Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/22805
DOI URL: http://dx.doi.org/10.21474/IJAR01/22805

RESEARCH ARTICLE

A STUDY ON THE IMPLEMENTATION OF AI IN EXAMINATION AND ACADEMIC MODERATION

Parimala Mani[1] and Prasanna Kulkarni[2]

1. Controller of Examinations, Pimpri Chinchwad University, Pune, Maharashtra, India.
2. Professor, Department of Physics, Global Academy of Technology, Bangaluru, Karnataka, India.

Manuscript Info

Manuscript History
Received: 12 December 2025
Final Accepted: 14 January 2026
Published: February 2026

Key words:-
Artificial Intelligence, Academic Moderation, Automated Question Generation, Online Proctoring, Natural Language Processing, Educational Ethics, Human-in-the-Loop.

Abstract

The educational landscape is undergoing a fundamental transformation as institutions integrate Artificial Intelligence (AI) and Natural Language Processing (NLP) into assessment workflows. Traditional manual processes for question generation and academic moderation are increasingly hindered by human error, inconsistent difficulty levels, and faculty burnout. This study provides a comprehensive analysis of the implementation of AI tools in automating question paper design, enhancing examination proctoring, and refining grading through human-in-the-loop systems. By evaluating a taxonomy of tools ranging from Large Language Models (LLMs) to biometric proctoring systems, this paper identifies significant gains in efficiency—including reports of a 36% decrease in instructor workload—while addressing critical concerns regarding algorithmic bias, data sovereignty, and the "black-box" nature of automated decision-making. The findings suggest that while AI can revolutionize assessment, success depends on a hybrid model that maintains rigorous human oversight and ethical governance.

Introduction:-

Assessment remains the foundational pillar of education, providing evidence of student learning and ensuring curriculum alignment. However, the shift toward digital learning environments and large-scale online courses (MOOCs) has placed an unprecedented burden on instructors. Manual question paper setting is time-consuming and often lacks the cognitive diversity required by frameworks like Bloom's Taxonomy. Furthermore, the rise of generative AI tools like ChatGPT has created an "assessment crisis," with 56% of students admitting to using AI to complete assignments, thereby challenging traditional evaluation methods.AI and NLP offer transformative solutions by automating repetitive tasks such as grading multiple-choice questions, providing feedback on essays, and identifying learning gaps. Recent research indicates that faculty using AI daily are far more likely to refocus their time on high-impact student interactions. This study explores the technical methodologies, efficacy, and ethical frameworks necessary for responsible AI integration in examinations.

Corresponding Author:-Parimala Mani
Address:-Controller of Examinations, Pimpri Chinchwad University, Pune, Maharashtra, India.

Methodology:-

This study adopts a qualitative, analytical research approach to examine the implementation of Artificial Intelligence (AI) in examination systems and academic moderation. The methodology is designed to provide a structured understanding of how AI technologies are currently being integrated into assessment workflows, the benefits they offer, and the practical and ethical challenges associated with their use. Rather than relying on a single method, the study synthesizes existing scholarly research, documented institutional practices, and technological evaluations to build a comprehensive perspective on AI-enabled assessment.The research is primarily based on secondary data sources, including peer-reviewed journal articles, conference proceedings, institutional reports, technical documentation of AI assessment tools, and policy literature related to digital ethics and governance in education. These sources were selected to ensure credibility, relevance, and coverage across the different stages of the assessment lifecycle, including question generation, grading, proctoring, and academic moderation.

A thematic analytical framework guided the study. The collected materials were systematically reviewed and categorized into key functional domains**:** automated question generation using Natural Language Processing, AI-supported grading and feedback systems, online proctoring technologies, post-examination moderation and item analysis, and ethical and governance considerations such as fairness, transparency, and data privacy. This structure enabled a comparative examination of how AI tools operate across different institutional contexts and assessment tasks.In addition, a taxonomy-based evaluation was used to analyze various AI tools referenced in the study. These tools were examined based on their functional roles, implementation models, level of human involvement, and reported outcomes in educational settings. Particular attention was given to human-in-the-loop systems, where AI-generated outputs are reviewed and validated by educators, ensuring a balance between automation and pedagogical oversight.

The method of analysis involved qualitative synthesis and interpretive comparison of documented findings. Patterns relating to efficiency gains, workload reduction, consistency in grading, integrity monitoring, and accessibility improvements were identified and discussed. At the same time, recurring concerns—including algorithmic bias, the opacity of AI decision-making, data sovereignty, and ethical governance—were critically examined to provide a balanced perspective on AI adoption.Ethical considerations formed an integral part of the methodological approach. The study critically engaged with issues of student privacy, biometric data collection in proctoring systems, consent mechanisms, and the need for transparent and explainable AI models in academic decision-making processes.While the study does not involve primary data collection such as surveys or experimental trials, it provides a rigorous synthesis of current practices, technological developments, and institutional experiences. This approach enables a comprehensive understanding of the evolving role of AI in examination and academic moderation while also identifying areas requiring further empirical investigation. The methodology therefore supports both conceptual clarity and practical relevance, aligning with the study's objective of evaluating AI as a transformative yet ethically complex tool in educational assessment.

Taxonomy of Ai Tools In Assessment:-

AI tools for educational assessment are categorized by their functional role in the academic lifecycle:

- **Content Generation & Design:** Platforms like Magic School AI and Eduaide AI help educators draft quizzes, rubrics, and lesson plans aligned with national standards.
- **Grading & Feedback:** Tools such as Gradescope and Writable use OCR and AI to group similar student responses and provide scaffolded feedback on written work.
- **Integrity & Proctoring:** Systems like Honorlock and Proctorio utilize behavioral biometrics (eye movement, gaze direction) and room scans to safeguard exam integrity.
- **Learning Management Integration:** AI assistants like D2L Lumi are embedded directly into Learning Management Systems (LMS) to provide instant quiz feedback and automated grades.

Automated Question Generation (AQG) and Alignment:-

NLP Techniques for Keyword Extraction:-

Effective AQG relies on extracting significant concepts from course materials using NLP methods such as Term Frequency-Inverse Document Frequency (TF-IDF) and N-grams. By analysing lecture notes, these systems can identify keywords that capture the primary topics of a lesson, allowing for the generation of factually accurate Multiple-Choice Questions (MCQs). Validation studies show that auto-generated keywords align closely with those manually extracted by teachers, ensuring relevant examinable content.

Bloom's Taxonomy and Differentiated Design:-

Generative AI allows for the creation of questions that span the spectrum of cognitive demand. Educators can prompt AI to generate specific question sets for:

• **Remembering/Understanding:** Basic recall and explanation of concepts.
• **Applying/Analyzing:** Scenario-based prompts where students must apply knowledge to new contexts.
• **Creating:** Designing experiments or original solutions. This differentiation ensures that assessments do not merely test rote memorization but evaluate higher-order thinking skills.

AI In Online Proctoring And Integrity:-

Behavioral and Physiological Biometrics

Modern proctoring systems capture sensitive personal data, including physiological biometrics (face/voice characteristics, fingerprints) and behavioral biometrics (typing patterns, mouse movements, eye gaze). Systems like ProctorTrack provide 360-degree room scans and multi-angle monitoring to prevent unauthorized aid.

Automated Flags vs. Automated Decisions

A critical distinction in implementation is the level of autonomy granted to AI:

1. **Automated Flags:** Machine-generated alerts (e.g., detecting multiple faces or background voices) that are referred to a human proctor for review.
2. **Automated Decisions:** Algorithms that can unilaterally end an exam session or deny access based on detected behavior. While many systems (e.g., Honorlock) prioritize human review of flagged incidents, some (e.g., HirePro) aim to eliminate manual reviews through AI-powered auto-detection of fraud.

Automated Essay Scoring (AES) And Grading Systems:-

Transformer Models and Data Augmentation:-

Transformer-based models like BERT and RoBERTa have demonstrated high proficiency in Automated Essay Scoring (AES). Research indicates that these models significantly outperform traditional LSTMs. Data augmentation—such as inserting topic summaries into essays at regular intervals—encourages internal representations to align more closely with the subject matter, improving accuracy when evaluating essays on diverse topics.

Human-in-the-Loop Calibration

To address concerns over AI's ability to evaluate nuanced thinking, systems like Avalon utilize an iterative calibration phase. Instructors provide corrective feedback on small samples, training the AI to align with their specific expectations. Preliminary results show a 93% agreement rate between students and AI after calibration, reducing the subset of submissions requiring manual instructor review to fewer than 16%.

Academic Moderation And Personalized Feedback

Item Analysis and Quality Control:-

AI analytics support post-examination moderation by evaluating question effectiveness through two primary metrics:

• **Facility (Difficulty) Index:** Measures if a question was appropriately challenging.
• **Discrimination Index:** Distinguishes between high-performing and low-performing students. These data-driven insights allow moderators to identify "weak" questions for refinement or removal.

Student Self-Assessment and Engagement:-

Innovative grading frameworks require students to perform self-grading before viewing AI feedback. This process encourages students to revisit rubric criteria and reflect on their reasoning. Discrepancy reports—where students explain differences between their self-score and the AI score—have been shown to uncover hidden misconceptions and increase student engagement with feedback.

Ethical Considerations and Governance:-

Algorithmic Bias and Fairness:-

AI models are susceptible to biases embedded in historical training data, which can lead to the disproportionate penalization of students from specific socioeconomic or demographic backgrounds. Addressing these inequalities requires diverse dataset representation and the implementation of fairness-aware algorithms.

Transparency and the "Black-Box" Problem:-
The opaque nature of complex neural networks makes it difficult for stakeholders to scrutinize how a specific grade or proctoring flag was determined. Explainable AI (XAI) techniques, such as SHAP and LIME, are essential for breaking down AI predictions into understandable components for educators and students.

Data Sovereignty and Privacy:-
Online proctoring raises serious concerns about the invasive nature of data collection. The concept of data sovereignty suggests that students should own and control their proctoring profiles, granting explicit consent for the extraction of biometrics. Institutions must also clarify roles under GDPR, distinguishing between the university as the "data controller" and the AI provider as the "data processor".

Digital Accessibility and Future Directions:-
Enhancing Accessibility:-
AI-driven frameworks can automate the remediation of digital content to comply with standards like WCAG 2.1. These systems can auto-generate alt-text for images and adjust contrast, achieving a 92% accuracy rate in detecting accessibility barriers. This automation has improved user satisfaction for 94% of students with disabilities.

The Return of Oral Exams:-
To combat AI-assisted cheating in written work, many institutions are re-introducing oral exams (viva voce). Real-time interaction provides immediate authentication of knowledge that AI cannot easily simulate. Future assessment models will likely be "blended," combining AI-supported written tasks with authentic oral defense.

Field Observations:-
Field observations from institutions that have begun integrating AI into examination and academic moderation processes reveal a mixed yet evolving experience. In several cases, AI-supported tools were used to assist in grading objective assessments, detecting anomalies in student responses, and streamlining moderation workflows. Academic staff reported that these systems helped reduce manual workload and improved turnaround time in evaluation.
However, the implementation process was not without challenges. Observations indicated that faculty members initially expressed hesitation in trusting automated outputs, particularly in high-stakes assessments. There were also practical concerns related to data preparation, system compatibility with existing academic platforms, and the need for training to ensure effective use. In some instances, institutions adopted a hybrid approach, where AI recommendations were reviewed and validated by human moderators before final decisions were made.Overall, field-based insights suggest that while AI has the potential to enhance efficiency and consistency in examination and moderation practices, its successful adoption depends heavily on institutional readiness, staff engagement, and ongoing monitoring of system performance. These observations highlight the importance of combining technological innovation with contextual understanding and human oversight.

Practical Hurdles Institutions May Face When Adopting AI Systems:-
Despite the growing enthusiasm around artificial intelligence and its transformative potential, the institutional adoption of AI systems is rarely straightforward. In practice, organizations often encounter a range of structural, technical, financial, and cultural challenges that can significantly influence the success of implementation efforts.
One of the most immediate barriers is the limitation of existing infrastructure. Many institutions, particularly in the public and academic sectors, operate on legacy systems that are not designed to integrate with modern AI technologies. Upgrading hardware, digitizing records, and ensuring system interoperability can require substantial time and resources before AI solutions can even be piloted. Closely related to this is the issue of data readiness. AI systems depend on large volumes of high-quality, well-structured data, yet institutional data are often fragmented, inconsistent, or stored in incompatible formats, which complicates deployment and reduces model reliability.

Financial constraints also represent a major hurdle. The adoption of AI involves not only initial procurement costs but also long-term investments in system customization, staff training, cybersecurity, maintenance, and periodic upgrades. For many institutions, particularly those operating under tight budgets, demonstrating a clear return on investment remains a challenge and can slow decision-making processes.Human and organizational factors further complicate implementation. Workforce readiness is a critical concern, as staff may lack the technical skills required to effectively use and manage AI-driven tools. At the same time, resistance to change—driven by uncertainty, fear of job displacement, or lack of trust in automated systems—can impede adoption. Institutional culture, leadership support, and clear communication strategies play an important role in mitigating these challenges but are not always

adequately addressed. Ethical, legal, and governance issues also demand careful attention. Institutions must navigate complex regulatory environments related to data protection, privacy, accountability, and transparency. Concerns about algorithmic bias, fairness, and explainability can further slow adoption, particularly in sectors such as healthcare, education,and public administration where decisions directly affect individuals and communities.Finally, sustainability and scalability pose ongoing challenges. Implementing an AI system is not a one-time effort; it requires continuous monitoring, retraining of models, technical support, and alignment with evolving institutional needs. Without long-term strategic planning and governance frameworks, even well-designed AI initiatives risk underperformance or abandonment.Addressing these practical hurdles is essential for moving beyond conceptual discussions of AI and toward meaningful, responsible, and sustainable institutional adoption. A more grounded understanding of these barriers enables institutions to design implementation strategies that are realistic, context-sensitive, and better aligned with organizational capacities.

Conclusion:-

The implementation of AI in examination and academic moderation offers a paradigm shift in educational efficiency and standardization. While tools can reduce administrative burdens and provide personalized feedback, they cannot replace the pedagogical judgment of human educators.

Successful adoption requires:

1. A **Human-AI Hybrid Model** where AI serves as a "drafter" and humans as final validators.
2. **Robust Governance Frameworks** to ensure data privacy and mitigate bias.
3. **Open Standards** for proctoring data and flag transparency to empower students. Ultimately, AI must be viewed as a partner in advancing assessment design, ensuring that integrity and fairness remain at the heart of the educational experience.

References:-

1. M. Masudi, "The 5 Best AI Assessment Tools for Tutors," 5 AI Assessment Tools every Tutor should use in 2026, Feb. 2026.
2. Gradescope, "AI-powered feedback and assessment platform," 2026.
3. MagicSchool AI, "Comprehensive tool suite for teachers," 2026.
4. N. M. Jackson, "5 Ways to Implement AI Effectively into Faculty Work," EdTech Magazine, Apr. 2025.
5. E. Heinrich, "A Systematic-Narrative Review of Online Proctoring Systems and a Case for Open Standards," Open Praxis, vol. 17, no. 3, pp. 485–499, 2025.
6. H. Raza, "AI-Driven Assessment: Reliability, Bias, and Ethical Implications," AI Edify Journal, vol. 1, no. 2, 2024.
7. J. P. Costa et al., "AI-Driven Solutions for Enhancing Digital Accessibility in Higher Education," Eminent Association of Researchers (EAR), 2024.
8. A. Dhandapani, "AI-driven test automation for microservices," World Journal of Advanced Engineering Technology and Sciences, vol. 15, no. 2, pp. 1868-1881, 2025.
9. C. A. Nwafor and I. E. Onyenwe, "An Automated Multiple-Choice Question Generation Using Natural Language Processing Techniques," arXiv, 2024.
10. D. Armfield et al., "Avalon: A Human-in-the-Loop LLM Grading System with Instructor Calibration and Student Self-Assessment," Carnegie Mellon University, 2024.
11. K. Castagnoli, "Daily AI Use May Ease Faculty Burnout," Insight into Academia, Aug. 2025.
12. K. Gupta, "Data Augmentation for Automated Essay Scoring using Transformer Models," arXiv, 2024.
13. The Case HQ Online, "Powerful Guide to Writing Exam Questions Using Gen AI Effectively," Apr. 2025.
14. M. Park, "Top 10 Best AI Proctoring Software of 2026," Gitnux, Feb. 2026.
15. Burak Aydın,et al(2025), “Automated scoring in the era of artificial intelligence: An empirical study with Turkish essays” ,Systems, Vol 133,ISSN 0346-251X.
16. Haile Misgna, et all, Springer (2025),“An automated english essay scoring system based on deep learning and the internet of things”, Springer's Applied Intelligence journal.
17. Jessie S. Barrot , “Generative artificial intelligence for automated essay scoring: Exploring teacher agency through an ecological perspective”, Assessing Writing, Volume 67,2026,100990,ISSN 1075-2935, https://doi.org/10.1016/j.asw.2025.100990.