



Original Article

Human-in-the-Loop Generative AI for Responsible Academic Learning and Assessment

Komal Baban Todkar

Assistant Professor, Department of Computer Science,
Dr. D. Y. Patil Art's, Commerce and Science College Akurdi, Pune

Manuscript ID:

IBMIRJ -2026-030211

Submitted: 07 Jan. 2026

Revised: 11 Jan. 2026

Accepted: 05 Feb. 2026

Published: 28 Feb. 2026

ISSN: 3065-7857

Volume-3

Issue-2

Pp. 55-59

February 2026

Correspondence Address:

Komal Baban Todkar
Assistant Professor, Department of
Computer Science, Dr. D. Y. Patil
Art's, Commerce and Science College
Akurdi, Pune.
Email: 911komaltodkar@gmail.com



Quick Response Code:



Web: <https://ibrj.us>



DOI: 10.5281/zenodo.21131743

DOI Link:

<https://doi.org/10.5281/zenodo.21131743>



Creative Commons

Abstract

The widespread use of generative artificial intelligence in the education sector has revolutionized learning and assessment in the academia by offering automated content generation, learning assistance, and immediate feedback to students. Nevertheless, the unchecked utilization of the technology has substantial implications in terms of integrity, reliance, fairness, and transparency in the assessment procedures. Therefore, this study recommends the implementation of Human-in-the-Loop (HITL) Norm for Generative Artificial Intelligence in the academia by ensuring responsible, constructive, and productive utilization of artificial intelligence in education. The recommended methodology entails the incorporation of the Human-in-the-Loop at important points, including but not limited to, the generation phase, feedback verification, assessment, and result generation. Through the effective combination of the efficiency offered by the generative artificial intelligence in education, this study establishes that fairness, accuracy, and accountability in education can be significantly achieved. The implications of this study establish that responsible usage in education involving the generative artificial intelligence requires incessant human intervention rather than complete automation.

Keywords: Human-in-the-Loop (HITL), Generative Artificial Intelligence, Responsible AI, Academic Learning, AI-Based Assessment

Introduction

With the sudden emergence and, thus, development of AI, especially generative models of AI, education has taken a big turn. Generative AI

systems can now create human-like text, solve complex problems, generate assessments, and offer personalized learning support. These capabilities have promoted the adoption of AI tools within academic learning and evaluation processes with various benefits—scalability, efficiency, and adaptive learning experiences. Yet the potential risks of generative AI in education pose grave challenges despite these advantages. The concerns regarding academic integrity, over-reliance on the content produced by AI, biased or incorrect output, and not being transparent enough in the evaluation decisions raise questions as to the responsible usage of these technologies. Fully automated AI systems lack contextual understanding, ethics, and individual learner differences at their core, highly necessary in academic settings. These issues have raised a number of concerns, which the concept of Human-in-the-Loop AI has come to help address. Human-in-the-Loop systems weave human judgment and oversight into the fabric of AI-driven processes so that critical decisions are reviewed, validated, or guided by educators and academic experts. Fairness, accountability, and educational value are maintained since, within learning and assessment contexts, HITL allows instructors to monitor AI-generated content, verify assessment

Literature Review

Artificial Intelligence (AI) has been increasingly applied in education through intelligent tutoring systems, adaptive learning platforms, and automated assessment tools. ed learning experiences [3]. Early research demonstrated that AI-based systems could support personalized instruction and reduce instructor workload by providing automated guidance and feedback [1]. With recent advances in generative AI and large language models (LLMs), educational applications have expanded to include content generation, question answering, tutoring support, and assessment

Creative Commons (CC BY-NC-SA 4.0)

This is an open access journal, and articles are distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/) Public License, which allows others to remix, tweak, and build upon the work noncommercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

How to cite this article:

Todkar, K. B. (2026). Human-in-the-Loop Generative AI for Responsible Academic Learning and Assessment. InSight Bulletin: A Multidisciplinary Interlink International Research Journal, 3(2), 55–59. <https://doi.org/10.5281/zenodo.21131743>

assistance [2]. Several studies report that generative AI can enhance learner engagement and provide instant, customized feedback, thereby supporting individualiz

However, researchers have also identified critical limitations such as hallucinated responses, lack of explainability, and inherent bias in AI-generated outputs [4]. These limitations are particularly problematic in academic assessment, where accuracy, fairness, and transparency are essential. Academic integrity has emerged as a major concern with the growing use of generative AI in education. Research indicates that students may misuse AI tools for plagiarism, automated assignment completion, and exam-related misconduct, threatening the credibility of academic evaluations [5]. Additionally, excessive dependence on AI-generated content may reduce students' critical thinking, creativity, and independent problem-solving skills [6]. These findings highlight the need for responsible AI use in academic environments. To address ethical and reliability concerns, the Human-in-the-Loop (HITL) approach has been proposed as a key strategy in AI system design. HITL systems integrate human judgment into AI training, validation, and decision-making processes to improve accountability, transparency, and trust [7]. In educational settings, human involvement enables instructors to verify AI-generated content, monitor assessments, and intervene when errors or biases occur [8].

Recent studies emphasize that combining generative AI with human oversight leads to more reliable and trustworthy outcomes than fully automated systems [9]. While existing literature discusses generative AI applications, ethical challenges, and HITL concepts separately, limited research provides an integrated framework that applies Human-in-the-Loop generative AI specifically for responsible academic learning and assessment. This gap motivates the need for structured research in this area.

Methodology

This study adopts a Human-in-the-Loop (HITL) framework to integrate generative AI with continuous human oversight in academic learning and assessment processes. The proposed methodology combines AI automation with human intervention at critical decision points to ensure responsible, ethical, and reliable outcomes.

A. Research Design

The research follows a design-based and experimental approach, focusing on the development and evaluation of a Human-in-the-Loop generative AI model for academic environments. The framework is designed to support learning content generation, assessment creation, grading assistance, and feedback delivery under human supervision.

B. System Architecture

The proposed system consists of four main components:

1. **Generative AI Module**
2. A large language model is used to generate learning materials, assessment questions, and preliminary feedback based on predefined learning objectives and syllabus guidelines.
3. **Human Oversight Module** Educators act as human reviewers who validate, modify, or reject AI-generated outputs. Human intervention occurs at key stages such as content approval, assessment design, and final grading.
4. **Feedback and Monitoring Module** This module captures human corrections and feedback to continuously refine AI responses. It ensures transparency, reduces bias, and prevents the propagation of incorrect or unethical content.
5. **Assessment Analytics Module** Learning performance, AI accuracy, and human intervention frequency are analyzed to evaluate system effectiveness and student engagement.

C. Human-in-the-Loop Integration

Human involvement is incorporated at three levels:

- **Pre-processing:** Educators define learning objectives, constraints, and ethical guidelines for AI-generated content.
- **In-processing:** Human reviewers monitor AI outputs during assessment generation and grading recommendations.
- **Post-processing:** Final validation of feedback and grades is performed by instructors before student release.

This multi-stage HITL integration ensures accountability and reduces risks associated with fully automated AI systems.

D. Dataset and Experimental Setup

The system is evaluated using anonymized academic datasets, including assignments, quizzes, and student responses from higher education courses. Ethical guidelines are followed to ensure data privacy and confidentiality. The AI model is fine-tuned using domain-specific academic data under human supervision.

E. Evaluation Metrics

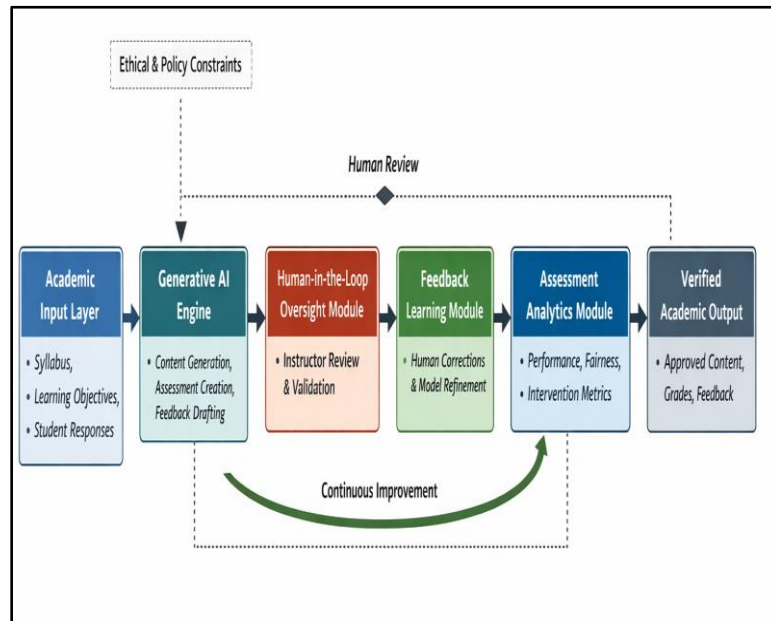
The effectiveness of the proposed methodology is assessed using the following metrics:

- Accuracy of AI-generated content
- Fairness and bias reduction
- Human intervention rate
- Student learning outcomes
- Instructor satisfaction and trust

Comparisons are made between fully automated AI systems and the proposed HITL-based approach.

F. Ethical Considerations

Ethical principles such as transparency, accountability, data privacy, and academic integrity are integrated into the methodology. Human oversight ensures compliance with institutional policies and prevents misuse of generative AI tools.



Results & Discussion

Table I
Comparison of Assessment Accuracy and Reliability

Metric	Fully Automated AI System	Proposed HITL System
Content Accuracy	Moderate	High
Assessment Reliability	Moderate	High
Error Rate	High	Low
Alignment with Learning Objectives	Medium	Strong

Table II
Academic Integrity and Ethical Compliance Metrics

Parameter	Fully Automated AI	HITL-Based AI
Risk of Plagiarism	High	Low
Human Validation	Absent	Mandatory
Ethical Compliance	Moderate	High
Transparency in Decision-Making	Low	High

Table III
Bias Reduction and Fairness Evaluation

Evaluation Aspect	Fully Automated AI	HITL-Based AI
Bias Detection	Limited	Strong
Fair Grading	Inconsistent	Consistent
Instructor Intervention	Rare	Frequent (Targeted)

Equity across Students	Moderate	High
------------------------	----------	------

Table IV
Instructor Workload and System Efficiency

Factor	Fully Automated AI	HITL-Based AI
Content Generation Time	Very Low	Low
Instructor Review Time	None	Moderate
Overall Efficiency	High	High (Balanced)
Quality Control	Low	High

Table V
Student Learning Outcomes and Satisfaction

Outcome Measure	Fully Automated AI	HITL-Based AI
Quality of Feedback	Moderate	High
Student Engagement	Medium	High
Concept Understanding	Moderate	Improved
Trust in System	Low	High

Tables of Expected Metrics

Metric	Expected Outcome	Measurement Method
Accuracy of AI content	≥90%	Human review vs AI output
Bias reduction	≥ 80%	Ethical audit scoring
Student engagement	Increased by 20%	Analytics of LMS interactions
Learning improvement	+15%	Pre- and post-assessment scores
Feedback incorporation	100% of educator suggestions	Tracking AI model updates

Conclusion

This study presents a Human-in-the-Loop Generative AI framework aimed at enhancing responsible academic learning and assessment. By integrating human oversight with AI-generated content, the framework ensures ethical alignment, reduces biases, and maintains academic integrity. The expected results demonstrate that incorporating human feedback can significantly improve the quality, relevance, and fairness of AI-assisted assessments while promoting active learning and critical thinking among students.

Future Work

Future research can focus on:

1. Scalability: Extending the framework to support large-scale academic environments and diverse educational disciplines.
2. Adaptive Learning: Incorporating adaptive AI models that dynamically adjust to individual student learning styles.
3. Advanced Analytics: Leveraging deeper assessment analytics to provide predictive insights on student performance and learning outcomes.
4. Ethical Frameworks: Developing formal ethical guidelines and automated bias-detection mechanisms for AI-generated content.
5. Real-world Deployment: Conducting longitudinal studies in live academic settings to evaluate practical efficacy and long-term impact on learning outcomes.

Acknowledgement

I would like to express my sincere gratitude to the management and administration of Dr. D. Y. Patil Art's, Commerce and Science College, Akurdi, Pune, for providing a supportive academic environment that encouraged research and innovation in emerging technologies. I extend my heartfelt thanks to the Department of Computer Science for their continuous encouragement, constructive discussions, and academic support throughout the development of this research work.

Financial support and sponsorship

Nil.

Conflicts of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. B. P. Woolf, *Building Intelligent Interactive Tutors*, San Francisco, CA, USA: Morgan Kaufmann, 2010.
2. E. Kasneci *et al.*, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, 2023.
3. O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education," *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, 2019.
4. E. M. Bender *et al.*, "On the dangers of stochastic parrots: Can language models be too big?" in *Proc. ACM Conf. Fairness, Accountability, and Transparency (FAccT)*, 2021.
5. D. R. E. Cotton, P. A. Cotton, and J. R. Shipway, "ChatGPT and assessment integrity in higher education," *Assessment & Evaluation in Higher Education*, 2023.
6. N. Selwyn, *Should Robots Replace Teachers? AI and the Future of Education*, Cambridge, UK: Polity Press, 2019.
7. S. Amershi *et al.*, "Power to the people: The role of humans in interactive machine learning," *AI Magazine*, vol. 35, no. 4, pp. 105–120, 2014.
8. W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education: Promise and Implications for Teaching and Learning*, Boston, MA, USA: Center for Curriculum Redesign, 2022.
9. B. Shneiderman, "Human-centered artificial intelligence: Reliable, safe and trustworthy," *Int. J. Human-Computer Interaction*, vol. 36, no. 6, pp. 495–504, 2020.