American Journal of Applied Research and AI (AJARAI)

VOLUME 1 ISSUE 1 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of
Applied Research and AI (AJARAI))

Volume 1 Issue 1, Year 2026
https://journals.e-palli.com/home/index.php/ajarai

Trust and Ethics for Artificial Intelligence in Authentic Education

Clark Dominic L. Alipasa[1*]

Article Information

Received: November 20, 2025
Accepted: January 03, 2026
Published: January 22, 2026

Keywords

AI Ethics, AI Integration, Artificial Intelligence, Authentic Assessment, Educational Leadership, Fairness, Higher Education, Human Oversight, Liberal Arts, Philippine Higher Education, Transparency, Trustworthiness

ABSTRACT

The integration of Artificial Intelligence (AI) into authentic assessment has generated renewed ethical concerns in higher education. This study examined teachers' and students' perceptions of the ethical dimensions of AI tools used in authentic assessment, focusing on trustworthiness, transparency, fairness, and human oversight. Guided by the Ethics of AI in Education framework proposed by Holmes *et al.* (2022), the study employed a quantitative descriptive design using a Likert-scale survey administered to faculty members and students. Descriptive statistics, including means and standard deviations, were used to analyze perceptions across ethical domains. Results indicate that both groups perceived the ethical use of AI in authentic assessment at a moderate extent, with faculty members consistently reporting higher levels of trust, fairness, and confidence in human oversight than students. Students expressed more cautious views, particularly regarding the trustworthiness and transparency of AI-generated feedback and scoring. The findings suggest that while AI is generally accepted as a support tool in assessment, ethical concerns persist, underscoring the need for clearer transparency mechanisms, strengthened fairness safeguards, and visible human oversight. The study contributes empirical evidence to ethical AI discourse in education and offers insights for higher education institutions seeking to integrate AI into authentic assessment while preserving trust, integrity, and pedagogical authenticity.

INTRODUCTION

Artificial Intelligence (AI) is increasingly embedded in higher education, particularly in authentic assessment practices such as portfolios, essays, case analyses, and project-based tasks. In Liberal Arts education, these assessment forms are critical for developing creativity, critical thinking, ethical reasoning, and communication skills (Roth, 2017). The growing use of generative AI tools for drafting, feedback, and presentation support has enhanced learning efficiency but has also raised concerns regarding authorship, assessment validity, and academic integrity. This emphasis on stakeholder perceptions is consistent with Philippine educational research demonstrating that users' trust and experiences are critical to the effective implementation of technology-mediated learning systems (Alipasa, 2026).

As AI mediates student work, questions emerge about whether assessment outcomes remain trustworthy, transparent, and fair, and whether human judgment continues to guide evaluative decisions. While international policy frameworks advocate ethical and human-centered AI use in education (UNESCO, 2023), they often provide limited insight into how ethical principles are experienced by teachers and students within everyday assessment contexts.

In the Philippine higher education setting, institutions have begun adopting AI-related policies and exploring AI-supported learning. However, there is limited empirical evidence on how key stakeholders perceive the ethical dimensions of AI use in authentic assessment, particularly within higher education programs where human expression and judgment are central.

This study is grounded in the Ethics of AI in Education framework by Holmes *et al.* (2022), which conceptualizes ethical AI integration through four dimensions: trustworthiness, transparency and explainability, fairness and non-discrimination, and human oversight and co-creation. Using this framework, the study addresses the central research question: How do teachers and students perceive the trustworthiness, transparency, fairness, and human oversight of AI tools used in authentic assessment?

By examining stakeholder perceptions, the study contributes empirical evidence to discussions on ethical AI in education and provides practical insights for higher education institutions seeking to integrate AI into authentic assessment while preserving trust, fairness, and meaningful human oversight.

LITERATURE REVIEW

Recent international frameworks from UNESCO and the OECD provide essential foundations for the integration of Artificial Intelligence (AI) in education, offering guidance that is crucial to the development of context-sensitive and ethically grounded guidelines for AI-supported authentic assessment. UNESCO's AI Competency Frameworks for Students and Teachers (2024) outline a comprehensive set of competencies across dimensions such as a human-centred mindset, ethics of AI, AI techniques and applications, and AI system design. These

[1] University of Perpetual Help System DALTA-Molino, Bacoor, Cavite, Philippines
[*] Corresponding author's email: clarkdominic22@gmail.com

frameworks emphasize a progression from understanding to application and creation, aiming to equip both learners and educators with the knowledge and ethical awareness necessary to engage responsibly with AI tools.

In parallel, the OECD's AI Capability Indicators (2025) present a structured approach to evaluating AI's capabilities in comparison to human competencies, particularly in areas such as critical thinking, creativity, and problem solving. This framework enables policymakers and educators to better understand the potential and limitations of AI integration, thereby informing curriculum design and authentic assessment strategies. Furthermore, the OECD AI Literacy Framework (AILit, 2025) underscores the importance of cultivating AI literacy among educators and learners, providing guidance for embedding AI competencies into learning materials, teaching practices, and institutional policies. Such comprehensive approaches promote transparency, fairness, and accountability in AI-powered education.

Existing literature underscores that trust and ethical practice are central to the effective integration of innovations in education, particularly when human judgment, transparency, and contextual sensitivity are involved. Qualitative research on women's decision-making in situations of intimate partner violence demonstrates how trust, perceived fairness, and institutional support influence critical choices, highlighting ethical dimensions of human-centered systems (Reyes *et al.,* 2022). In technology-mediated learning environments, parents' perceptions of virtual learning implementation reveal that transparency, accountability, and institutional readiness are crucial in fostering trust in digital systems (Alipasa, 2026). Contextualized instructional interventions further show that learning experiences grounded in learners' cultural and linguistic realities enhance authenticity and engagement, reinforcing ethical considerations in educational design (Kabigting & Alipasa, 2025). Similarly, phenomenological studies on bullying experiences among Filipino students emphasize the role of fairness, care, and supportive oversight in shaping learners' sense of safety and trust within educational spaces (Alipasa & Alipasa, 2025). Collectively, these studies provide a conceptual foundation for examining trust, ethics, and human oversight in AI-supported authentic education, emphasizing that technological adoption must remain aligned with transparency, fairness, and human-centered values.

Building on these international frameworks and the AI passion-driven pedagogy in enhancing student engagement and learning outcomes of Prestoza and Banatao (2024), this study addresses significant research gaps in the integration of AI in authentic assessment, particularly within the Philippine higher education context. While UNESCO and OECD provide comprehensive principles and competency models, there remains a lack of localized, evidence-based guidelines that operationalize these frameworks specifically for authentic assessment practices in higher education education. Moreover, there is limited research on how instructional leadership can support such operationalization. By developing context-sensitive guidelines grounded in these global frameworks and informed by leadership practices that promote 21st-century skills, this study seeks to fill this gap. The resulting framework will serve as a structured, ethically informed, and practically applicable roadmap for AI integration that enhances authenticity, fairness, transparency, and trust in higher education assessment practices.

MATERIALS AND METHODS

This study employed a quantitative descriptive design to examine teachers' and students' perceptions of the ethical dimensions of AI tools used in authentic assessment. Guided by the Ethics of AI in Education framework by Holmes *et al.* (2022), the survey instrument measured four core domains: trustworthiness, transparency and explainability, fairness and non-discrimination, and human oversight and co-creation. An additional item examined perceptions of AI's role in maintaining academic integrity, as this concern is closely associated with trust and oversight in assessment contexts.

Respondents rated five statements using a five-point Likert scale ranging from 1 (least extent) to 5 (greatest extent). Mean scores and standard deviations were computed separately for students and faculty members. The interpretation of results followed the established scale descriptors: least extent (1.00–1.80), limited extent (1.81–2.60), moderate extent (2.61–3.40), considerable extent (3.41–4.20), and greatest extent (4.21–5.00). The analysis focused on identifying patterns and differences in perceptions between the two groups across the ethical domains.

RESULTS AND DISCUSSION

The results illustrate the perceptions of students and faculty members regarding the trustworthiness, transparency, fairness, and human oversight of AI tools used in authentic assessment. Across all domains, results indicate ratings within the moderate extent, with faculty members consistently reporting slightly higher mean

Table 1: Artificial intelligence in authentic assessments

Domains	Students		Faculty Members		Interpretation
	Mean	SD	Mean	SD	
I trust AI-generated feedback and scoring in authentic assessments.	2.59	0.745	2.8	0.919	LIE and ME
AI tools provide transparent explanations for grading and feedback.	2.73	0.681	2.8	1.033	ME and LIE

AI-assisted assessments are fair across different student abilities and contexts.	2.67	0.730	2.9	0.994	ME
Human oversight is adequately integrated in AI-assisted assessment processes.	2.87	0.662	3.1	0.994	ME
AI tools help maintain academic integrity in assessments.	2.64	0.776	2.9	0.658	ME

Legend: 1.00–1.80 – Least extent (LE), 1.81–2.60 – Limited extent (LIE), 2.61–3.40 – Moderate extent (ME), 3.41–4.20 – Considerable extent (CE), 4.21–5.00 – Greatest extent (GE)

scores than students. This pattern suggests cautious acceptance of AI-supported assessment, accompanied by reservations about its ethical implications.

With respect to trustworthiness, students reported a mean score of 2.59, interpreted at a limited extent, while faculty members reported a higher mean of 2.80, corresponding to a moderate extent. This discrepancy reflects students' hesitancy to fully trust AI-generated feedback and scoring, particularly in assessments that carry evaluative consequences. Similar patterns of cautious trust have been observed in earlier studies examining stakeholder experiences in digitally mediated education, where limited understanding of technological decision-making processes contributed to uncertainty and skepticism (Alipasa *et al.,* 2021). Faculty members' comparatively higher trust may stem from their familiarity with assessment criteria and their role in supervising AI use, reinforcing Holmes *et al.*'s (2022) assertion that trust is shaped by both control and contextual understanding.

In terms of transparency and explainability, students (M = 2.73) and faculty members (M = 2.80) both rated AI tools at a moderate extent. These findings suggest that while AI systems provide some explanatory feedback, such explanations may be insufficiently clear or pedagogically meaningful. Transparency is particularly crucial in higher education, where assessment often involves interpretive judgment rather than fixed answers. Prior research on curriculum and assessment in the Philippine context emphasizes the importance of clarity, language sensitivity, and contextual alignment in evaluation processes (Alipasa, 2016). Limited transparency in AI-supported assessments may therefore undermine students' ability to understand and act on feedback effectively.

Perceptions of fairness likewise fell within the moderate range for both students (M = 2.67) and faculty members (M = 2.90). While respondents did not view AI-assisted assessments as overtly biased, the absence of higher ratings indicates lingering concerns about equity across diverse student abilities and learning contexts. This concern mirrors broader findings in Philippine education research, which highlight structural and contextual factors that shape learners' authentic experiences in technology-enhanced environments (Alipasa, 2021). From the perspective of Holmes *et al.* (2022), fairness in AI systems requires continuous human monitoring to ensure that algorithmic outputs do not disadvantage particular groups of learners.

The highest mean scores were observed in human oversight, with students rating this dimension at a moderate extent (M = 2.87) and faculty members reporting the strongest agreement (M = 3.10). These results indicate a shared perception that educators remain actively involved in AI-assisted assessment processes. The prominence of human oversight aligns with ethical expectations in educational leadership, particularly in contexts that prioritize values formation, reflective judgment, and learner development (Gatdula *et al.,* 2021). Faculty members' confidence in this domain underscores the importance of maintaining educator authority and accountability when integrating AI into assessment practices.

Finally, perceptions of AI's role in maintaining academic integrity were rated at a moderate extent by both students (M = 2.64) and faculty members (M = 2.90). While AI is seen as having potential to support integrity—through feedback consistency or plagiarism detection—respondents expressed uncertainty about its effectiveness in preventing misuse. This tension reflects broader debates in higher education, where AI simultaneously functions as a learning aid and a challenge to originality and authentic performance. As noted in studies on student preparedness and professional readiness, integrity and ethical judgment remain human competencies that cannot be fully delegated to technological systems without The teachers' readiness and administrators' support for the implementation (Alipasa, 2021).

The experiences of students and teachers with authentic performance tasks, such as those in the GRASPS model, and insights from parents on virtual learning implementation highlight the need for clear guidance, stakeholder engagement, and ethical oversight to ensure AI-supported assessments are trustworthy, transparent, and fair (Alipasa, 2020; Alipasa, 2026). Overall, the findings reveal a pattern of guarded acceptance. Faculty members appear more confident in the ethical deployment of AI, while students express more restrained trust, particularly in high-stakes assessment contexts. This divergence highlights the need for clearer communication, shared ethical norms, and institutional transparency to bridge perception gaps.

CONCLUSIONS

This study examined teachers' and students' perceptions of the trustworthiness, transparency, fairness, and human oversight of AI tools used in authentic assessment, grounded in the Ethics of AI in Education framework proposed by Holmes *et al.* (2022). Findings revealed that the ethical acceptability of AI integration in authentic

assessment is perceived at a moderate level overall, with faculty members consistently expressing greater confidence than students across all ethical dimensions. This disparity highlights the need to address student concerns and strengthen shared understanding of AI-supported assessment practices.

The results emphasize the importance of enhancing transparency mechanisms, reinforcing fairness safeguards, and making human oversight more explicit in AI-assisted assessment processes. Trust in AI cannot be presumed; rather, it must be deliberately cultivated through clear institutional policies, explicit disclosure of AI use, and sustained human involvement in evaluative decision-making. In the context of Liberal Arts education, where authenticity, creativity, and ethical judgment are central, AI should serve as a supportive instructional tool rather than an authoritative evaluator.

Based on these findings, higher education institutions should prioritize ethical capacity-building alongside technological adoption. Faculty training, student orientation on responsible AI use, reflective learning activities, and continuous evaluation of AI tools are essential to ensuring that AI-supported assessments uphold trust, integrity, and authentic learning outcomes.

REFERENCES

Alipasa, C. D. (2016). *Issues in the Philippine English curriculum today.* De La Salle University.

Alipasa, C. D. (2020). Challenges in implementing the Goal, Role, Audience, Situation, Product/ Performance, and Standards (GRASPS) model in selected schools in Cavite. *Academia Lasalliana Journal of Education and Humanities, 1*(2).

Alipasa, C. D. (2026). The parents' perception on the extent of implementation of virtual learning: Basis for a faculty eLearning training program. *Perspectives in Sustainable Development Studies, 1*(1), 1–11. https://psds-journal.com/publications/642700/the-parents-perception-on-the-extent-of-implementation-of-virtual-learning-basis

Alipasa, J. B., & Alipasa, C. D. (2025). *Bullying experiences and coping mechanisms of Filipino high school students: A phenomenological inquiry.* Research Square.

Alipasa, C. D., Baroja, M., Gutierrez, C., Ramirez, M. T., & Sulse, L. (2021). The experiences of the parents involved in their children's online distance learning during the COVID-19 pandemic: A phenomenological study. *Psychology and Education Journal, 58*(4), 1807–1817.

Gatdula, G. M., Ramirez, M. T., Regis, M. L., Domingo, J., Jr., & Sebastian, R. S. (2021). Non-Catholic Lasallians' experiences in living out the Lasallian values. *Turkish Online Journal of Qualitative Inquiry, 12*(7).

Holmes, W., Bialik, M., Fadel, C., & Vossoughi, S. (2022). *Ethics of AI in education: A framework for trust, transparency, fairness, and human oversight.* Publisher.

Kabigting, M. J., & Alipasa, C. D. (2025). Enhancing the English language proficiency of Capampangan pupils in a public school: Basis for an instructional intervention. *International Journal of Research and Innovation in Social Science.*

Organisation for Economic Co-operation and Development. (2025). *AI capability indicators.* OECD.

Organisation for Economic Co-operation and Development. (2025). *AI literacy framework (AILit).* OECD.

Prestoza, A., & Banatao, J. (2024). AI passion-driven pedagogy in enhancing student engagement and learning outcomes.

Reyes, A., Hijalda, L., Gatdula, G., Ramirez, M. T., & Alipasa, C. D. (2022). LIVE or LEAVE: Processes Caviteña women undergo in intimate partner violence. *Journal of Positive School Psychology.*

Roth, J. (2017). *Authentic assessment in liberal arts education.* Publisher.

UNESCO. (2023). *AI competency frameworks for students and teachers.* United Nations Educational, Scientific and Cultural Organization.