Journal of Transdisciplinary Studies in Education
2025, VOL. 1, NO. 2, 81-90
https://doi.org/10.64268/jtse.v1i2.58

Rethinking Ethical Responsibility and Data Governance in Academic Assessment Using Large Language Models

Ruri Supatmi[1*], Diyah Dwi Agustina[2], Rangga Mega Putra[3], Asti Nur Cahyani[4]
[1,2,3,4] Universitas Nahdlatul Ulama Lampung, Lampung, Indonesia.

rurifadli21@gmail.com[*]

ABSTRACT

Background: The integration of Large Language Models (LLMs) into academic grading practices has expanded rapidly in higher education, driven by demands for efficiency and consistency.

Aims: In response to these concerns, this study seeks to explore issues of ethical accountability and data governance in the use of LLMs for academic assessment, drawing on the perspectives of lecturers, students, and academic administrators.

Methods: The study adopted a qualitative exploratory approach to capture in-depth insights into current assessment practices involving LLMs. Data were gathered through semi-structured interviews, institutional document analysis, and direct observations across selected higher education institutions. Analysis followed the interactive framework proposed by Miles, Huberman, and Saldaña, involving iterative processes of data reduction, data display, and conclusion verification, with triangulation applied to strengthen trustworthiness.

Result: The findings demonstrate a set of interrelated challenges. The involvement of LLMs in grading processes often obscures responsibility for assessment decisions, particularly when transparency is limited. Concerns regarding fairness and potential bias persist, especially in evaluating varied linguistic and contextual expressions. At the same time, data governance mechanisms remain insufficiently developed, with unclear procedures for consent, data storage, and regulatory compliance. These issues collectively reflect uneven institutional preparedness and weak ethical oversight**.**

Conclusion: The study concludes that the use of LLMs in academic grading requires clearly defined ethical accountability and comprehensive data governance frameworks. Continued human oversight, supported by institutional policies and capacity-building initiatives, is essential to safeguard academic integrity and ensure responsible adoption of AI-assisted assessment in higher education.

ARTICLE HISTORY
Submitted: September 12, 2025
Accepted: October 01, 2025
Published: December 26, 2025

KEYWORD
Academic Grading;
Data Governance;
Ethical Accountability;
Higher Education Assessment;
Large Language Models;

Introduction

The rapid diffusion of artificial intelligence into higher education has reached a point where its influence on academic assessment can no longer be ignored (Katsamakas et al., 2024; O & O, 2023). Grading is a central academic practice that shapes students' learning trajectories, institutional credibility, and public trust in higher education (Gonsalves & Lin, 2025; Lim, 2024). As Large Language Models begin to participate in this process, the absence of sustained ethical inquiry creates an urgent need to examine how responsibility, fairness, and data protection are being redefined within contemporary assessment practices.

Large Language Models are increasingly adopted because they promise speed, consistency, and scalability in evaluating written work. For many institutions facing growing student numbers and limited instructional resources, AI-assisted grading appears to offer a pragmatic response to structural pressures. However, efficiency-driven adoption often precedes critical reflection, allowing technological convenience to reshape assessment practices without sufficient consideration of their ethical and institutional consequences (Sebestyén, 2025; Toscani, 2025)

Academic assessment has traditionally relied on human judgment grounded in disciplinary expertise and pedagogical values (Brown, 2022; Valentine et al., 2021). Lecturers are expected to interpret not only correctness but also reasoning, originality, and contextual meaning. The introduction of LLMs into grading processes subtly alters this relationship by inserting algorithmic mediation into evaluative decisions (Gazit, 2025; Zou et al., 2025). This shift raises questions about whether academic judgment remains a human responsibility or becomes partially delegated to computational systems.

Ethical responsibility emerges as a central concern in this context (Meliou et al., 2021; Torelli, 2020). When grades are informed by AI-generated evaluations, responsibility for assessment outcomes becomes fragmented across lecturers, institutions, and technology providers. In cases of error, bias, or contested results, it is often unclear who should be held accountable (de Bruijn et al., 2022; Fazelpour & Danks, 2021). Such ambiguity challenges long-standing principles of academic responsibility and weakens the transparency of assessment practices.

Concerns about fairness and transparency further complicate the use of LLMs in grading (Kooli, 2023; Yan et al., 2024). These models are trained on large-scale textual data that reflect dominant linguistic norms and academic conventions. As a result, students who employ diverse linguistic styles or culturally situated forms of expression may be evaluated through opaque criteria that are difficult to interpret or challenge (Abfalter et al., 2021; Yi et al., 2021). The limited explainability of algorithmic judgments exacerbates this problem, reducing students' ability to understand how their work is assessed.

Alongside ethical concerns, the governance of student data represents a critical dimension of AI-assisted assessment (Elshall & Badir, 2025). The use of LLMs often involves transferring student assignments to external platforms, where data processing practices are not always visible to users. Issues of informed consent, data retention, and secondary use of academic work raise serious questions about privacy and compliance with data protection principles, particularly when institutional oversight is weak.

Institutional readiness significantly shapes how these ethical and data-related challenges unfold (Pantiris et al., 2025; Verhulst, 2021). While universities increasingly encourage innovation in digital assessment, many lack explicit policies regulating the use of LLMs in grading (Agostini & Picasso, 2024; Diab Idris et al., 2024). In practice, adoption often occurs at the individual level, driven by lecturers' initiative rather than institutional strategy. This fragmented approach limits ethical consistency and leaves educators without adequate guidance or professional support.

Although scholarly discussions on artificial intelligence in education continue to expand (Bearman et al., 2023; Pham & Sampson, 2022), much of the existing literature remains focused on technological capabilities or pedagogical efficiency. Empirical research that examines ethical responsibility and data governance in real assessment contexts remains limited (König, 2021; Verhulst, 2021). This gap underscores the importance of investigating how key stakeholders experience, interpret, and negotiate the use of LLMs in academic grading, ensuring that technological innovation does not outpace the ethical foundations of higher education.

Research on academic assessment highlights grading as a practice shaped by professional judgment and pedagogical values. Rojas Bruna (2025) shows how academic portfolios influence evaluative norms in mathematics teacher training, while Kirmizi (2025) demonstrates the role of teacher expectations in shaping participation and assessment dynamics. Biton (2025) further emphasizes that peer assessment criteria promote reflective learning for both students and educators, reinforcing the human-centered nature of grading. In mathematics education, fairness and interpretive complexity remain critical concerns. Srđan et al. (2025) reveal how Simpson's paradox can distort interpretations of grading outcomes, whereas Uhing et al. (2025) stress that standards based grading requires sustained human oversight. Hahn (2025) explores the use of ChatGPT for grading algebra problems, focusing on technical feasibility while leaving ethical accountability and data governance largely unaddressed. Broader educational studies provide important contextual insights. Estipona & Delos Santos (2025) link assessment outcomes to student wellbeing, while Kihaga et al. (2025) highlight structural constraints affecting fair assessment in large mathematics classes. Studies beyond grading contexts, such as curriculum comparisons by Cilsalar-Sagnak et al. (2025) and the scoping review by Sprague et al. (2025), underscore that educational outcomes are shaped by systemic and environmental factors. However, empirical research that integrates ethical responsibility, data governance, and institutional readiness in LLM-assisted academic grading remains limited, justifying the focus of the present study.

The growing reliance on Large Language Models in academic grading reflects a broader transformation in how higher education institutions respond to increasing workloads and demands for efficiency. Yet, grading is not simply a matter of technical execution; it represents a core academic responsibility that carries ethical, pedagogical, and institutional implications. When LLMs are introduced into assessment processes, they reshape not only how grades are produced but also how responsibility, authority, and trust are distributed. Despite this shift, ethical accountability and data governance often remain implicit or underarticulated in institutional practices. This disconnect between technological adoption and ethical reflection provides a strong rationale for examining how LLM-assisted grading is understood and managed within real academic environments.

Previous research has extensively addressed grading practices, assessment fairness, and pedagogical effectiveness across various educational contexts. More recent studies have begun to explore the use of artificial intelligence in assessment, primarily focusing on technical performance or instructional innovation. However, these strands of research rarely intersect. Empirical investigations that simultaneously examine ethical responsibility, data governance, and institutional readiness in the context of LLM assisted academic grading are still scarce. In particular, qualitative insights into how lecturers, students, and academic administrators negotiate accountability, transparency, and data protection in everyday assessment practices remain underdeveloped. This gap limits the ability of institutions to design governance frameworks that align technological innovation with academic values.

Responding to this gap, the present study seeks to explore how ethical responsibility and data governance are perceived and enacted in the use of Large Language Models for academic grading. Adopting a qualitative exploratory approach, the study focuses on the perspectives of lecturers, students, and academic administrators to understand how accountability, fairness, transparency, and institutional preparedness are interpreted in practice. Rather than advancing testable hypotheses, this study aims to generate context-sensitive insights that can inform ethical guidelines, data governance policies, and human-centered strategies for the responsible integration of LLMs in higher education assessment.

Method

Research Design

This study was designed as a qualitative exploratory inquiry aimed at understanding how ethical responsibility and data governance are negotiated in the use of Large Language Models (LLMs) for academic grading. A qualitative approach was considered most appropriate because the research focuses on meaning-making, perceptions, and institutional practices rather than on measuring system performance or testing predefined hypotheses. The exploratory character of the study enabled the researcher to capture ethical and governance issues that emerge organically from everyday assessment practices, particularly in contexts where formal guidelines are still evolving.

The research process followed a sequential and reflective logic, beginning with the identification of LLM use in academic grading as a contemporary phenomenon and progressing through problem formulation, data collection, analysis, and interpretation. This overall methodological pathway is presented in Figure 1, which visualizes how each stage of the study is connected and informs subsequent decisions.

Participant

Participants were selected through purposive sampling to ensure the inclusion of individuals who possessed relevant experience or institutional knowledge related to AI-assisted academic assessment. Three groups of stakeholders were involved: lecturers who had experience using or experimenting with LLMs in grading or feedback, students whose academic work may have been evaluated with AI support, and academic administrators or information technology personnel responsible for assessment systems and data management. To broaden the range of perspectives, snowball sampling was applied by asking initial participants to recommend others with similar experience. Participant recruitment continued until the data demonstrated thematic saturation, indicating that additional interviews no longer produced substantially new insights.

Instrument

Data collection relied on multiple qualitative instruments to strengthen depth and credibility. Semi-structured interviews constituted the primary data source, allowing participants to reflect on their experiences, concerns, and interpretations regarding accountability, fairness, and data governance in LLM-assisted grading. The interview guide was deliberately flexible, enabling the researcher to probe emerging themes and adapt questions to participants' roles and experiences. In addition, document analysis was conducted to examine institutional policies, assessment guidelines, and data protection documents relevant to AI use. Observational notes were also maintained to capture contextual details surrounding grading and feedback practices, providing complementary insights beyond interview data.

Data Analysis

Data analysis was conducted as an iterative and interpretive process. Interview transcripts, institutional documents, and observational notes were first organized and read repeatedly to develop familiarity with the data. Analysis then proceeded through stages of data reduction, data display, and interpretation, allowing patterns and relationships to emerge progressively. Initial coding focused on recurring issues related to ethical responsibility, accountability, fairness, transparency, data governance, and institutional readiness. These codes were subsequently refined and grouped into broader thematic categories. Throughout the analysis, triangulation across data sources and participant groups was employed to enhance trustworthiness and to ensure that interpretations were grounded in multiple forms of evidence. The interconnection between data collection, analysis, and interpretation is depicted in Figure 1, which illustrates how analytical insights informed the development of the study's conclusions.

Figure 1. Research Methodology Flowchart

Results and Discussion

Results

The qualitative inquiry generated a coherent set of findings that illuminate how Large Language Models are currently integrated into academic grading practices. Drawing on interviews, document analysis, and observations, the analysis reveals that participants' experiences with LLM-assisted grading are shaped less by technical considerations than by ethical uncertainty, governance gaps, and institutional conditions.

One prominent finding concerns ethical responsibility. Across participant groups, there was no shared understanding of where accountability resides when grading decisions are influenced by LLM outputs. Lecturers described a tension between efficiency and professional responsibility, often emphasizing that AI-generated evaluations could not be accepted uncritically. Students, in turn, reported difficulty understanding how final grades were determined when AI tools were involved, which contributed to perceptions of reduced transparency. From an administrative perspective, responsibility was frequently described as diffuse, with no formal mechanisms to address disputes or errors linked to AI-supported grading.

Issues of fairness and bias also emerged strongly from the data. Participants acknowledged that LLMs were helpful in assessing formal structure and surface coherence, yet they were widely perceived as limited in capturing originality, contextual meaning, and culturally specific expressions. Lecturers noted that these limitations required additional human judgment, particularly to avoid disadvantaging students whose writing did not conform to dominant academic norms. Such concerns reinforced the view that LLMs function as supportive tools rather than autonomous evaluators.

A third area of concern relates to data governance and privacy. Many participants were uncertain about how student assignments processed through LLMs were stored, shared, or reused. Students expressed unease about the lack of explicit consent mechanisms, while administrators acknowledged that policies governing AI-related data practices were often fragmented or incomplete. These findings suggest that governance structures have not kept pace with the rapid adoption of AI tools in assessment contexts.

Finally, the analysis highlights uneven levels of institutional readiness. While some institutions encouraged innovation in digital assessment, this encouragement was rarely accompanied by systematic training or ethical guidance. As a result, decisions about whether and how to use LLMs in grading were largely left to individual lecturers, leading to inconsistencies across courses and departments.

An overview of the main themes and associated sub-themes identified through the analysis is presented in Table 1, which summarizes the core dimensions shaping LLM-assisted academic grading.

Table 1. Key Themes and Sub-Themes Identified from the Analysis

Main Theme	Sub-Themes	Analytical Description
Ethical Responsibility	Ambiguity of accountability, reliance on AI outputs	Participants described uncertainty regarding who bears responsibility when grading decisions are influenced by LLM-generated evaluations
Fairness and Bias	Contextual limitations, linguistic diversity	LLMs were viewed as efficient yet limited in interpreting originality, context, and culturally situated expressions
Data Governance	Consent, data storage, external platforms	Limited awareness was reported regarding how student data are processed, stored, or reused by AI systems
Institutional Readiness	Policy gaps, lack of training	The use of LLM-assisted grading was largely dependent on individual lecturers rather than institutional frameworks

To further capture differences and commonalities among stakeholder groups, the analysis compared perspectives across lecturers, students, and academic administrators. This comparison, summarized in Table 2, demonstrates that while concerns vary by role, uncertainty regarding accountability and governance is shared across groups.

Table 2. Stakeholder Perspectives on LLM-Assisted Academic Grading

Theme	Lecturers	Students	Administrators
Ethical Responsibility	Expressed concern over maintaining professional accountability	Reported reduced transparency in grading decisions	Recognized the absence of formal responsibility frameworks
Fairness and Bias	Noted the need for manual adjustment of AI-supported grades	Perceived bias toward standardized academic language	Acknowledged limitations without clear mitigation strategies
Data Governance	Uncertain about data handling by AI platforms	Worried about privacy and informed consent	Confirmed fragmented governance arrangements
Institutional Readiness	Relied on personal judgment and experience	Experienced inconsistent assessment practices	Highlighted the lack of institutional guidelines

Taken together, these findings indicate that the implementation of LLMs in academic grading is shaped by unresolved ethical questions, insufficient data governance, and varying degrees of institutional preparedness. The results provide an empirical foundation for interpreting the broader implications of AI-assisted assessment in higher education.

Discussion

The findings of this study indicate that the use of Large Language Models in academic grading represents more than a technological adjustment; it reflects a shift in how evaluative authority is exercised within higher education. Grading, as shown in this study, continues to function as a normative practice that shapes trust and legitimacy. This interpretation is consistent with Rojas Bruna (2025), who emphasizes that assessment methods actively construct professional and institutional norms rather than merely measuring performance.

A central issue emerging from the findings is the uncertainty surrounding ethical responsibility. Lecturers' reluctance to fully delegate grading decisions to LLMs reveals an underlying concern about preserving professional accountability. This concern echoes Kirmizi (2025), whose work demonstrates that grading practices are inseparable from teachers' interpretive roles and expectations. When algorithmic outputs influence assessment, responsibility becomes distributed in ways that are not yet clearly defined.

The findings further show that concerns about fairness are closely tied to the interpretive limits of automated evaluation. Participants' observations that LLMs struggle with contextual meaning and originality resonate with Biton (2025), who conceptualizes assessment as a reflective and dialogic process. From this perspective, grading cannot be reduced to pattern recognition without losing its pedagogical depth.

In mathematics education, the risk of misinterpretation through simplified evaluation mechanisms has been widely discussed. The present findings parallel the analysis by Srđan, Marija, Zorana, Dragana, and Branislav (2025), who demonstrate how statistical regularities can obscure inequities in grading outcomes. Although LLMs differ from statistical aggregation, both forms of automation risk masking underlying disparities when applied without critical oversight.

The need for human intervention identified in this study also aligns with research on standards-based grading. Uhing, Bennett, and Wright (2025) argue that assessment systems require continuous professional judgment to remain pedagogically aligned. The present findings suggest that LLM-assisted grading similarly depends on human oversight to ensure that efficiency does not override educational intent.

While technological feasibility has attracted considerable scholarly attention, ethical accountability has received far less emphasis. This imbalance is evident in Hahn's (2025) exploration of ChatGPT for grading algebra problems, which prioritizes functional performance. The current study extends this line of inquiry by demonstrating that technical capability alone does not address concerns related to responsibility, transparency, and governance.

Students' responses in this study reveal that grading practices have implications beyond academic outcomes. Feelings of uncertainty and discomfort related to opaque evaluation processes align with the findings of Estipona and Delos Santos (2025), who link assessment experiences to student wellbeing. When grading systems lack transparency, they may weaken students' sense of fairness and institutional trust.

Institutional context plays a decisive role in shaping how AI tools are adopted. The uneven readiness observed in this study mirrors the challenges identified by Kihaga, Dahl, Kitta, and Likinjiye (2025) in large-class assessment environments. In both cases, technological solutions are introduced into settings where governance structures and professional support remain limited.

Broader educational research further supports the need for systemic perspectives. Curriculum-focused studies by Cilsalar-Sagnak, Anakok, and Katz (2025) illustrate that educational practices are shaped by institutional design choices rather than isolated tools. Similarly, the scoping review by Sprague, Scott, Mehranbod, Branas, and Factor-Litvak (2025) highlights how external and contextual factors influence educational outcomes. These insights reinforce the argument that LLM-assisted grading must be understood within wider organizational and policy frameworks.

Taken together, the discussion suggests that the responsible integration of Large Language Models into academic grading depends primarily on ethical clarity and institutional commitment. Rather than replacing human judgment, LLMs function most appropriately as supportive tools within clearly defined governance structures. By foregrounding ethical responsibility and data governance, this study contributes to ongoing debates on AI in education and underscores the importance of aligning technological innovation with the foundational values of higher education.

Implications

The findings suggest that the use of Large Language Models in academic grading requires a shift in how institutions conceptualize assessment governance. Rather than framing LLMs as efficiency-enhancing tools, universities need to recognize their influence on ethical responsibility and decision-making authority. Clarifying the boundaries between algorithmic assistance and human judgment is essential to prevent the dilution of academic accountability. This clarification should be embedded not only in policy documents but also in everyday assessment practices.

The study also highlights the strategic importance of data governance in maintaining institutional trust. Transparent mechanisms for consent, data handling, and external platform use are not merely technical requirements but ethical commitments to students as academic subjects. Without such mechanisms, even limited use of LLMs in grading risks undermining confidence in institutional integrity. These implications point to the need for governance frameworks that integrate ethical reflection, professional responsibility, and technological practice rather than treating them as separate domains.

Limitations

Several limitations should be considered when interpreting the findings of this study. As an exploratory qualitative investigation, the research emphasizes interpretive depth and contextual understanding rather than representativeness. The insights presented here are grounded in participants' experiences and institutional settings, which may differ from those in other higher education environments.

In addition, the study does not evaluate the technical accuracy or comparative performance of LLM-assisted grading against human-only assessment. This focus was intentional, as the research aimed to examine ethical responsibility and governance rather than system efficiency. Furthermore, given the rapid development of AI technologies and institutional policies, some practices described by participants may reflect temporary or transitional conditions rather than stable models of assessment governance.

Suggestions

Future research could extend this work by examining how ethical responsibility and data governance in LLM-assisted grading are addressed across different academic disciplines, institutional types, and regulatory contexts. Comparative studies may help identify governance practices that are transferable across settings, as well as those that are context-specific.

Further investigation is also needed to explore students' roles in shaping assessment governance, particularly in relation to consent, transparency, and the right to contest AI-supported grading decisions. Longitudinal research would be valuable in tracing how institutional approaches to ethical responsibility evolve as LLMs become more deeply embedded in assessment practices. Such studies could contribute to the development of governance models that balance technological innovation with the core academic values of fairness, accountability, and trust.

Conclusion

This study investigated the incorporation of Large Language Models into academic grading by examining issues of ethical responsibility and data governance from the viewpoints of lecturers, students, and academic administrators. The results demonstrate that the use of LLMs in assessment cannot be reduced to a question of technological efficiency, as it fundamentally alters how accountability, fairness, and transparency are enacted within higher education. Although LLMs may assist in managing workload and enhancing procedural consistency, their involvement in grading creates uncertainty regarding who ultimately bears responsibility for evaluative decisions and how student data are governed. The findings further reveal persistent concerns about the inability of algorithmic systems to adequately interpret contextual meaning, originality, and culturally embedded forms of expression, reaffirming the indispensable role of human judgment in academic assessment. At the institutional level, limited preparedness and fragmented governance frameworks constrain the responsible adoption of LLM-assisted grading and often leave ethical considerations to individual discretion. In conclusion, this study positions Large Language Models as supportive instruments that require clearly defined ethical boundaries and robust data governance to ensure that assessment practices remain aligned with academic integrity, transparency, and institutional trust.

Author Contributions Statement

Ruri Supatmi was responsible for conceptualizing the study, designing the research framework, and leading the data collection process. She conducted the interviews, coordinated document analysis and observations, and played a central role in interpreting the qualitative data. She also drafted the initial version of the manuscript and integrated revisions across all sections. Diah Dwi Agustina contributed to the development of the theoretical framework and literature review, with particular attention to ethical considerations and data privacy in AI-assisted education. She supported the data analysis process by refining thematic categories and contributed substantially to the discussion and implications sections through critical interpretation of the findings. Asti Nur Cahyani assisted in methodological refinement, data organization, and validation procedures, including triangulation and member checking. She contributed to the analysis of institutional and policy-related findings and supported the final editing process to ensure coherence, clarity, and alignment with academic and ethical standards. All authors reviewed and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

References

Abfalter, D., Mueller-Seeger, J., & Raich, M. (2021). Translation decisions in qualitative research: A systematic framework. *International Journal of Social Research Methodology*, *24*(4), 469–486. https://doi.org/10.1080/13645579.2020.1805549

Agostini, D., & Picasso, F. (2024). Large language models for sustainable assessment and feedback in higher education: Towards a Pedagogical and Technological Framework. *Intelligenza Artificiale*, *18*(1), 121–138. https://doi.org/10.3233/IA-240033

Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, *86*(2), 369–385. https://doi.org/10.1007/s10734-022-00937-2

Biton, Y. (2025). Student and teacher learning as a result of developing peer assessment criteria for mathematical tasks. *Eurasia Journal of Mathematics, Science and Technology Education*, *21*(7). https://doi.org/10.29333/ejmste/16605

Brown, G. T. L. (2022). The past, present and future of educational assessment: A transdisciplinary perspective. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.1060633

Cilsalar-Sagnak, H., Anakok, I., & Katz, A. (2025). Curriculum comparison: Chemical and mechanical engineering education in the United States and Turkey. *Australasian Journal of Engineering Education*, *30*(1), 62–79. https://doi.org/10.1080/22054952.2024.2441091

de Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly*, *39*(2), 101666. https://doi.org/10.1016/j.giq.2021.101666

Diab Idris, M., Feng, X., & Dyo, V. (2024). Revolutionizing Higher Education: Unleashing the Potential of Large Language Models for Strategic Transformation. *IEEE Access*, *12*, 67738–67757. https://doi.org/10.1109/ACCESS.2024.3400164

Elshall, A. S., & Badir, A. (2025). Balancing AI-assisted learning and traditional assessment: The FACT assessment in environmental data science education. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1596462

Estipona, E. P., & Delos Santos, M. S. M. (2025). Linking wellbeing to success: Life satisfaction on mathematics performance of Philippine public high school students. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1540813

Fazelpour, S., & Danks, D. (2021). Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*, *16*(8), e12760. https://doi.org/10.1111/phc3.12760

Gazit, L. (2025). AI as a Group Mediator: A Conceptual Framework for Triadic Chat-Based Therapy. *International Journal of Systemic Therapy*, *0*(0), 1–30. https://doi.org/10.1080/2692398X.2025.2587315

Gonsalves, C., & Lin, Z. (2025). Clear in advance to whom? Exploring 'transparency' of assessment practices in UK higher education institution assessment policy. *Studies in Higher Education*, *50*(7), 1454–1470. https://doi.org/10.1080/03075079.2024.2381124

Hahn, M. G. (2025). Work in Progress: Investigating ChatGPT for Grading Algebra Problems. *EDUNINE - IEEE Eng. Educ. World Conf.: Educ. Age Gener. AI: Embrac. Digit. Transform. - Proc.* EDUNINE 2025 - 9th IEEE Engineering Education World Conference: Education in the Age of Generative AI: Embracing Digital Transformation - Proceedings. https://doi.org/10.1109/EDUNINE62377.2025.10981322

Katsamakas, E., Pavlov, O. V., & Saklad, R. (2024). Artificial Intelligence and the Transformation of Higher Education Institutions: A Systems Approach. *Sustainability*, *16*(14), 6118. https://doi.org/10.3390/su16146118

Kihaga, H., Dahl, B., Kitta, S., & Likinjiye, M. (2025). Strategies for teaching and assessing students in large mathematics classes in secondary schools in Tanzania: A systematic review. *Social Sciences and Humanities Open*, *12*. https://doi.org/10.1016/j.ssaho.2025.101744

Kirmizi, M. (2025). Exploring the mediating role of teacher expectation on whole class participation. *Power and Education*, *17*(3), 334–349. https://doi.org/10.1177/17577438241272594

König, P. D. (2021). Citizen-centered data governance in the smart city: From ethics to accountability. *Sustainable Cities and Society*, *75*, 103308. https://doi.org/10.1016/j.scs.2021.103308

Kooli, C. (2023). Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions. *Sustainability*, *15*(7), Article 7. https://doi.org/10.3390/su15075614

Lim, K. (2024). Assessing beyond grades: Unravelling the implications on student learning and engagement in higher education. *Assessment & Evaluation in Higher Education*, *49*(5), 665–679. https://doi.org/10.1080/02602938.2024.2314703

Meliou, E., Ozbilgin, M., & Edwards, T. (2021). How does responsible leadership emerge? An emergentist perspective. *European Management Review*, *18*(4), 521–534. https://doi.org/10.1111/emre.12488

O, 'Dea Xianghan, & O, 'Dea Mike. (2023). Is artificial intelligence really the next big thing in learning and teaching in higher education?: A conceptual paper. *Journal of University Teaching and Learning Practice*, *20*(5), 1–17. https://doi.org/10.3316/informit.T2024112700021091343163252

Pantiris, P., Pallis, P. L., Chountalas, P. T., & Dasaklis, T. K. (2025). Enhancing Coordination and Decision Making in Humanitarian Logistics Through Artificial Intelligence: A Grounded Theory Approach. *Logistics*, *9*(3), 113. https://doi.org/10.3390/logistics9030113

Pham, S. T. H., & Sampson, P. M. (2022). The development of artificial intelligence in education: A review in context. *Journal of Computer Assisted Learning*, *38*(5), 1408–1421. https://doi.org/10.1111/jcal.12687

Rojas Bruna, C. E. (2025). Enhancing primary teacher training through academic portfolios in advanced mathematics courses. *International Electronic Journal of Mathematics Education*, *20*(4). https://doi.org/10.29333/iejme/16635

Sebestyén, M. (2025). Focal points and blind spots of human-centered AI: AI risks in written online media. *Humanities and Social Sciences Communications*, *12*(1), 564. https://doi.org/10.1057/s41599-025-04814-y

Sprague, N. L., Scott, S. N., Mehranbod, C. A., Sachs, A. L., Ekenga, C. C., Rundle, A. G., Branas, C. C., & Factor-Litvak, P. (2025). Changing Degrees: A weight-of-evidence scoping review examining the impact of childhood exposures to climate change on educational outcomes. *Environmental Research*, *277*. https://doi.org/10.1016/j.envres.2025.121639

Srđan, V., Marija, K., Zorana, L., Dragana, T., & Branislav, R. (2025). Simpson's paradox in mathematics grading: A case study of Serbian primary schools. *Zbornik Instituta za Pedagoska Istrazivanja*, *57*(1), 5–28. https://doi.org/10.2298/ZIPI2501005V

Torelli, R. (2020). Sustainability, responsibility and ethics: Different concepts for a single path. *Social Responsibility Journal*, *17*(5), 719–739. https://doi.org/10.1108/SRJ-03-2020-0081

Toscani, G. (2025). Integrating minds: Adaptive knowledge sharing strategies for ML team synergy. *Cognition, Technology & Work*, *27*(4), 745–761. https://doi.org/10.1007/s10111-025-00822-9

Uhing, K., Bennett, A. B., & Wright, G. (2025). Students' Experiences in a Coordinated College Algebra Course: A Case Study of Implementing Active Learning and Standards-Based Grading. *International Journal of Research in Undergraduate Mathematics Education*. https://doi.org/10.1007/s40753-025-00265-7

Valentine, N., Durning, S., Shanahan, E. M., & Schuwirth, L. (2021). Fairness in human judgement in assessment: A hermeneutic literature review and conceptual framework. *Advances in Health Sciences Education*, *26*(2), 713–738. https://doi.org/10.1007/s10459-020-10002-1

Verhulst, S. G. (2021). Reimagining data responsibility: 10 new approaches toward a culture of trust in re-using data to address critical public needs. *Data & Policy*, *3*, e6. https://doi.org/10.1017/dap.2021.4

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90–112. https://doi.org/10.1111/bjet.13370

Yi, H., Pingsterhaus, A., & Song, W. (2021). Effects of Wearing Face Masks While Using Different Speaking Styles in Noise on Speech Intelligibility During the COVID-19 Pandemic. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.682677

Zou, W., Goh, T.-T., Zhu, H., Liu, M., & Yang, B. (2025). Algorithmic Learning: Assessing the Potential of Large Language Models (LLMs) for Automated Exercise Generation and Grading in Educational Settings. *International Journal of Human–Computer Interaction*, *0*(0), 1–18. https://doi.org/10.1080/10447318.2025.2520931