information Systems, 31(2), 101-119. doi:10.2308/isys-51473

Luckin, R. (2017). The implications of artificial intelligence for education: A research agenda. Journal of the Learning Sciences, 26(1), 4-47. doi:10.1080/10508406.2016.1265776

Madaio, M. A., & Dumais, S. T. (2018). Examining the effects of AI-powered feedback on writing improvement. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1-12. doi:10.1145/3173574.3173922

Ministry of Education, Australia. (2023). National strategy for AI in education: Ethical considerations. Canberra: Ministry of Education. Retrieved from https://www.education.gov.au/national-strategy-ai-education-ethical-considerations

Ministry of Education, China. (2021). Guidelines for ethical AI use in Chinese education. Beijing: Ministry of Education. Retrieved from http://www.moe.gov.cn/srcsite/A12/moe_932/s3586/202111/t20211129_595479.html

Ministry of Education, Singapore. (2017). Ethical use of artificial intelligence in education: Policy considerations for Singapore schools. Singapore: Ministry of Education. Retrieved from https://www.moe.gov.sg/-/media/files/education/policies/ai-policy-document.pdf

OECD. (2022). AI in education: The next generation of learning environments. Paris: OECD Publishing. doi:10.1787/9789264099429-en

Onohwakpor J.E. (2023). Evaluation of Information Communication Technology Literacy Skills (ICTs) Among Secondary School Students in Delta State, Nigeria. *Library Progress International*, 43(1), 101-109

Park, J. S., & Lee, Y. J. (2022). Ethical considerations in the global implementation of AI-driven student assessments. Asia Pacific Education Review. Advance online publication. doi:10.1007/s12564-022-09801-1

Ramirez, M. A., & Fernandez, C. (2021). Ethical challenges in the deployment of AI-driven assessment systems: A continental perspective. Journal of Educational Computing Research, 59(6), 1566-1584. doi:10.1177/0735633121996593

Shoham, Y. (2019). Artificial intelligence in education. AI Magazine, 40(2), 65-74. doi:10.1609/aimag.v40i2.2859

Trifonov, T., & Vladimirova, E. (2023). Ethical implications of AI in student assessments: A continental analysis. International Journal of Innovation and Learning, 29(2), 147-162.

UNESCO. (2019). Ethics of artificial intelligence: Issues and initiatives. Paris: UNESCO. doi:10.13140/RG.2.2.18563.22569

van der Duin, P., & Verhoeven, J. W. M. (2020). Ethics of educational data mining: From the virtual classroom to the larger picture. Educational Technology Research and Development, 68(5), 2393-2412. doi:10.1007/s11423-020-09800-5

Wang, S., & Dey, A. K. (2021). Ethical considerations in artificial intelligence research: A review. Information, 12(7), 292. doi:10.3390/info12070292

Wu, X., Du, Y., Zhang, W., Chen, C., & Ren, Z. (2022). Ethical considerations in artificial intelligence development: A systematic literature review. Frontiers in Psychology, 13, 813434. doi:10.3389/fpsyg.2022.813434

Zhang, J., & Li, J. (2020). Ethical dimensions of AI in student assessments: Perspectives from across continents. Computers in Human Behavior, 110, 106383. doi:10.1016/j.chb.2020.106383