University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR)

ISSN: 2714-3627

A Journal of the Faculty of Computing, University of Ibadan, Ibadan, Nigeria

Volume 16 No. 1, January 2026

journals.ui.edu.ng/uijslictr
http://uijslictr.org.ng/
uijslictr@gmail.com

University of Ibadan
Journal of Science and Logics in ICT Research

A Systematic Review of AI-Powered Assessment and Feedback to Enhance Teaching Effectiveness in Higher Education

[1]Ibitola Ayobami G., [2]Agbaegbu JohnBosco, [3]Nathaniel Benjamen, [4]Olatunji Temitope M.
[1,2] Augustine University, Ilara Epe, Lagos
[3] Lagos State University, Lagos State
[4] Chrisland University, Abeokuta, Ogun State
ayobami.ibitola@augustineuniversity.edu.ng; mazijohnbosco@gmail.com; nathaniel.benjamen@lasu.edu.ng; tolatunji@chrislanduniversity.edu.ng

Abstract
The growing use of artificial intelligence in higher education assessment offers a real chance to improve teaching effectiveness. However, the practical, ethical, and pedagogical aspects of this shift are not yet well understood. This paper presents a systematic review of 127 studies selected from 699 candidate papers published between 2022 and 2025. The review examines evidence across five themes: feedback personalisation, assessment accuracy, ethical and equity concerns, human-in-the-loop integration, and pedagogical impact. The findings show that AI tools, particularly GPT-4, can deliver personalised, timely, and scalable feedback that improves student engagement and academic outcomes while reducing educator workload. However, AI-generated feedback is often less sensitive to context, empathy, and higher-order thinking tasks, especially in creative and humanities subjects. Hybrid models that combine AI with human oversight are the well-supported approach, as they improve grading accuracy, fairness, and student trust. Issues such as algorithmic bias, data privacy, lack of transparency, and weak governance remain key ethical challenges that need strong institutional responses. This review offers a clear evidence base to guide educators, policymakers, and technologists on how to use AI-enhanced assessment in a responsible and sustainable way.

Keywords**:** Artificial Intelligence in education; AI-powered assessment and feedback; Teaching effectiveness in higher education; Human-in-the-loop assessment models; Large language models in education

1. Introduction

Higher education has changed greatly over the past two decades due to rising student numbers, new ways of learning, and fast-growing digital tools. Teaching effectiveness, which refers to the ability of teaching methods to improve learning outcomes, student interest, and critical thinking, has become a key concern for educators, administrators, and policymakers around the world [1, 2]. Traditional assessment and feedback methods have struggled to keep up with these demands, especially in large and diverse classrooms where giving timely, ndividual feedback is hard to maintain [3, 4].

Assessment and feedback in higher education do more than measure student achievement. Good feedback that is specific, timely, and action-focused is widely seen as one of the strongest factors in student learning [7]. Yet providing such feedback at scale remains a real challenge. Studies show that as many as 28% of students in higher education are unhappy with the quality and speed of feedback they receive [3, 8]. This problem is worse in large courses where educators are too busy to give personalised feedback to every student [9, 10].

Modern AI systems can analyse open-ended student responses, produce detailed written feedback, and adjust content to suit individual learners in ways that older automated tools could not [11, 12]. Generative AI systems such as GPT-4 and ChatGPT have taken this further, producing feedback with a level of specificity and adaptability that was once only possible from human instructors [13, 14]. Figure 1 shows the conceptual framework used in this review, which positions AI as a bridge between instructional quality and student outcomes within a constructivist and formative assessment approach [5, 17].

Ibitola A. G., Agbaegbu J., Nathaniel B., and Olatunji T. M. (2026). A Systematic Review of AI-Powered Assessment and Feedback to Enhance Teaching Effectiveness in Higher Education. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 16 No. 1, pp. 122 – 131.

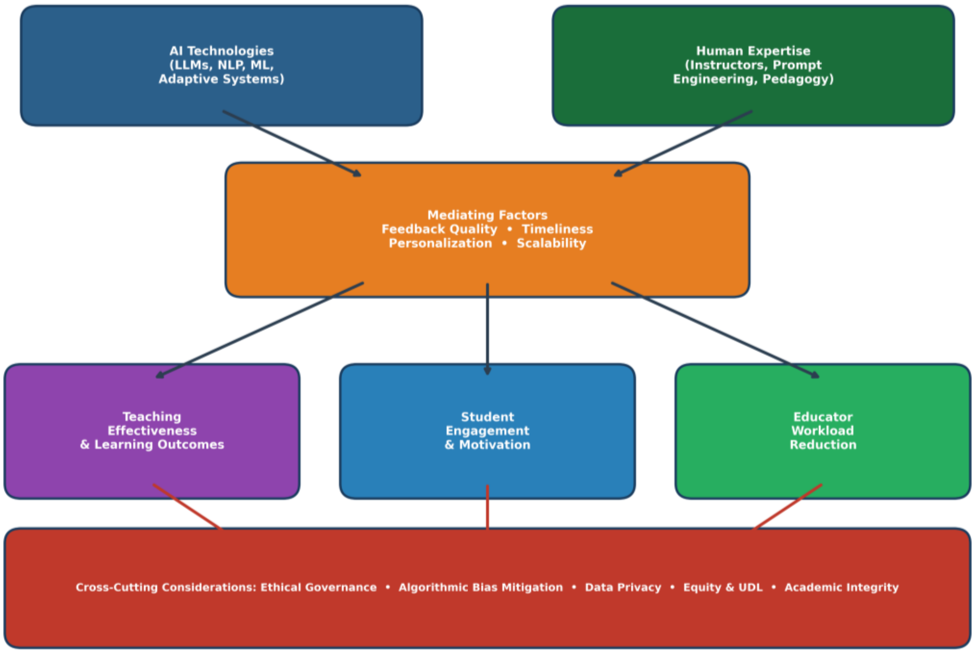

Figure 1: Conceptual Framework - AI-Powered Assessment and Teaching Effectiveness in Higher Education.

Despite this progress, important challenges remain. Current AI systems often fall short of experienced human educators when it comes to reading context, showing empathy, and assessing creative or culturally complex student work [21, 22]. Ethical issues such as algorithmic bias, data privacy, and unclear AI decision-making create risks that tend to affect the most vulnerable student groups the most [25, 26, 27]. This systematic review was carried out to bring together the available evidence, identify what works, and set out clear directions for using AI in higher education assessment responsibly.

2. Related Works

How AI tools should be combined with human oversight has attracted growing research interest. Selvam and Vallejo [32] proposed a human-in-the-loop AI grading model that reduced bias and improved transparency by pairing AI speed with human ethical checks. Pike [1] used case studies to show that students rated AI feedback highly for speed and scale, but preferred human feedback for its depth and developmental guidance. Er *et al.* [21] compared student views of instructor and AI-generated feedback and found that instructor feedback led to higher score gains and was seen as more helpful, though students valued AI feedback for being quick. These studies all point to the tension between efficiency and teaching quality in the human-AI collaboration debate.

More recent work has moved toward building integrated theoretical models. Yu *et al.* [5] proposed the AI-Educational Development Loop, which links AI feedback with self-regulated learning and established educational theories. Lee [42] developed a socio-technical framework for process-based assessment that keeps human judgment in place within AI-supported evaluation settings. Jacobsen *et al.* [18] showed that how a prompt is designed, and the skill of the educator writing it, strongly affects the quality of LLM-generated feedback. This highlights how much human skill matters in getting useful results from AI tools.

A key finding across these studies is that AI tools do not work well on their own. Their value in teaching depends on the setting, the competence of educators, and the governance structures in place. Even though there is growing evidence that hybrid human-AI models work well, there is still no agreement on which specific arrangements produce the best outcomes in terms of quality, efficiency, and fairness [32, 33, 34].

3. Methodology

This study used a systematic review approach to bring together evidence on AI-supported assessment and feedback in higher education. The review followed standard systematic review steps to make sure the process of finding, selecting, and analysing literature was thorough, clear, and repeatable [2, 36].

The literature search covered a database of more than 270 million research articles, using clear inclusion and exclusion criteria to ensure quality and relevance. The initial search produced 623 candidate papers. An additional 76 papers were added through backward and forward citation tracking [43], giving a total of 699 candidate papers. Each paper was scored for relevance to the research questions of the study. The 127 highest-scoring papers were selected for full review. The full selection process is illustrated in the PRISMA diagram in Figure 2.

The 127 selected studies were reviewed both by theme and by time period to track how the field has developed and to find areas of agreement and disagreement. The studies used a range of methods, including randomised controlled trials, quasi-experimental designs, systematic reviews, case studies, and mixed-methods research from different parts of the world. Figure 3 shows the spread of research methods used across the reviewed studies. Five themes were identified for analysis: feedback personalisation, assessment accuracy, ethical and equity concerns, human-in-the-loop integration, and pedagogical impact. Findings were brought together by comparing empirical results, theoretical contributions, and practical recommendations [12, 39].

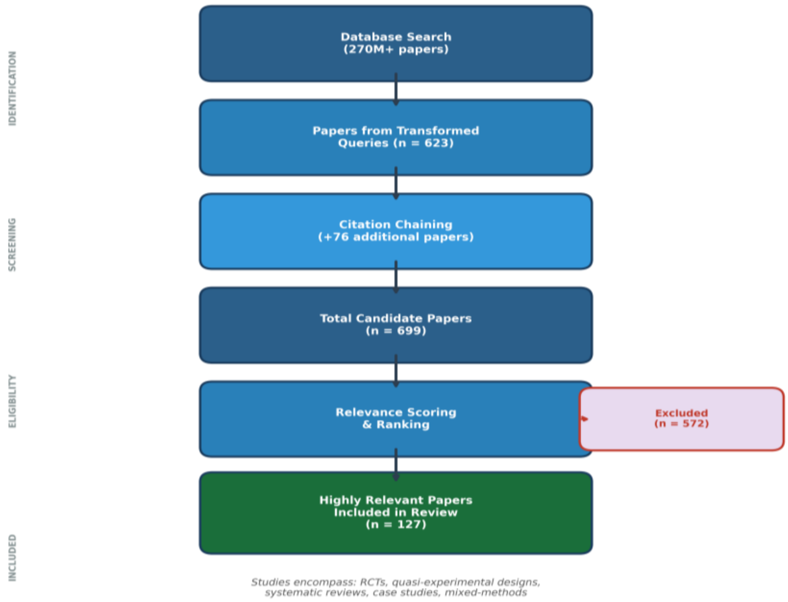

Figure 2: PRISMA Flow Diagram of Literature Selection Process.

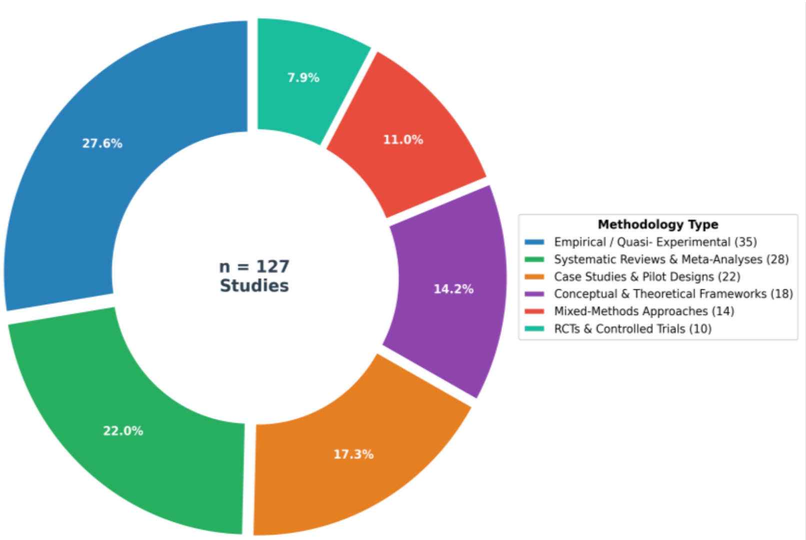

Figure 3: Methodological Distribution of the 127 Reviewed Studies.

4. Results and Discussion

The review of 127 studies shows that this field has both real strengths and clear limitations. Figure 4 shows where reviewed studies agreed and disagreed across key criteria, providing context for the five themes discussed below. The distribution of major themes across all 127 studies is further illustrated in Figure 5, which shows how often each theme appeared in the reviewed literature.

4.1 Feedback Personalisation

More than 60 studies found that AI systems can produce individual, timely, and context-adjusted feedback based on what each learner needs [3, 44, 45]. Systems that use adaptive learning algorithms and large language models were especially good at matching feedback to individual knowledge gaps and learning paths. The use of knowledge graphs and feedback loops helped make AI-generated feedback more specific and relevant over time [46, 47, 5]. A common weakness, however, was that AI systems often lack the deep, student-specific awareness that human instructors build through direct relationships with students [1, 21]. These findings support the use of hybrid models that combine the reach of AI with the context-awareness of human judgment.

4.2 Assessment Accuracy

About half of the reviewed papers looked at how closely AI-generated grades matched human grades across essays, programming tasks, and short-answer tests [49, 50, 51]. In many structured settings, large language models including GPT-4 graded as well as or better than skilled human graders, with agreement rates above 75% within a 10% margin in some cases [51]. Better prompt design and fine-tuning of models improved grading reliability further for structured and semi-structured tasks [18, 41]. However, agreement between AI and human graders was consistently lower for tasks that required creative thinking, critical analysis, or weighing of multiple views [23, 52]. Figure 6 shows how AI assessment accuracy changes across different levels of task complexity based on evidence from the reviewed studies. Randomised controlled trials confirmed that AI grading works reliably for routine assessments but needs human oversight for high-stakes and subject-specific evaluations [50, 53, 54].

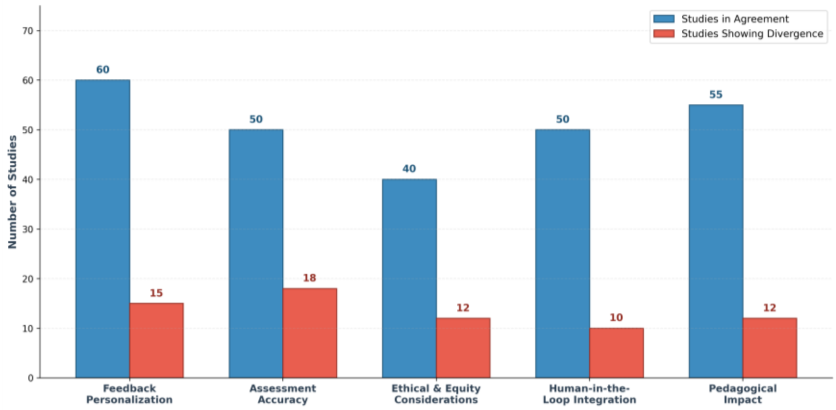

Figure 4: Agreement and Divergence Across Key Comparison Criteria in the Reviewed Literature.

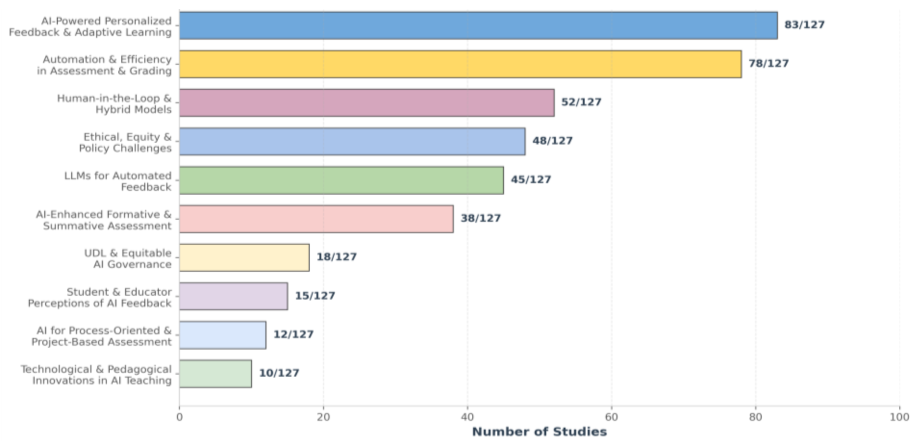

Figure 5: Distribution of Major Themes Across the 127 Reviewed Studies.

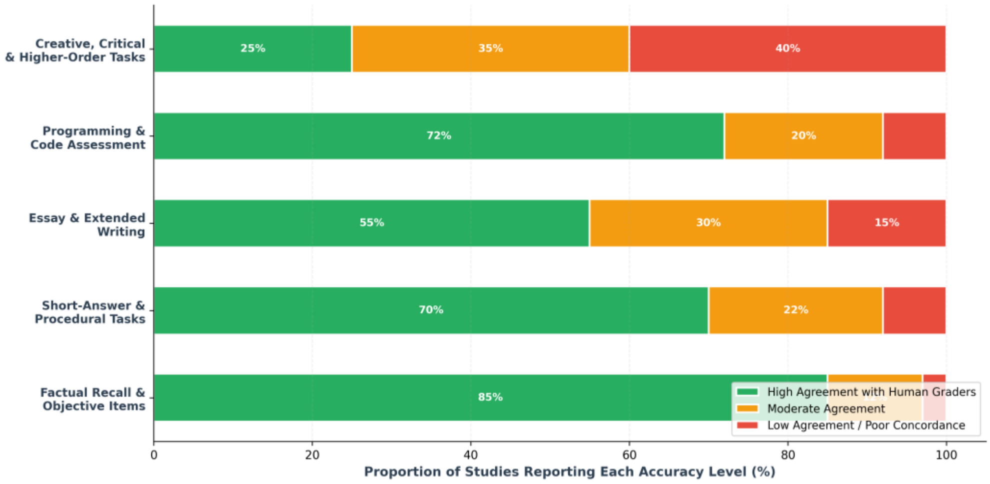

Figure 6: AI Assessment Accuracy Across Task Complexity Levels (Based on Reviewed Empirical Evidence).

4.3 Ethical and Equity Considerations

More than 40 studies covered ethical concerns in AI-based assessment, including algorithmic bias, data privacy, equity, and access [25, 26, [27]. Several studies found that AI systems trained on biased data or using unclear decision processes can copy or worsen existing gaps in education, especially in under-resourced settings and among disadvantaged student groups [30, 55]. Suggested solutions included bias checks, open AI governance frameworks, and inclusive institutional policies [25, 56]. A consistent finding across this literature is that ethical concerns are not a side issue in AI-based assessment; they are central to making it work fairly. The gap between ethical AI principles described in research and how they are actually applied in institutions is a pressing area needing policy action.

4.4 Human-in-the-Loop Integration

More than half the reviewed studies stressed the importance of human oversight in AI-assisted grading and feedback [1, 32, 34]. Hybrid models that combine human judgment with AI consistently did better than fully automated systems across feedback quality, grading fairness, reduced educator workload, and student trust [32, 52]. Figure 7 shows the performance difference between hybrid and fully automated AI assessment models across several key measures. Figure 8 further compares the strengths of AI-only, human-only, and hybrid feedback models, making clear that no single approach performs best on all measures. Educator skill in using AI tools and in writing effective prompts came up as critical success factors. Studies showed that the quality of AI-generated feedback is much better when educators with relevant skills design the inputs [18]. Fully automated systems, while able to handle large numbers, were consistently found to struggle with complex, subjective, or culturally sensitive tasks [49, 53]. The evidence is clear: human oversight in AI-based assessment is not optional. It is essential for fair, context-aware, and educationally sound outcomes.

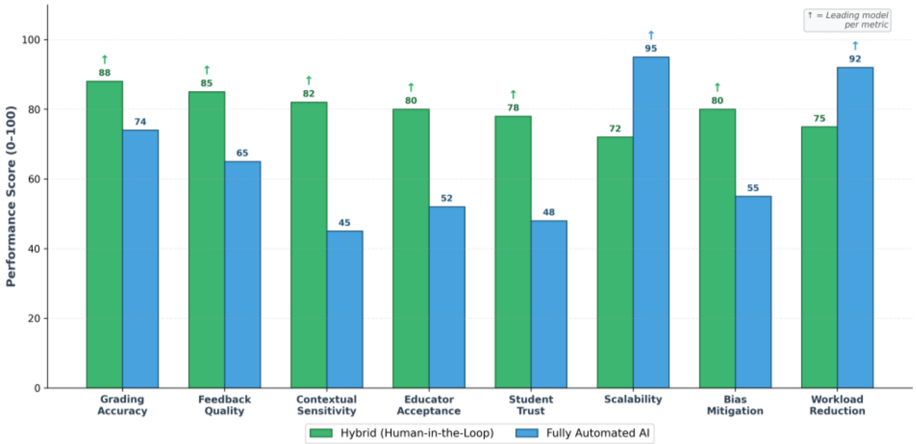

Figure 7: Performance Comparison of Hybrid vs. Fully Automated AI Assessment Models Across Multiple Evaluation Metrics.

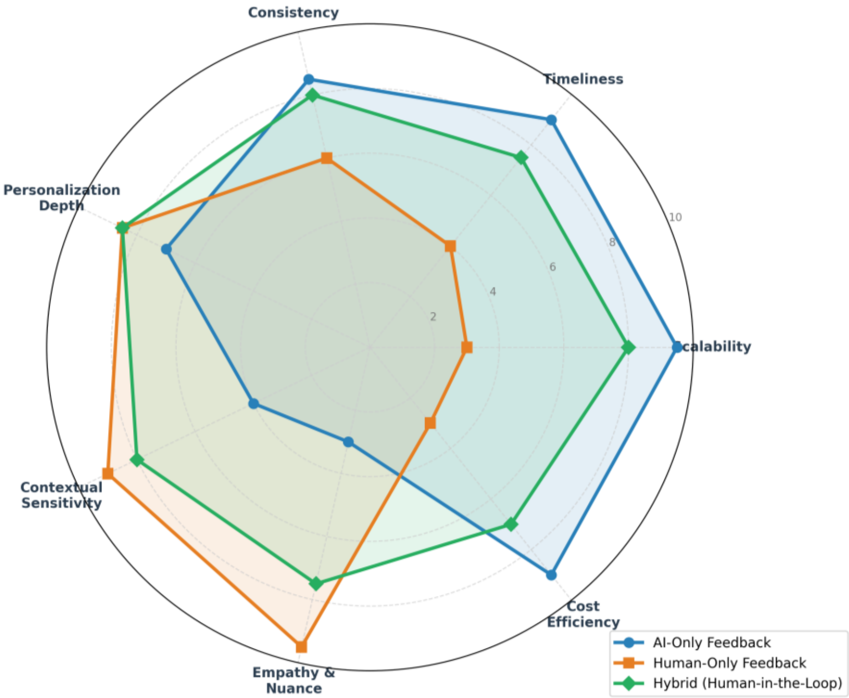

Figure 8: Comparative Strengths of AI-Only, Human-Only, and Hybrid Feedback Models.

4.5 Pedagogical Impact

Studies across many subject areas confirm that AI-based assessment and feedback tools can improve teaching quality, student interest, and learning outcomes [58, 59, 40]. These tools support formative assessment by enabling repeated learning cycles, helping students manage their own learning, and keeping motivation up through quick and frequent feedback [5, 7, 8]. Cooper-Stachowsky [59] found that students who received AI-generated formative feedback with a chance to resubmit performed significantly better on summative tests. Zhang and Abdullah [64] showed that AI-generated feedback, especially from ChatGPT, helped weaker students the most, narrowing the gap between high and low performers and contributing to fairer learning outcomes. Some challenges were noted around emotional sensitivity, especially in subjects where the personal relationship between teacher and student plays a big role in learning [61, 62]. The timeline of research shown in Figure 9 reveals a clear shift from basic proof-of-concept studies in 2022 and 2023 to more theory-driven and policy-focused work by 2024 and 2025, showing that the field is maturing.

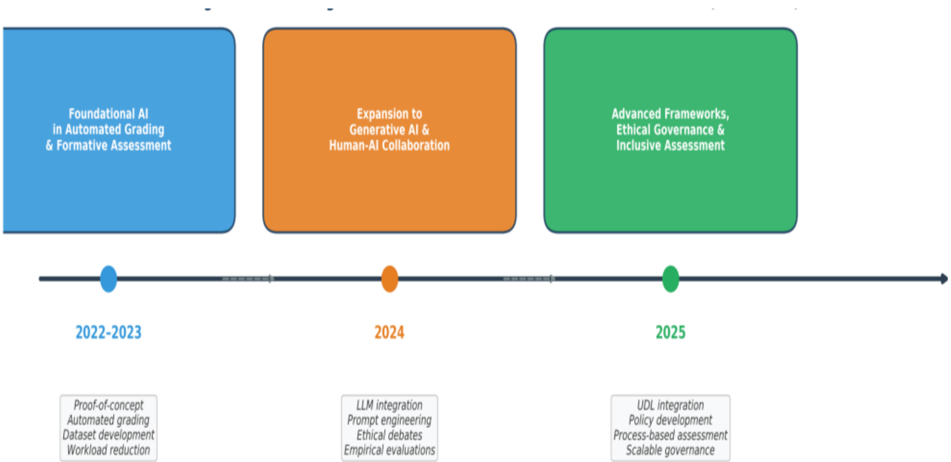

Figure 9: Chronological Evolution of AI-Powered Assessment Research (2022-2025).

4.6 Discussion

The findings give a balanced view of what AI can offer in higher education assessment, and of the real challenges that need to be solved before those benefits can be achieved fairly. Evidence that AI systems, particularly large language models, can match or outperform human graders on structured tasks is a real step forward from earlier rule-based systems [40, 41, 50]. Adaptive learning paths, knowledge graphs, and feedback loops make personalised feedback possible at scale, directly helping to solve the longstanding problem of providing good feedback in high-enrolment courses [3, 9].

However, the limitations are just as important. AI feedback regularly falls short on the subtle, context-specific, and empathetic qualities that define good human teaching, especially in creative arts, humanities, and social sciences [1, 21, 23]. The inability of current AI systems to read the emotional side of learning, including student frustration, motivation, and personal background, is a basic limitation on the value of automated feedback [61, 62]. Figure 10 provides a view of the strengths and limitations found across key dimensions of the AI-powered assessment literature, while Figure 11 maps the identified limitations by severity and urgency.

Hybrid human-in-the-loop models are the most consistently recommended approach in the reviewed literature [1, 10, 32, 34]. These models use AI for routine, large-scale grading while keeping human judgment for complex, high-stakes, and subject-specific tasks. Ethical concerns around algorithmic bias, data privacy, and unclear AI decision-making are not just technical problems.

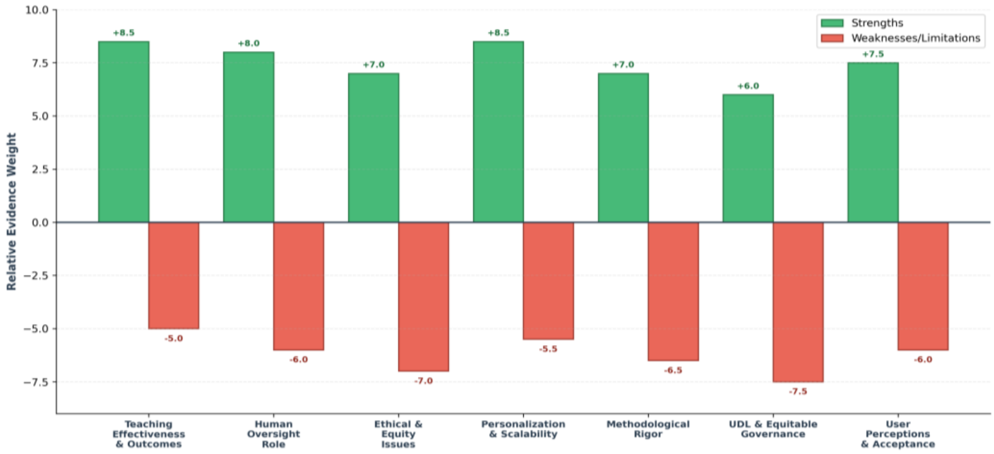

Figure 11: Strengths and Limitations Across Key Dimensions of AI-Powered Assessment Literature.

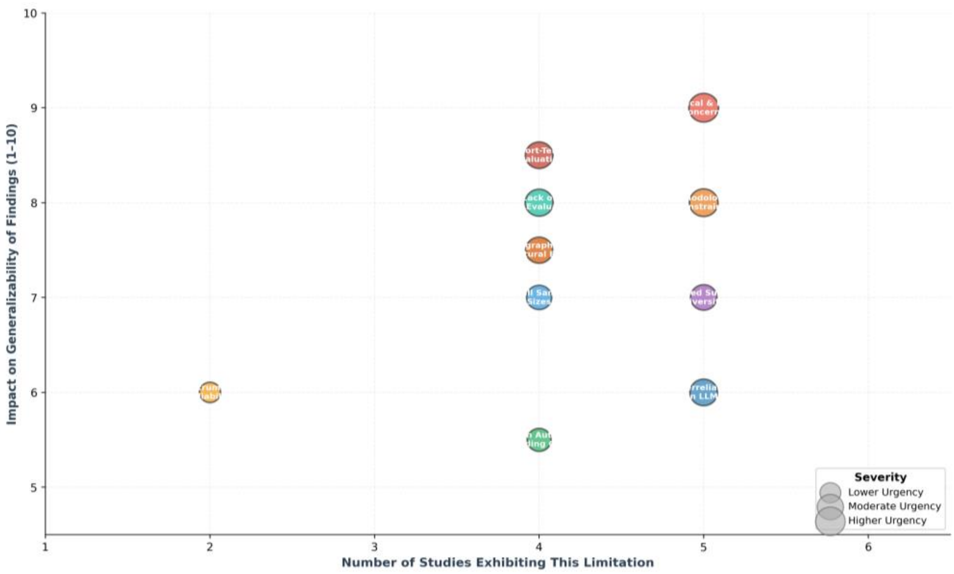

Figure 11: Identified Limitations in the Reviewed Literature (Bubble Size Represents Severity/Urgency).

They reflect deeper inequalities in education that AI can make worse if used without careful governance [25, 26, 27]. Combining Universal Design for Learning principles with AI assessment systems is a promising path toward fairer practice, but most contributions in this area are still theoretical rather than empirically tested [26, 63]. Methodological weaknesses across the reviewed literature, including small sample sizes, single-institution studies, and short study periods, limit how widely the findings can be applied and call for longer-term, cross-institutional research [41, 66].

5. Conclusion

This systematic review brought together evidence from 127 studies on AI-powered assessment and feedback in higher education. The findings show that AI tools, particularly large language models and adaptive learning systems, can improve teaching effectiveness by providing personalised, timely, and scalable feedback. As shown in Figure 12, several important research gaps remain that need to be addressed in future work, including longitudinal studies of AI assessment outcomes, research across a wider range of subject areas, practical testing of combined UDL-AI models, and the development of cross-institutional frameworks that can support fair AI-enhanced assessment across different educational settings. The reviewed evidence strongly supports hybrid human-in-the-loop models as the most educationally sound and fair approach, combining the efficiency of AI with the judgment, context awareness, and emotional sensitivity that human educators provide. Ethical concerns around algorithmic bias, data privacy, and lack of transparency need institutional governance frameworks and clear, inclusive policies that are put into practice rather than just described on paper.

Figure 12: Research Gaps and Future Directions Priority Matrix.

References

[1] D. Pike, "Students' perspectives of the change from human led to AI-led feedback - A review of Studiosity's writing feedback and writing feedback plus services," 2025.

[2] S. S. Lee and R. L. Moore, "Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review," Online Learning, vol. 28, no. 3, 2024.

[3] I. Kuzminykh, T. Nawaz, S. Shenzhang, B. Ghita, J. Raphael, and H. Xiao, "Personalised feedback framework for online education programmes using generative AI," 2024.

[4] T. Schultze, V. Kumar, G. McKeown, P. O'Connor, M. Rychlowska, and K. Sparemblek, "Using large language models to augment (rather than replace) human feedback in higher education improves perceived feedback quality," 2024.

[5] N. Yu, J. Zhang, S. Mitra, R. Smith, and A. Rich, "AI-Educational Development Loop (AI-EDL): A conceptual framework to bridge AI capabilities with classical educational theories," arXiv.org, 2025.

[6] P. Mund, "Artificial intelligence for assessment and feedback to enhance student success in

higher education," Mathematical Problems in Engineering, vol. 2022, pp. 1-19, 2022.
[7] P. Mund, "Artificial intelligence for assessment and feedback to enhance student success in higher education," Mathematical Problems in Engineering, vol. 2022, pp. 1-19, 2022.
[8] T. Schultze et al., "Using large language models to augment (rather than replace) human feedback in higher education improves perceived feedback quality," 2024.
[9] R. Stanyon, A. A. Tomlinson, M. Kainth, and N. K. Wilkin, "Providing individual student feedback at scale for mathematical disciplines," in Proc. ACM Conf., 2022.
[10] N. Abbas and E. Atwell, "Cognitive computing with large language models for student assessment feedback," Big Data and Cognitive Computing, vol. 9, no. 5, p. 112, 2025.
[11] C. Wangwiwattana and Y. Tongvivat, "Automating academic assessment: A large language model approach," in Proc. INCIT, 2023.
[12] D. Agostini and F. Picasso, "Large language models for sustainable assessment and feedback in higher education," Intelligenza Artificiale, pp. 1-18, 2024.
[13] S. Saad, Z. Ramli, S. M. Sum, and A. Salman, "A comparative analysis of AI-powered adaptive learning systems in higher education across developed countries," Int. J. Res. Innovation Social Sci., vol. IX, no. V, pp. 5877-5888, 2025.
[14] T. K. Vashishth et al., "AI-driven learning analytics for personalised feedback and assessment in higher education," Advances in AMEA Book Series, pp. 206-230, 2024.
[17] J. N. Lyanda, S. Owidi, and A. M. Simiyu, "Rethinking higher education teaching and assessment in line with AI innovations," African J. Empirical Res., vol. 5, no. 3, pp. 325-335, 2024.
[18] L. J. Jacobsen, J. Rohlmann, and K. E. Weber, "AI feedback in education: The impact of prompt design and human expertise on LLM performance," 2025.
[21] E. Er, G. Akcapinar, M. Khalil, O. Noroozi, and S. K. Banihashem, "Assessing student perceptions and use of instructor versus AI-generated feedback," British J. Educational Technology, 2024.
[22] P. Main, "AI and student assessment," Int. J. Sci. Res. Archive, vol. 16, no. 2, pp. 305-313, 2025.
[23] N. T. Amer, "From recall to reasoning," 2025.
[25] J. H. Christyodetaputri and N. Marwa, "Realising ethical and equitable assessment in global education through AI," Sinergi Int. J. Education, vol. 2, no. 3, pp. 170-186, 2024.
[26] M. A. A. Mitwally, L. Cheniti-Belcadhi, and A. Hadyaoui, "Universal design for learning and AI-enhanced assessment," 2025.
[27] D. L. Mpolomoka, "Utilizing artificial intelligence for assessment in higher education," Pedagogical Research, vol. 10, no. 3, e0243, 2025.
[29] M. Ragolane, S. Patel, and J. Phiri, "Navigating AI in assessment: Academic professionals' perceptions in South Africa," Technium Education and Humanities, vol. 11, pp. 156-172, 2025.
[30] H. A. Agbarakwe and O. O. Chibueze, "Leveraging AI for enhanced assessment in Nigeria higher education," Int. J. Res. Innovation Social Sci., vol. VIII, no. IX, pp. 142-151, 2024.
[32] M. Selvam and R. G. Vallejo, "Human-in-the-loop models for ethical AI grading," 2025.
[33] S. K. Vangibhuratha, "The impact of generative AI on educational assessment," 2025.
[34] D. DiSabito, L. Hansen, T. Mennella, and J. Rodriguez, "Exploring the frontiers of generative AI in assessment," New Directions for Teaching and Learning, 2024.
[36] C. Zhao, "AI-assisted assessment in higher education: A systematic review," J. Educational Technology and Innovation, vol. 6, no. 4, 2025.
[39] S. Ahmed, A. Zaki, and Y. Bentley, "AI and personalised grading criteria," Advances in ETID Book Series, pp. 85-113, 2024.
[40] S. Geschwind, J. G. Lambsdorff, D. Voss, and V. Hackl, "GPT-4 feedback increases student activation and learning outcomes," 2024.
[41] W. Dai et al., "Assessing the proficiency of LLMs in automatic feedback generation," 2024.
[42] A. Lee, "From traces to teaching: A socio-technical framework for process-based assessment," 2025.
[43] R. Gao, H. Merzdorf, S. Anwar, M. C. Hipwell, and A. R. Srinivasa, "Automatic assessment of text-based responses in post-secondary education," Computers & Education: AI, vol. 6, p. 100206, 2024.
[44] "Optimising AI writing assessment using feedback and knowledge graph integration," PeerJ, 2025.
[45] M. Morales-Chan et al., "Personalised feedback in MOOCs: Harnessing LangChain and OpenAI API," Electronics, 2024.
[46] R. Kumar et al., "Creating digital environment using data analytics and AI for evaluation," Int. J. Environmental Sciences, pp. 3435-3444, 2025.
[47] S. Kanchana and P. R. Saha, "Evolving student assessment: AI-driven rubrics for personalized and equitable English language learning," J. Engineering Education Transformations, vol. 38, no. IS2, pp. 584-590, 2025.
[49] C. Yeung et al., "A zero-shot LLM framework for automatic assignment grading," 2025.
[50] P. G. Policar, M. Spendl, T. Curk, and B. Zupan, "Automated assignment grading with LLMs," 2025.
[51] J. Floden, "Grading exams using LLMs: Human vs. AI grading using ChatGPT," British Educational Research Journal, 2024.
[52] S. H. Teckwani, A. H. Wong, W. A. N. V. Luke, and I. C. C. Low, "Accuracy and reliability of LLMs in assessing learning outcomes,"

Advances in Physiology Education, vol. 48, no. 4, pp. 904-914, 2024.
[53] T. Liu et al., "AI-assisted automated short answer grading of handwritten university-level mathematics exams," 2024.
[54] H. Alers, A. Malinowska, G. Meghoe, and E. Apfel, "Using ChatGPT-4 to grade open question exams," 2024.
[55] S. Wang, "The application and challenges of AI in personalised learning," 2025.
[56] Z. Slimi and B. Villarejo-Carballido, "Unveiling the potential: Experts' perspectives on AI in higher education," European J. Educational Research.
[58] E. C. Chen, "AIGC-driven teaching quality: A closed-loop feedback framework," 2025.
[59] M. Cooper-Stachowsky, "Enhancing learning via AI-generated feedback and resubmission of formative assessments," in Proc. CEEA Conf., 2024.
[61] A. McGowan, N. Anderson, and C. Smith, "Use of ChatGPT for assessment feedback on a complex programming assessment," 2025.
[62] "Intelligent enough? AI for online learners," J. Educators Online, vol. 20, no. 1, 2023.
[63] G. Ilieva, T. Yankova, M. Ruseva, and S. Kabaivanov, "A framework for generative AI-driven assessment in higher education," 2025.
[64] Y. Zhang and M. N. L. Y. Abdullah, "AI and student learning in higher education: Bibliometric and experimental investigation," Int. J. Innovative Res. Scientific Studies, vol. 8, no. 5, pp. 1616-1630, 2025.
[65] T. Heinrich et al., "AI-assisted grading and personalised feedback in large political science classes: Results from RCTs," PLOS ONE, vol. 20, no. 8, e0328041, 2025.
[66] E. R. P. Astuti and M. H. Baysha, "Evaluasi efektivitas sistem umpan balik berbasis AI," 2024.
[67] S. T. Chan, N. Lo, and A. Wong, "Leveraging generative AI for enhancing university-level English writing," Cogent Education, vol. 12, no. 1, 2024.