Cole-Puli International Conference Publishing Group

Business and Social Sciences Proceedings

Review

2025 2nd International Conference on Global Economics, Education and the Arts (GEEA 2025)

AI-Based Automated Grading Systems: Opportunities, Challenges, and Future Directions

Junyao Wang [1], Yasmin Hussain [1,*], Xijia Zhang [2], Chencheng Mao [1] and Zhao Wu [3]

[1] City Graduate School, City University Malaysia, Kuala Lumpur, Malaysia
[2] City Graduate School, Universiti SEGi Malaysia, Kuala Lumpur, Malaysia
[3] Shanghai Industrial and Commercial Polytechnic, Xinfeng North Road, Huaxin Town, Qingpu District, Shanghai, China
* Correspondence: Yasmin Hussain, City Graduate School, City University Malaysia, Kuala Lumpur, Malaysia

Abstract: Automated grading systems powered by artificial intelligence (AI) have significantly transformed education by enhancing efficiency, reducing grading biases, and providing instant feedback. This paper explores the opportunities, challenges, and future directions of AI-based automated grading systems. Opportunities include increased scalability, cost reduction, and enhanced student learning experiences. However, challenges such as bias in AI models, lack of human touch, ethical concerns, and data security issues remain prevalent. Future directions involve the integration of explainable AI, improved natural language processing (NLP) capabilities, and a focus on enhanced fairness and transparency. This paper concludes by emphasizing the need for interdisciplinary collaboration to optimize AI-driven grading solutions.

Keywords: AI grading; automated assessment; artificial intelligence; machine learning; education technology

Received: 02 April 2025
Revised: 11 April 2025
Accepted: 06 May 2025
Published: 08 May 2025

1. Introduction

Artificial Intelligence (AI) has become a transformative force in education, particularly in the realm of assessment and grading. AI-based automated grading systems utilize advanced technologies such as machine learning (ML) and natural language processing (NLP) to evaluate student work efficiently and with improved consistency. These systems are increasingly adopted in various educational settings, including standardized testing, online learning platforms, and higher education institutions, to streamline assessment processes and provide timely feedback.

The expansion of digital education and Massive Open Online Courses (MOOCs) has highlighted the limitations of traditional grading methods, which can be time-consuming and inconsistent, especially when dealing with large student populations. AI-based grading systems address these challenges by offering instant evaluation and standardized assessments across diverse student groups. Research indicates that these systems have the potential to enhance grading fairness, increase learning engagement, and support personalized education pathways.

However, the integration of AI in grading is not without challenges. Concerns regarding algorithmic bias, transparency, fairness, ethical considerations, and the absence of human judgment in nuanced assessments persist. For instance, automated grading sys-

tems may struggle to assess creativity, complex argumentation, and context-driven responses, leading to potential grading inaccuracies. Additionally, issues related to data security, student and educator resistance, and model errors raise questions about the reliability of AI in educational contexts.

This paper explores the opportunities, challenges, and future directions of AI-based automated grading systems, discussing how emerging AI technologies can optimize educational assessments while addressing critical concerns.

2. Opportunities of AI-Based Automated Grading Systems

AI-based automated grading systems have gained significant attention in recent years due to their potential to enhance assessment accuracy, reduce grading workloads, and provide immediate feedback to students. These systems leverage advanced machine learning (ML) and natural language processing (NLP) techniques to assess assignments, quizzes, and even complex essay responses with increasing precision [1]. The following sections explore the key opportunities offered by AI-based grading, including enhanced efficiency, cost reduction, reduced bias, real-time feedback, and scalability in large-scale online education [2].

2.1. Increased Efficiency and Scalability

One of the most significant advantages of AI-based grading systems is their ability to process large volumes of student submissions rapidly and consistently. Traditional human grading can be time-consuming, particularly in large courses, and may lead to inconsistencies due to grader fatigue. In contrast, AI-driven grading models can analyze large volumes of responses rapidly, allowing educators to focus more on instructional support rather than spending excessive time on assessment tasks [3].

In higher education institutions and online learning platforms, automated grading systems have been widely implemented to support Massive Open Online Courses (MOOCs), where thousands of students enroll in a single course. The ability of AI to provide instant assessment in such environments has revolutionized the educational landscape by ensuring timely feedback and progress tracking [4]. Moreover, in subjects requiring frequent evaluation—such as mathematics, programming, and language learning—AI-powered grading systems enable repeated practice and consistent feedback, which contribute to improved learning outcomes [5].

Furthermore, AI-driven grading is not limited to objective assessments such as multiple-choice questions. Advances in NLP and deep learning have enabled automated essay scoring (AES) systems to evaluate written assignments based on coherence, grammar, argument structure, and logical reasoning. These AI models are designed to improve through machine learning, aiming to align more closely with human evaluations over time.

2.2. Cost Reduction and Resource Optimization

Educational institutions spend significant resources on grading, including hiring teaching assistants and exam evaluators. AI-based grading systems help reduce operational costs by minimizing the need for additional grading personnel, particularly in large-scale courses. Universities and online education providers can reallocate resources to other academic support services, such as student advising, tutoring, and curriculum development [6].

From an administrative perspective, AI-driven assessment tools also streamline the grading process by reducing errors and logistical complexities. Traditional grading involves manual entry of scores, verification, and result compilation, which can be labor-intensive and prone to errors [7]. AI-based grading systems automate these steps, ensuring that scores are accurately recorded and processed without human intervention.

Additionally, the adoption of AI in grading enhances educator productivity. Teachers often spend a substantial portion of their workload on grading assignments and exams, limiting the time available for personalized instruction and student mentoring. AI can alleviate this burden, allowing educators to focus on fostering critical thinking, creativity, and interactive discussions rather than routine grading [8].

2.3. Reduced Bias and Improved Fairness

A key concern in traditional grading is the presence of grading inconsistencies and biases. Human evaluators, even when following structured rubrics, may unintentionally grade have based on subjective factors such as handwriting, personal biases, or emotional state. AI grading systems, when trained on diverse and representative datasets, offer a solution by ensuring objective and standardized evaluations across all student submissions.

Research suggests that AI-based grading can reduce racial, gender, and linguistic biases when properly implemented [9]. For example, human graders may unknowingly favor writing styles or dialects they are more familiar with, whereas AI can be trained to recognize and assess responses based on predefined evaluation criteria rather than personal preference. Furthermore, AI models continuously refine their assessment techniques based on extensive data analysis, making them more reliable over time [10].

However, it is essential to recognize that AI fairness is highly dependent on the training data used. If an AI grading system is trained on biased datasets, it may replicate and amplify existing biases rather than eliminate them. Therefore, continuous bias detection, dataset diversification, and algorithmic transparency are necessary to ensure equitable grading outcomes for all students.

2.4. Enhanced Learning Through Instant Feedback

AI-powered grading systems significantly enhance the learning experience by providing real-time feedback, enabling students to track their progress and make necessary improvements [11]. Unlike traditional grading, where students may wait days or weeks for feedback, AI-driven assessment tools allow learners to receive immediate evaluations, which is crucial for effective learning.

Immediate feedback can help students correct misunderstandings early, reducing the risk of reinforcing incorrect knowledge over time. This is particularly beneficial in subjects requiring progressive skill development, such as writing, coding, and problem-solving.

For example:

1) In essay writing, AI feedback can highlight grammatical errors, structural weaknesses, and argument inconsistencies, allowing students to refine their work before final submission.
2) In computer programming, AI-based grading can detect coding errors and suggest alternative solutions, helping students improve their coding logic.
3) In mathematics, AI-driven tools can break down step-by-step problem-solving processes, allowing students to identify specific mistakes.

Furthermore, AI-generated personalized feedback tailor's recommendations to individual student needs. By analyzing patterns in student responses, AI can provide targeted suggestions, such as additional resources, practice exercises, or concept explanations. This level of adaptive learning enhances student engagement and ensures that learning experiences are more tailored and effective [12].

2.5. Supporting Large-Scale Online Education

The increasing demand for online learning and remote education has amplified the need for scalable grading solutions. AI-based grading systems play a vital role in handling large-scale assessments in online courses, certification programs, and virtual universities.

For institutions offering online degrees or certification programs, AI-based grading ensures timely assessment of student progress without the logistical constraints of hiring graders in different time zones. Platforms such as Coursera, edX, and Khan Academy have adopted AI-based grading to manage thousands of student assessments efficiently. Additionally, AI tools are instrumental in evaluating open-ended responses, such as essay-based assignments and peer-reviewed coursework [13].

Beyond academic courses, another area where AI-based grading has proven beneficial is language learning and assessment. AI tools now power automated speech recognition and essay scoring systems that help non-native speakers improve their language proficiency. This enables learners worldwide to access high-quality education regardless of their location while ensuring fair and consistent grading [14,15].

2.6. Data-Driven Insights for Educators

Beyond grading, AI assessment tools generate valuable insights into student performance trends, enabling educators to make informed decisions about curriculum design and teaching strategies. By analyzing aggregate student performance data, AI can identify:

1) Common misconceptions across a class or cohort.
2) Areas where students struggle the most, allowing teachers to adjust instructional approaches.
3) Patterns in student progress, helping institutions improve curriculum pacing and instructional materials.

These insights can be integrated into learning analytics dashboards, which provide educators with real-time data on student engagement, strengths, and weaknesses. This empowers teachers to intervene early when students show signs of falling behind, ultimately improving learning outcomes.

3. Challenges of AI-Based Automated Grading Systems

Despite the numerous benefits of AI-based automated grading systems, several challenges hinder their effectiveness and widespread adoption. These challenges range from algorithmic bias, ethical and privacy concerns, and lack of human judgment, to technical limitations and resistance from educators and students. Addressing these issues is essential for ensuring that AI-driven grading systems are fair, transparent, and effective.

3.1. Algorithmic Bias and Fairness Issues

One of the most critical challenges in AI-based grading is algorithmic bias, where the system may favor certain groups of students while disadvantaging others. Since AI models are trained on historical datasets, they inherit the biases present in these datasets, leading to potentially unfair grading outcomes.

For example, research has found that AI essay grading systems may favor students who use more complex vocabulary and structured grammar, penalizing those whose writing style deviates from the norm [16]. This can disproportionately affect non-native speakers, students from underprivileged backgrounds, or those with unique writing styles [17,18]. Additionally, AI grading tools trained on datasets primarily composed of English-language essays may perform poorly when applied to multilingual or culturally diverse student populations.

Moreover, biases may emerge in subjective assessments. For instance, in creative subjects such as literature, AI might struggle to fairly evaluate originality, rhetorical strategies, and contextual meaning, resulting in inconsistent grading outcomes.

Key Issues:

1) AI may favor writing styles similar to its training data, disadvantaging non-traditional styles.
2) Non-native speakers and students from underprivileged backgrounds may receive lower scores.

3) AI struggles to evaluate creativity and context, leading to grading inconsistencies.
4) Lack of transparency makes it difficult to challenge unfair grading.

3.2. Lack of Human Judgment and Contextual Understanding

While AI excels at grading structured responses such as multiple-choice questions and objective short answers, it struggles with nuanced responses that require human interpretation. Many subjects — such as literature, philosophy, and social sciences — require an understanding of context, tone, and deeper meaning, which AI lacks.

For example:

1) Argumentative essays: AI cannot fully assess the persuasiveness of an argument or recognize rhetorical nuances.
2) Creative writing: AI often penalizes unconventional writing that does not follow standard grammar rules, even when intended for artistic effect.
3) Cultural and contextual understanding: AI models, trained primarily on Western datasets, may misinterpret cultural references, leading to biased assessments.

Additionally, exclusive reliance on data-driven metrics may result in assessments that overlook individual learning journeys, ignoring the personal effort, learning progress, or improvement of a student.

Key Issues:

1) AI struggles with subjective assignments, such as essays and creative works.
2) Lack of contextual understanding leads to misinterpretation of arguments and creativity.
3) AI does not account for personal effort or learning progression.
4) Over-reliance on AI may eliminate valuable human feedback that helps students grow.

3.3. Ethical and Privacy Concerns

AI grading systems require large amounts of student data, including writing samples, test scores, and learning behaviors. The collection and storage of this data pose serious ethical and privacy concerns [18].

Data Security Risks: AI grading platforms often store student responses on centralized servers, making them vulnerable to data breaches and cyberattacks.

Lack of Transparency: Many AI grading models function as "black boxes", meaning educators and students cannot see how scores are determined. This raises concerns about unjustified grading and inability to appeal AI decisions.

Surveillance and Student Profiling: Some AI grading systems track keystrokes, writing habits, and response patterns, raising concerns about student privacy and autonomy.

Bias in AI Governance: The increasing reliance on AI may disempower educators, shifting control over grading to private tech companies, whose algorithms may prioritize efficiency over fairness.

Key Issues:

1) Student data security risks due to AI-driven assessments.
2) AI grading lacks transparency, making it difficult to challenge scores.
3) Surveillance concerns arise from AI tracking student writing behavior.
4) Over-reliance on private AI models reduces educator control over grading.

3.4. Technical Limitations and Model Errors

Despite advancements in AI, automated grading systems still face technical challenges that can lead to incorrect or inconsistent grading. Since AI models rely on pattern recognition, they may fail to assess responses that deviate from expected patterns.

Common Technical Limitations:

1) Over-Reliance on Patterns: AI models are designed to identify structures and patterns, but they may struggle with unexpected or unique responses, marking them incorrect even when valid.
2) Grammar and Syntax Bias: Many AI grading tools prioritize grammatical correctness over logical argumentation, leading to unfair penalization of students with minor grammatical errors.
3) Handwriting Recognition Issues: AI grading is primarily optimized for typed responses and may struggle to process handwritten work, limiting its applicability in traditional exams.

Moreover, AI requires frequent updates to stay relevant. As language evolves and educational methods change, AI models trained on older data may become outdated, leading to inconsistent grading standards.

Key Issues:

1) AI struggles with unique or unconventional responses.
2) Grammar-heavy grading may penalize students for minor language errors.
3) AI cannot easily assess handwritten assignments.
4) Frequent updates are required to maintain grading accuracy.

3.5. Resistance from Educators and Students

The integration of AI-based grading has been met with resistance from educators and students, primarily due to concerns about fairness, reliability, and depersonalization.

Concerns from Educators:

1) Fear of Job Displacement: Many teachers worry that AI will replace human graders, reducing their role in assessment.
2) Loss of Control: AI grading limits educator discretion, making it difficult for teachers to exercise professional judgment when considering factors such as student effort, progress over time, or unique circumstances.
3) Doubt About Reliability: Many educators question whether AI can truly match human intuition in assessing essays and creative responses.

Concerns from Students:

1) Distrust in AI Fairness: Students often feel that AI grading lacks transparency and may not fairly assess creativity or complex ideas.
2) Reduced Personalized Feedback: AI-generated feedback tends to be generic and lacks the depth of personalized teacher comments.
3) Fear of "Automation Bias": Some students worry that AI may standardize responses too rigidly, discouraging originality.

Key Issues:

1) Educators fear AI replacing traditional grading roles.
2) AI limits teacher discretion in assessing student performance.
3) Students distrust AI fairness and question grading transparency.
4) AI lacks personalized feedback that human teachers provide.

While AI-based automated grading systems offer significant benefits, several challenges must be addressed to ensure their successful integration into education. Algorithmic bias, lack of human judgment, ethical concerns, technical limitations, and resistance from educators and students present substantial hurdles. Future developments should focus on improving AI transparency, reducing bias, integrating human oversight, and addressing privacy concerns. AI should be seen as a tool to assist educators rather than replace them, ensuring that grading remains fair, accurate, and supportive of student learning [19].

4. Future Directions

The rapid advancements in artificial intelligence (AI) and machine learning (ML) have significantly improved the efficiency and effectiveness of automated grading systems [20]. However, to fully realize the potential of AI-based grading while addressing existing challenges, several key areas require further research and development. These future directions include improving transparency and explainability, enhancing natural language processing (NLP) capabilities, integrating AI with human oversight, ensuring ethical AI development and fairness, expanding multimodal assessment capabilities, and developing adaptive learning feedback mechanisms. Addressing these areas will help build more reliable, ethical, and effective AI-driven assessment systems.

4.1. Improving Transparency and Explainability

One of the most critical issues in AI-based grading systems is their lack of transparency. Many existing AI grading models function as black boxes, generating scores without offering clear rationales. This opacity undermines trust among students and educators, making it difficult to challenge or learn from the grading process.

Key Areas for Improvement:

1) Explainable AI (XAI) in Grading: Future AI grading models should incorporate explainable AI (XAI) techniques, allowing students and educators to understand why a particular grade was assigned.
2) Grading Justifications: AI systems should provide detailed justifications for scores, highlighting areas of improvement and strengths in student responses.
3) Interactive Grading Reports: Implementing visualized grading analytics can help students track their progress and better understand AI evaluations.

Future research should focus on developing AI models that not only deliver accurate grades but also communicate their reasoning in an accessible manner.

4.2. Enhancing Natural Language Processing (NLP) for Subjective Assessments

While AI-based grading systems perform well on structured assessments like multiple-choice and short-answer questions, they still struggle with subjective and complex tasks, such as essay grading, creative writing evaluation, and argument analysis. The limitations of current natural language processing (NLP) models make it difficult for AI to assess nuance, context, creativity, and rhetorical strategies effectively.

Key Areas for NLP Improvement:

1) Better Contextual Understanding: Advanced NLP models should be trained to assess coherence, logical flow, and deep argumentation in essays.
2) Multilingual AI Grading: AI systems should be optimized to fairly assess responses in multiple languages, reducing bias against non-native speakers.
3) Creativity Recognition: Future AI grading tools must strengthen their ability to understand and evaluate originality and creative expression, areas where current rule-based models remain limited.

The integration of state-of-the-art NLP models (such as BERT, GPT-4, and future AI architectures) can improve AI grading accuracy for essay-based and subjective assessments.

4.3. Integrating AI with Human Oversight

Instead of replacing human graders, AI-based grading should be used as a supportive tool for educators, ensuring greater accuracy, fairness, and flexibility in assessments. A hybrid grading system that integrates AI automation with human oversight can minimize errors and preserve the role of educator judgment in grading complex assignments.

Key Future Developments in Human-AI Collaboration:

1) AI-Assisted Grading Platforms: AI should pre-grade assignments and provide initial assessments, which educators can then review and adjust.

2) Grading Moderation by Educators: Teachers should have the ability to override AI scores, especially for subjective or borderline cases.
3) Training AI with Human Feedback: AI models should learn from teacher adjustments, improving their future grading performance.

This human-AI collaboration will ensure that grading remains fair, transparent, and adaptable to individual learning needs.

4.4. Ensuring Ethical AI Development and Fairness

Addressing bias in AI-based grading systems is a crucial future direction. AI models can unintentionally replicate biases present in their training data, leading to unfair grading outcomes for students from diverse linguistic, cultural, and socioeconomic backgrounds.

Key Areas for Ethical AI Development:

1) Bias Auditing and Fairness Testing: AI grading models should undergo regular audits to ensure they do not favor certain student demographics over others.
2) Inclusive Training Datasets: Future AI models should be trained on diverse and representative datasets, ensuring fair assessment across different populations.
3) Ethical AI Guidelines in Education: Educational institutions and AI developers must establish clear ethical guidelines for the responsible deployment of AI grading systems.

Prioritizing ethical AI development will help future grading systems mitigate bias and strengthen trust within educational communities.

4.5. Expanding Multimodal Assessment Capabilities

Current AI grading systems are largely focused on text-based assessments, such as essays and multiple-choice questions. However, the future of AI-based grading should involve the ability to assess multimodal student work, including handwritten assignments, spoken responses, diagrams, and coding projects.

Future Developments in Multimodal AI Grading:

1) Handwriting Recognition: AI should be able to accurately analyze and grade handwritten responses, making it applicable to traditional pen-and-paper exams.
2) Speech and Oral Assessments: AI should be able to assess spoken language assignments, especially in subjects like foreign language learning and public speaking.
3) Diagram and Visual Content Evaluation: AI should be trained to evaluate diagrams, mathematical equations, and visual projects, expanding its utility in STEM education.

By supporting multimodal assessments, AI can become more inclusive and adaptable across different academic disciplines.

4.6. Developing Adaptive Learning Feedback Mechanisms

AI-powered grading systems should go beyond assigning scores — they should provide personalized and adaptive learning feedback to help students improve their skills.

Key Areas for AI-Driven Personalized Learning:

1) Dynamic Feedback Adjustments: AI should adapt its feedback based on student progress, ensuring that struggling students receive additional guidance and resources.
2) Real-Time Learning Recommendations: AI can suggest specific study materials, practice exercises, and personalized learning plans based on student weaknesses.
3) Progress Tracking Dashboards: AI should provide interactive learning dashboards, helping students track their improvements over time.

By integrating adaptive learning features, AI-based grading systems can evolve into intelligent tutoring assistants that actively support student learning.

4.7. Standardizing AI Grading Across Educational Institutions

Currently, AI grading standards vary widely across different platforms and institutions. Future research should focus on establishing universal grading frameworks, ensuring consistency, reliability, and fairness in AI-based assessments.

Key Steps for Standardization:

1) Developing Universal AI Grading Rubrics: AI grading systems should adhere to common educational frameworks, ensuring consistency in grading across different schools and universities.
2) AI Certification and Accreditation: AI-based grading models should undergo certification by educational authorities, ensuring their fairness and reliability before adoption.
3) Cross-Institution AI Data Sharing: Secure and anonymized data-sharing practices can help institutions develop better AI grading models through collaborative learning.

By implementing standardized grading criteria, AI-based assessment can become more trustworthy and widely accepted.

5. Conclusion

AI-based automated grading systems have emerged as a transformative force in modern education, offering significant advantages in efficiency, scalability, cost reduction, bias reduction, instant feedback, and support for large-scale online education. These systems leverage advanced machine learning (ML), natural language processing (NLP), and deep learning techniques to assess student responses rapidly and consistently. Their ability to process vast amounts of student submissions in real time makes them an invaluable tool, particularly in massive open online courses (MOOCs) and digital learning environments.

Despite these benefits, AI-based grading systems face substantial challenges, including algorithmic bias, fairness issues, ethical concerns, lack of human judgment, technical limitations, and resistance from educators and students. Algorithmic bias, in particular, remains a critical issue, as AI models may favor certain linguistic styles, cultural expressions, or socioeconomic backgrounds, leading to unfair grading outcomes for marginalized students. Furthermore, AI's inability to assess creativity, originality, and complex argumentation limits its effectiveness in grading essays, creative writing, and higher-order thinking assessments. Additionally, privacy concerns and ethical dilemmas related to data security, student surveillance, and lack of grading transparency necessitate stricter governance and regulatory frameworks.

Given these opportunities and challenges, future advancements in AI-based grading must prioritize fairness, transparency, explainability, and human-AI collaboration. Future research should focus on:

1) Enhancing Explainable AI (XAI): Ensuring AI models provide clear justifications for grading decisions so students and educators can understand and challenge automated scores when necessary.
2) Improving NLP for Subjective Assessments: Developing more sophisticated AI models that can interpret nuance, context, and creative expression, making them more suitable for grading essays, debates, and literature assignments.
3) Integrating Human Oversight with AI: Instead of fully automating the grading process, AI should act as an assistive tool for educators, enabling a hybrid grading system where teachers review and refine AI-generated assessments.

4) Ensuring Ethical and Fair AI Development: AI models must be trained on diverse and inclusive datasets to eliminate bias, and they should be regularly audited to maintain fairness and reliability.
5) Expanding Multimodal Assessment Capabilities: Future AI grading systems should support handwritten responses, spoken language assessments, diagram-based answers, and project-based learning, making them adaptable across various disciplines.
6) Developing Personalized Learning Feedback Mechanisms: AI-based grading should go beyond just assigning scores by offering adaptive, real-time feedback, helping students track their progress and improve in specific subject areas.

The integration of AI into education must be guided by ethical principles, robust governance frameworks, and continuous improvement through research and interdisciplinary collaboration. AI should not be seen as a replacement for human educators but rather as a complementary tool that enhances teaching efficiency, grading fairness, and student learning experiences.

Ultimately, the future of AI in education lies in creating balanced, human-centered AI systems that empower both educators and learners. By addressing technical limitations, fairness concerns, and ethical dilemmas, AI-driven grading can evolve into a powerful force for positive educational transformation, making learning more efficient, inclusive, and adaptive to individual student needs. However, realizing this potential requires a commitment from AI developers, educators, policymakers, and researchers to ensure that AI remains a trustworthy and equitable tool in education.

The future of AI-based automated grading lies in enhancing transparency, improving NLP capabilities, integrating human oversight, ensuring ethical AI deployment, supporting multimodal assessments, and developing adaptive learning mechanisms. AI should not replace human graders but rather augment their capabilities, making assessments more efficient, scalable, and personalized.

To achieve these advancements, continuous interdisciplinary research, ethical AI governance, and collaboration between educators, developers, and policymakers are necessary. AI grading must be fair, interpretable, and supportive of student learning, ensuring that it becomes an equitable and trusted tool in education.

The continued integration of AI in education requires a proactive approach to addressing both its benefits and its limitations. While AI-based grading holds immense potential to streamline assessments, enhance student feedback, and support large-scale learning, its effectiveness will ultimately depend on how responsibly it is designed, implemented, and regulated. Moving forward, the education sector must embrace AI with caution, ensuring that it remains fair, transparent, and aligned with the broader goal of fostering meaningful and equitable learning experiences for all students.

References

1. A. Gobrecht et al., "Beyond human subjectivity and error: a novel AI grading system," *arXiv Prepr. arXiv:2405.04323*, 2024, doi: 10.48550/arXiv.2405.04323.
2. C. Xiao et al., "Human-AI collaborative essay scoring: A dual-process framework with LLMs," in *Proc. 15th Int. Learn. Anal. Knowl. Conf.*, Mar. 2025, pp. 293–305, doi: 10.1145/3706468.3706507.
3. S. Sonkar, N. Liu, D. B. Mallick, and R. G. Baraniuk, "Marking: Visual grading with highlighting errors and annotating missing bits," in *Proc. Int. Conf. Artif. Intell. Educ.*, Jul. 2024, pp. 309–323. Cham, Switzerland: Springer. ISBN: 9783031643019.
4. A. Dimari et al., "AI-based automated grading systems for open book examination system: implications for assessment in higher education," in *Proc. 2024 Int. Conf. Knowl. Eng. Commun. Syst. (ICKECS)*, Apr. 2024, vol. 1, pp. 1–7, doi: 10.1109/ICKECS61492.2024.10616490.
5. Z. Mo, "AI-Assisted Strategies for Improving Chinese Proficiency in Non-Native AP Exam Takers", *J. Comput. Signal Syst. Res.*, vol. 2, no. 1, pp. 115–122, Mar. 2025, doi: 10.71222/4tksvc16.
6. I. Saputra, A. Kurniawan, M. Yanita, E. Y. Putri, and M. Mahniza, "The evolution of educational assessment: How artificial intelligence is shaping the trends and future of learning evaluation," *Indones. J. Comput. Sci.*, vol. 13, no. 6, 2024, doi: 10.33022/ijcs.v13i6.4465.

7. Q. Xia, X. Weng, F. Ouyang, T. J. Lin, and T. K. Chiu, "A scoping review on how generative artificial intelligence transforms assessment in higher education," *Int. J. Educ. Technol. High. Educ.*, vol. 21, no. 1, p. 40, 2024, doi: 10.1186/s41239-024-00468-z.
8. I. Gambo, F. J. Abegunde, O. Gambo, R. O. Ogundokun, A. N. Babatunde, and C. C. Lee, "GRAD-AI: An automated grading tool for code assessment and feedback in programming course," *Educ. Inf. Technol.*, pp. 1–41, 2024, doi: 10.1007/s10639-024-13218-5.
9. Z. Mo, "Artificial Intelligence in Lifelong Learning: Enhancing Chinese Language Instruction for Non-Native Adult Learners", *GBP Proc. Ser.*, vol. 2, pp. 141–146, Mar. 2025, doi: 10.71222/vxzcka39.
10. J. Gardner, M. O'Leary, and L. Yuan, "Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?'," *J. Comput. Assist. Learn.*, vol. 37, no. 5, pp. 1207–1216, 2021, doi: 10.1111/jcal.12577.
11. C. Figueras, A. Farazouli, T. Cerratto Pargman, C. McGrath, and C. Rossitto, "Promises and breakages of automated grading systems: a qualitative study in computer science education," *Educ. Inq.*, pp. 1–22, 2025, doi: 10.1080/20004508.2025.2464996.
12. S. Caspari-Sadeghi, "Artificial intelligence in technology-enhanced assessment: A survey of machine learning," *J. Educ. Technol. Syst.*, vol. 51, no. 3, pp. 372–386, 2023, doi: 10.1177/00472395221138791.
13. L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: A review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020, doi: 10.1109/ACCESS.2020.2988510.
14. M. Murtaza, Y. Ahmed, J. A. Shamsi, F. Sherwani and M. Usman, "AI-Based Personalized E-Learning Systems: Issues, Challenges, and Solutions," *IEEE Access*, vol. 10, pp. 81323-81342, 2022, doi: 10.1109/ACCESS.2022.3193938.
15. V. Suresh, R. Agasthiya, J. Ajay, A. A. Gold and D. Chandru, "AI based Automated Essay Grading System using NLP," in *2023 7th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Madurai, India, 2023, pp. 547-552, doi: 10.1109/ICICCS56967.2023.10142822.
16. Z. Mo, "AI-powered Language Learning: Empowering Non-Native Children to Master Chinese", *J. Linguist. Cult. Stud.*, vol. 2, no. 1, pp. 24–33, Feb. 2025, doi: 10.70088/ev8tgr31.
17. K. Ercikan and D. F. McCaffrey, "Optimizing implementation of artificial-intelligence-based automated scoring: An evidence centered design approach for designing assessments for AI-based scoring," *J. Educ. Meas.*, vol. 59, no. 3, pp. 272–287, 2022, doi: 10.1111/jedm.12332.
18. Z. Mo, "Artificial Intelligence Empowering Online Teaching of Chinese as a Foreign Language: Opportunities, Challenges, and Future Prospects", *Educ. Insights*, vol. 1, no. 5, pp. 18–29, Dec. 2024, doi: 10.70088/nwwqch86.
19. A. M. Naeem, A. Gad, M. Alhalabi, M. Yaghi, A. Khelifi and M. Ghazal, "AI-Based Mobile Paper Grading: Trends and Challenges," in *2023 10th Int. Conf. Future Internet Things Cloud (FiCloud)*, Marrakesh, Morocco, 2023, pp. 404-409, doi: 10.1109/FiCloud58648.2023.00065.
20. K. Twabu and M. Nakene-Mginqi, "Developing a design thinking artificial intelligence driven auto-marking/grading system for assessments to reduce the workload of lecturers at a higher learning institution in South Africa," *Front. Educ.*, vol. 9, 2024.