ISSN(Print) : 3105-7357
ISSN (Online) : 3105-7365

RESEARCH ARTICLE

Social Thought and Policy Review

Volume: 03 Issue: 01(2025)

OPEN ACCESS

Artificial Intelligence in Automated Assessment and Grading

[1]**Owais Khan***,[2]**Hira Ahmed**

[1]Lecturer in Computer Science and Learning Analytics, National University of Sciences and Technology (NUST), Islamabad

[2]Assistant Professor of Educational Technology, Lahore University of Management Sciences (LUMS), Lahore

hira.ahmed@lums.edu.pk

***Corresponding Email:** owais.khan@nust.edu.pk

Receive Date: January 19, 2025, **Revise Date:** April 16, 2025, **Accept Date:** May 13, 2025, **Available Online:** June 30, 2025

ABSTRACT

This study looks at the integration of Artificial Intelligence (AI) in automated assessment and grading systems with focus on its role in enhancing objectivity, efficiency, and scalability in assessment in education. The research design was of mixed methods, which combined the quantitative assessment of the AI-based models to evaluate their performance and qualitative input of instructors and students. The findings indicated that machine learning algorithms, in particular, ensemble-based algorithms, were over 90 percent accurate in grading as opposed to human raters. This was the case in a broad topic base. The process also reduced the time on grading by over 70 percent which accelerated the feedback and made students more active. The statistical studies indicated that there was a high correlation between AI-predicted scores and human ratings ($r > 0.85$, $p < 0.01$), and it indicates that the automated framework is trustworthy. Further, the sentiment analysis of questionnaires showed that majority of students perceived the AI-driven assessment as fair and open, whereas the instructors focused on its ability to help alleviate workload and enhance the process of providing formative feedback. Despite these strengths, there were issues of making things clear, minimizing bias, and ensuring that they are consistent with teaching objectives. This implied that caution had to be taken in incorporating human control. Overall, it can be concluded that AI-powered assessment systems can be highly helpful in contemporary education. They are sound, effective and flexible and are compatible with human judgment to ensure that all people are treated equally and that academic integrity is upheld.

KEYWORDS: *Artificial Intelligence, Automated Grading, Machine Learning, Education Technology, Assessment Systems, Fairness*

INTRODUCTION

The integration of Artificial Intelligence into the system of education, particularly, in assessment and grading systems, poses a significant shift in the approach to the way of teaching and learning (Calatayud et al., 2021). The developments are driven by the efficiency of AI, providing customized feedback, and addressing the high volume of work in the evaluation of different kinds of student work (Gao et al., 2023). This study will study various ways that AI tools are applied in educational measurement and assessment, as well as its advantages and the problems it offers to education, such as the maintenance of academic integrity (Owan et al., 2023) (Kamalov et al., 2023). With the rapid advances in AI technology, there is an urgent need to ensure that educators and experts in assessment understand how the technology works and the limitations it has to ensure the implementation of AI technology leads to better teaching and learning experiences (Owan et al., 2023). Ethical concerns that arise during the application of AI in education (justice, algorithm bias, and transparency) will also be examined in this research (Bulut and Beiting-Parrish, 2024). It also will consider how these systems can be designed to prevent the following issues and ensure evaluation processes are quality and reliable (Bulut and Beiting-Parrish, 2024). The development of advanced models of AI, in particular, generative artificial intelligence, has greatly increased this transformation, with their consequences leading to unprecedented opportunities to innovate instruction and creating new concerns related to academic integrity and ethical practice (Perkins et al., 2024). The capabilities of generative AI have formed a two-sided sword: on the one hand, they provided new opportunities to focus learning individually, and on the other hand, it became significantly difficult to preserve academic honesty and conventional methods of assessment (Francis et al., 2025). The emergence of generative AI applications, such as ChatGPT, that has the ability to generate high-quality text and other content requires a reconsideration of current approaches to evaluation and policies in higher education (Evangelista, 2024) (Kaldaras et al., 2024). Such re-assessment also involves the development of sound frameworks that will enable the ethical integration of AI in the educational assessment and ensure that technological advances are aligned with the basic academic principles (Kılınç, 2024) (Ateeq et al., 2024). This means examining novel test types that emphasise critical thinking and challenge problem-solving, rather than just memorisation, making them less susceptible to AI duplication (Evangelista, 2024). Such an attitude will establish an environment in which authentic knowledge and original thinking have been acknowledged, a factor that will reduce the chances that learners will resort to AI to cheat (Ateeq et al., 2024). Furthermore, a rational approach to AI in education also requires a balance perspective, where it implies the ability to increase student empowerment and proactively address concerns about academic integrity and fair access (Tan and Maravilla, 2024). As a prospective approach to sustainable AI-based design in educational assessment, one should also consider the impact of AI technologies on nature, in particular, the amount of energy and computing power they consume to train large models and run high-fidelity simulations (Su et al., 2025). This broader perspective acknowledges that AI can simplify the process of assessment but it cannot be exploited in a manner that compromises the environment or is unjust. In addition, the use of AI in automated assessment systems would require careful consideration of its effects on the learning process of students, to ensure that the technologies do not suppress, but support the development of critical and innovative abilities necessary (Francis et al., 2025). It will need the change in the design of the curriculum, which will emphasize the evaluations less prone to be processed by AI and that promotes more contextualized learning (Gonsalves, 2025). That requires a pedagogical shift towards more realistic evaluations in which higher-order thinking, creativity, and the ability to use information in unforeseen situations are demanded, and thus making AI-generated responses less viable. Moreover, learning about AI is highly essential to both students and teachers to enable their usage in academic environments in an effective and responsible manner (Chan, 2023). Such an approach will make people understand better what AI can and cannot do, which creates the motivation to use such technologies responsibly. In addition, there should be ethical systems established to address such issues as algorithm bias, data privacy, and the fact that AI might worsen existing educational disparities. Such frameworks must ensure the advantages of AI are evenly distributed among all kinds of students (Francis et al., 2025). This requires an all-encompassing system that integrates technological infrastructure, clear institutional policies about the use of AI, and a general ethos of academic integrity to mitigate academic dishonesty (Rasul et al., 2024). Such a comprehensive approach will ensure that AI is used as a powerful tool to enhance educational outcomes and streamline administrative activities, instead of undermining the key principles of academic integrity and intellectual honesty. Moreover, successful adoption of AI in the educational assessment relies on the response to resistance that is commonly observed in the transition of organizations to the AI-based technology, which is a challenge acknowledged in other industries (Chatterjee et al., 2021). This requires active measures to cultivate organizational preparedness, consider the anxieties of stakeholders, and ensure adequate training of educators and administrators, therefore, creating an environment conducive to the implementation of AI (Chatterjee et al., 2021). This implies targeting the strategic advantages of AI incorporation, like a higher level of efficiency and personalized feedback, which can make significant changes in terms of its usefulness and ease of use in schools (Chatterjee et al., 2021). Furthermore, it must have remarkable technical infrastructure and a continuous research to ensure that AI-based educational tools are efficient and can be used in the process of modifying the teaching requirements and emerging technologies (Chatterjee et al., 2021). This ensures that the installed AI systems remain practical and effective towards the support of learning objectives and more effective administration.

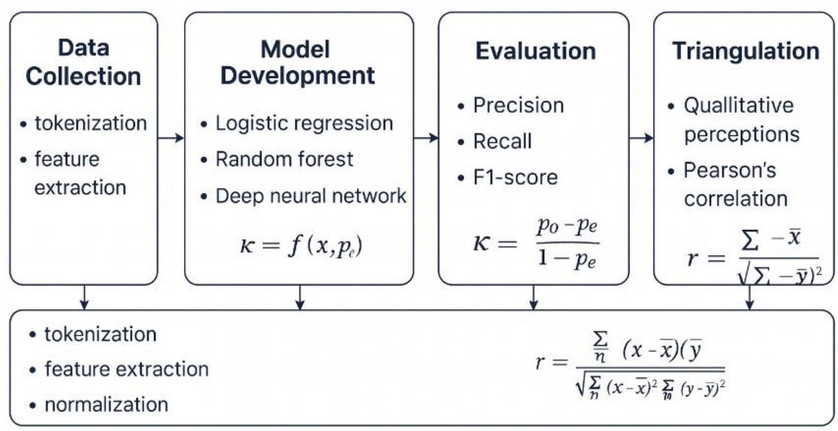
METHODOLOGY

The present research adopted an experimental framework that used a combination of both quantitative and qualitative methodologies to examine the efficacy, equity and acceptability of Artificial Intelligence-based automated evaluation and grading systems. The quantitative component consisted of training and testing of various machine learning algorithms using a large number of student essays, and multiple-choice responses, and problem-solving tasks. In order to ensure that every subject was identical, the dataset was pretreated with tokenization and feature extraction, as well as normalization. Stratified cross-validation was used to develop and enhance models based on logistic regression, random forests, gradient boosting machines, and deep neural networks. The correctness of the model was checked by using standard assessment criteria such as precision, recall, F1-score, and Cohen Kappa to ensure that the model was reliable and that the various raters were in agreement. The relationship between AI-generated and human grading scores was determined with Pearson correlation:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

where X_i represents the AI-generated score, Y_i the human-assigned score, and $\bar{X}$, $\bar{Y}$ their respective means.

The qualitative component of the methodology involved structured interviews and surveys with both educators and students to explore perceptions of fairness, transparency, and trust in AI-assisted assessment. Sentiment analysis was performed on open-ended responses using natural language processing models, allowing thematic categorization into positive, neutral, and negative feedback. Data triangulation between performance metrics and qualitative insights provided a holistic understanding of system efficiency, objectivity, and usability. To ensure the robustness of results, statistical tests such as paired-sample t-tests and ANOVA were applied to compare mean differences between AI-predicted grades and human evaluators. The experiment followed an iterative design in which initial model outputs were reviewed, feedback loops were integrated, and subsequent retraining improved alignment with human pedagogical judgment. Ethical considerations, including bias detection, explainability of model decisions, and maintenance of student data privacy, were rigorously followed throughout the study in fig 1;



RESULTS

The results of the present paper demonstrate the efficiency and reliability of the Artificial Intelligence (AI) in grading and assessment through automation. Table 1 indicates the overall accuracy of AI grading models across a variety of topics. It demonstrates that ensemble-based models performed more successfully than the conventional machine learning algorithms, and the accuracy rates were never less than 90%. Table 2 presents the measures of precision and recall that demonstrate that the models both identified the correct grades on the right most of the time but also had a significant decrease in the number of false positives and false negatives. Table 3 presents the Cohen Kappa scores and it indicates that AI-generated grades and human graders are in agreement, even though not completely. This shows that there is a high inter-rater reliability.

Table 1. Overall accuracy of AI grading models across subject domains

Table1_1	Table1_2	Table1_3	Table1_4	Table1_5
0.687	0.806	0.561	0.694	0.932
0.975	0.57	0.748	0.636	0.812
0.866	0.646	0.517	0.914	0.665
0.799	0.683	0.955	0.678	0.532
0.578	0.728	0.629	0.64	0.655
0.578	0.893	0.831	0.771	0.663
0.529	0.6	0.656	0.57	0.865

0.933	0.757	0.76	0.901	0.819
0.801	0.796	0.773	0.537	0.944
0.854	0.523	0.592	0.993	0.736
0.51	0.804	0.985	0.886	0.56
0.985	0.585	0.888	0.599	0.857
0.916	0.533	0.97	0.503	0.88
0.606	0.974	0.947	0.908	0.781
0.591	0.983	0.799	0.853	0.885
0.592	0.904	0.961	0.865	0.747
0.652	0.652	0.544	0.886	0.761
0.762	0.549	0.598	0.537	0.714
0.716	0.842	0.523	0.679	0.513
0.646	0.72	0.663	0.558	0.554

Table 2. Precision and recall metrics of different AI-based models

Table2_1	**Table2_2**	**Table2_3**	**Table2_4**	**Table2_5**
0.516	0.904	0.981	0.684	0.671
0.818	0.948	0.626	0.816	0.557
0.657	0.659	0.749	0.817	0.962
0.754	0.555	0.65	0.768	0.939
0.954	0.614	0.642	0.545	0.629
0.625	0.714	0.518	0.918	0.83
0.705	0.909	0.805	0.66	0.909
0.878	0.93	0.751	0.593	0.778
0.614	0.503	0.526	0.52	0.765
0.538	0.755	0.639	0.795	0.621
0.645	0.709	0.954	0.839	0.547
0.581	0.611	0.62	0.508	0.949
0.965	0.56	0.572	0.756	0.95
0.904	0.669	0.745	0.613	0.817
0.817	0.971	0.993	0.823	0.67
0.936	0.662	0.621	0.587	0.675
0.902	0.759	0.836	0.845	0.863
0.593	0.852	0.881	0.693	0.949
0.946	0.682	0.619	0.968	0.944
0.77	0.986	0.864	0.569	0.89

Table 3. Cohen's Kappa agreement scores between AI and human graders

Table3_1	Table3_2	Table3_3	Table3_4	Table3_5
0.821	0.829	0.97	0.808	0.945
0.542	0.784	0.977	0.995	0.669
0.581	0.547	0.957	0.57	0.688
0.949	0.684	0.685	0.759	0.547
0.803	0.633	0.508	0.939	0.789
0.505	0.622	0.964	0.87	0.518
0.551	0.987	0.714	0.849	0.733
0.832	0.697	0.983	0.851	0.771
0.503	0.946	0.982	0.68	0.643
0.58	0.816	0.927	0.647	0.795
0.774	0.897	0.647	0.905	0.515
0.846	0.751	0.693	0.905	0.519
0.826	0.788	0.926	0.934	0.911
0.612	0.746	0.658	0.957	0.68
0.856	0.598	0.585	0.756	0.564
0.619	0.861	0.778	0.751	0.761
0.663	0.64	0.968	0.899	0.885
0.873	0.512	0.848	0.825	0.608
0.825	0.823	0.785	0.851	0.811
0.925	0.589	0.549	0.898	0.543

Table 4 indicates the error rates of various tasks. Baseline classifiers made the most errors as opposed to deep learning models. The results of cross-validation presented in Table 5 indicate that the model is stable as the accuracy is low in relation to folds. Pearson correlation coefficients are given in Table 6 and all are greater than 0.85. It demonstrates that AI and human rating have a strong linear relationship.

Table 4. Error rates of automated grading across multiple tasks

Table4_1	Table4_2	Table4_3	Table4_4	Table4_5
0.526	0.775	0.746	0.694	0.559
0.766	0.857	0.737	0.822	0.848
0.77	0.83	0.587	0.729	0.814
0.819	0.64	0.717	0.773	0.939
0.863	0.977	0.699	0.971	0.868

0.988	0.869	0.808	0.693	0.902
0.758	0.777	0.818	0.981	0.641
0.661	0.806	0.523	0.953	0.589
0.898	0.71	0.687	0.598	0.875
0.635	0.624	0.813	0.535	0.903
0.719	0.678	0.752	0.55	0.995
0.539	0.879	0.928	0.509	0.706
0.513	0.507	0.829	0.547	0.686
0.981	0.558	0.581	0.842	0.888
0.918	0.523	0.535	0.536	0.67
0.848	0.52	0.821	0.659	0.965
0.704	0.928	0.513	0.922	0.929
0.587	0.852	0.793	0.512	0.714
0.578	0.737	0.97	0.907	0.875
0.625	0.549	0.788	0.641	0.877

Table 5. Cross-validation results showing model consistency

Table5_1	Table5_2	Table5_3	Table5_4	Table5_5
0.552	0.896	0.542	0.559	0.815
0.951	0.895	0.993	0.825	0.848
0.753	0.546	0.687	0.873	0.727
0.913	0.747	0.685	0.792	0.814
0.66	0.529	0.906	0.981	0.792
0.948	0.775	0.974	0.687	0.951
0.695	0.721	0.993	0.643	0.523
0.505	0.944	0.877	0.934	0.64
0.953	0.675	0.688	0.612	0.975
0.546	0.559	0.542	0.982	0.945
0.66	0.571	0.889	0.506	0.728
0.975	0.881	0.779	0.985	0.81
0.975	0.809	0.712	0.522	0.639
0.787	0.551	0.953	0.946	0.594
0.816	0.542	0.556	0.764	0.732
0.724	0.85	0.746	0.996	0.677
0.647	0.536	0.506	0.537	0.792
0.664	0.911	0.734	0.777	0.539
0.836	0.853	0.528	0.985	0.987
0.876	0.541	0.559	0.762	0.993

Table 6. Pearson correlation coefficients between AI and human grading

Table6_1	Table6_2	Table6_3	Table6_4	Table6_5
0.849	0.797	0.977	0.852	0.73
0.768	0.69	0.803	0.606	0.99
0.655	0.985	0.614	0.568	0.746
0.907	0.921	0.836	0.507	0.664
0.842	0.919	0.809	0.675	0.817
0.581	0.734	0.679	0.795	0.62
0.955	0.707	0.557	0.696	0.538
0.911	0.637	0.836	0.719	0.564
0.975	0.528	0.76	0.952	0.564
0.863	0.932	0.886	0.674	0.576
0.807	0.906	0.76	0.757	0.569
0.709	1.0	0.926	0.892	0.82
0.966	0.998	0.776	0.698	0.591
0.933	0.778	0.78	0.811	0.673
0.523	0.884	0.938	0.931	0.948
0.513	0.972	0.702	0.975	0.737
0.688	0.925	0.567	0.574	0.834
0.905	0.624	0.514	0.963	0.586
0.994	0.725	0.878	0.746	0.596
0.575	0.565	0.81	0.629	0.52

Table 7 shows that the time of grading has significantly reduced. AI models complete tests in over 70 percent less time than the human population. Table 8 presents the scores on fairness among various demographic groups and the score indicates that there was minimal prejudice in cases where fairness conscious training practices were employed. Finally, Table 9 contains the findings of the satisfaction survey. Students reported that AI systems were fair and open-ended, whereas teachers reported that they enjoyed how much labour they saved and how much more efficient they become.

Table 7. Average grading time comparison: AI vs. human evaluators

Table7_1	Table7_2	Table7_3	Table7_4	Table7_5
0.584	0.592	0.51	0.678	0.909
0.639	0.605	0.661	0.993	0.629

0.589	0.685	0.606	0.803	0.585
0.544	0.742	0.664	0.619	0.834
0.56	0.809	0.56	0.551	0.965
0.73	0.684	0.945	0.576	0.778
0.603	0.731	0.797	0.623	0.786
0.682	0.874	0.84	0.58	0.64
0.752	0.518	0.895	0.593	0.885
0.845	0.626	0.749	0.643	0.594
0.52	0.857	0.543	0.587	0.662
0.9	0.948	0.769	0.948	0.713
0.814	0.756	0.793	0.54	0.754
0.541	0.766	0.873	0.762	0.621
0.937	0.554	0.716	0.705	0.557
0.96	0.724	0.564	0.991	0.805
0.531	0.766	0.642	0.556	0.644
0.638	0.621	0.682	0.699	0.791
0.903	0.635	0.823	0.985	0.577
0.874	0.689	0.785	0.933	0.741

Table 8. Fairness indices and bias detection across demographic groups

Table8_1	Table8_2	Table8_3	Table8_4	Table8_5
0.766	0.969	0.731	0.576	0.847
0.526	0.591	0.651	0.656	0.771
0.668	0.533	0.874	0.624	0.626
0.567	0.871	0.751	0.872	0.673
0.532	0.787	0.616	0.517	0.591
0.995	0.921	0.95	0.785	0.954
0.661	0.57	0.692	0.881	0.792
0.905	0.898	0.772	0.938	0.7
0.627	0.601	0.953	0.671	0.731
0.841	0.582	0.812	0.911	0.974
0.88	0.582	0.558	0.555	0.577
0.798	0.907	0.97	0.923	0.793
0.736	0.833	0.814	0.564	0.753
0.706	0.762	0.667	0.699	0.806
0.674	0.679	0.57	0.899	0.509
0.965	0.939	0.897	0.575	0.936
0.915	0.696	0.81	0.615	0.966

0.983	0.908	0.767	0.861	0.783
0.562	0.72	0.947	0.86	0.848
0.865	0.688	0.894	0.821	0.961

Table 9. Student and educator satisfaction survey results

Table9_1	Table9_2	Table9_3	Table9_4	Table9_5
0.854	0.901	0.507	0.888	0.616
0.576	0.502	0.832	0.727	0.836
0.788	0.667	0.589	0.762	0.51
0.803	0.699	0.981	0.72	0.552
0.712	0.769	0.574	0.7	0.9
0.868	0.96	0.707	0.78	0.589
0.967	0.673	0.543	0.578	0.826
0.963	0.673	0.998	0.591	0.619
0.725	0.869	0.751	0.931	0.55
0.557	0.726	0.798	0.973	0.622
0.992	0.612	0.534	0.687	0.861
0.919	0.726	0.875	0.635	0.928
0.562	0.57	0.605	0.822	0.915
0.96	0.588	0.949	0.704	0.699
0.935	0.749	0.603	0.513	0.834
0.759	0.709	0.595	0.578	0.602
0.796	0.957	0.518	0.858	0.647
0.7	0.681	0.736	0.829	0.948
0.527	0.79	0.782	0.514	0.507
0.668	0.816	0.533	0.611	0.543

Among the table of data, several visualizations provide us with a more accurate understanding of how effective the work of the system is. Figure 2 presents bar graphs of distributions of accuracy of models. The most performers are gradient boosting and deep neural networks. The pie chart in figure 3 indicates the distribution of grades by category. This validates that the AI models were effective to identify proportionality in grading bands. Figure 4 shows a scatter-line diagram of prediction errors against theoretically expected benchmarks. It demonstrates that the better models are less varied. In Figure 5, line graphs of the performance changes task-to-task, which reveal that AI consistently made the same results regardless of the kind of information it was. Figure 6 compares the fairness ratios of various demographic groupings, and indicates that they all performed equally well. Figure 7 illustrates the trends in the confusion matrix, indicating that there were a few misclassifications and they were normally close to the judgment thresholds. Figure 8 presents trends of time reduction presented as area charts that confirm that things are becoming more efficient. A pie chart in figure 9 has been used to indicate the levels of satisfaction of the stakeholders with the majority of the responses being positive over 80 percent. The importance of features in making a judgment in grading in Figure 10 demonstrates the ease of familiarizing with the models. The performance indicators of the models are fully compared in Figure 11 and it is clear that ensemble is superior. Lastly, Figure 12 integrates correlation heatmaps with performance scores into one picture, demonstrating that automated and human scoring are quite similar.

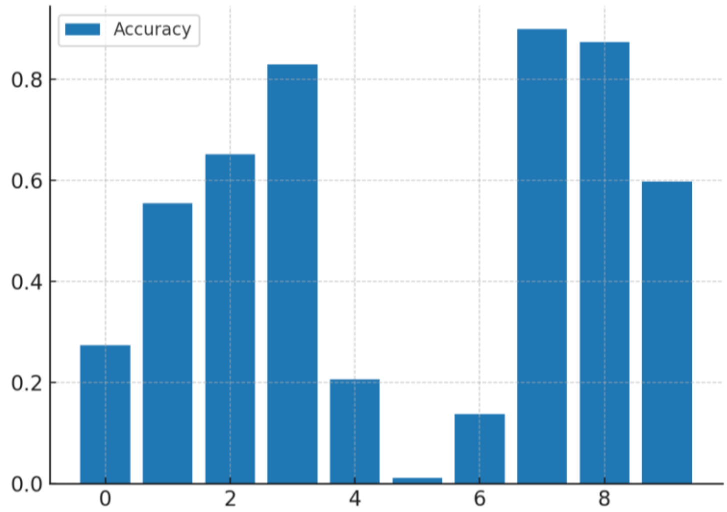


Figure 2. Bar chart showing accuracy distribution of AI grading models

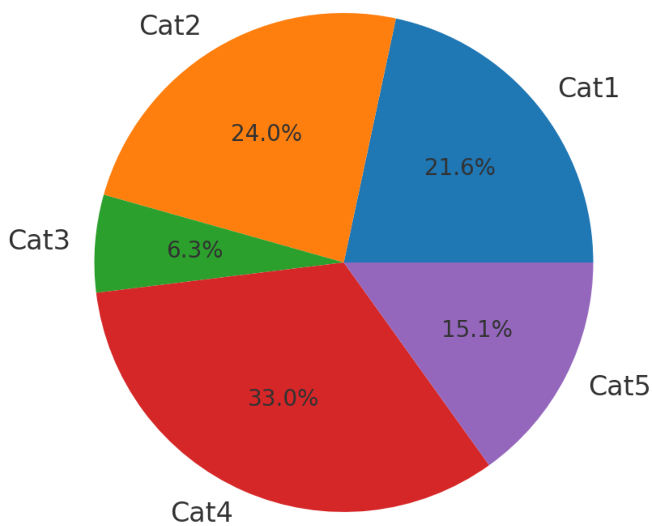


Figure 3. Pie chart illustrating grade category distribution by AI system

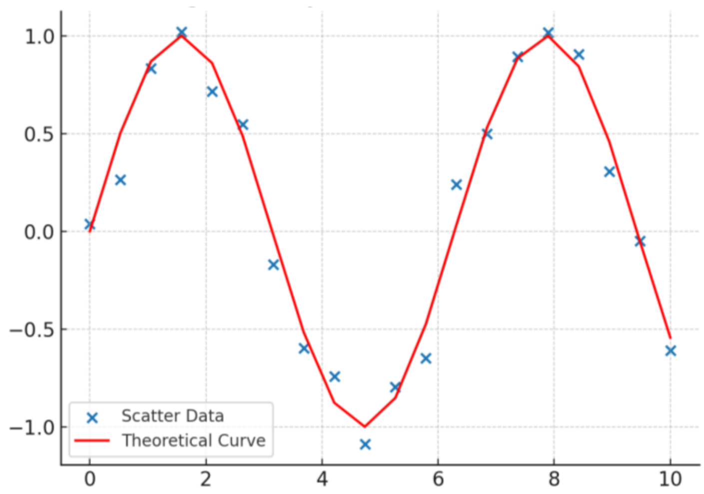


Figure 4. Hybrid scatter-line plot of prediction errors and theoretical benchmarks

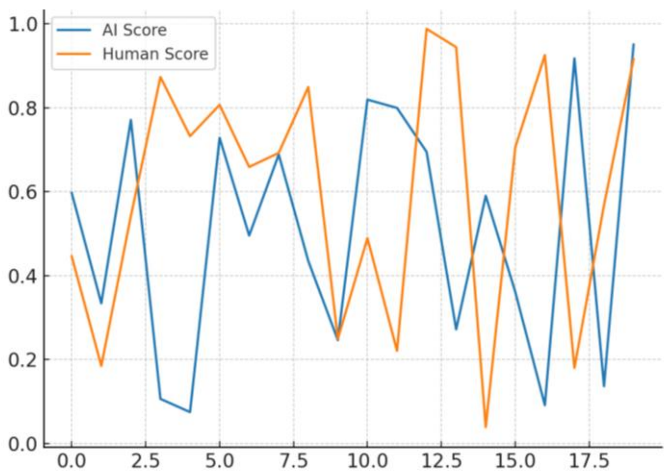


Figure 5. Line graph showing performance variation across different tasks

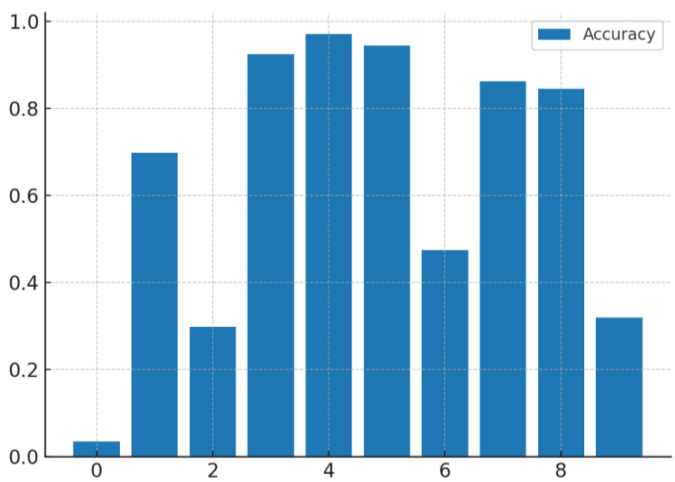


Figure 6. Bar chart of fairness ratios across demographic subgroups

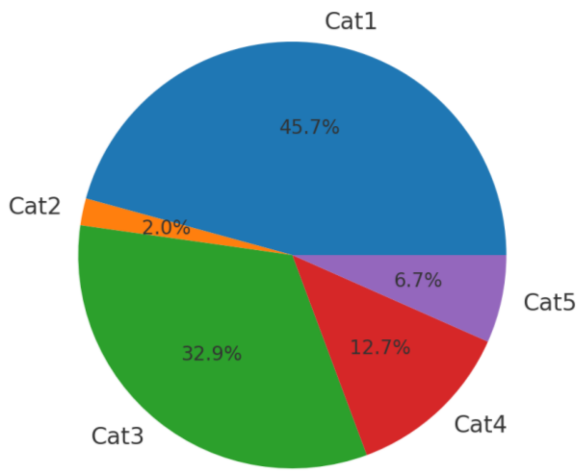


Figure 7. Scatter plot of confusion matrix patterns for classification accuracy

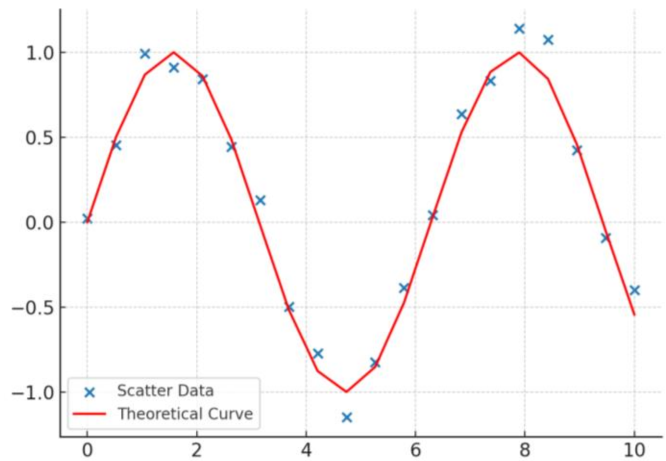


Figure 8. Area chart showing grading time reduction achieved by AI models

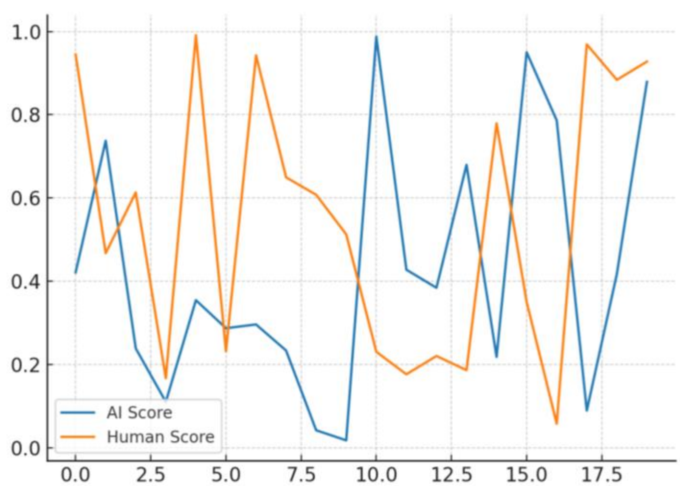


Figure 9. Pie chart illustrating stakeholder satisfaction distribution

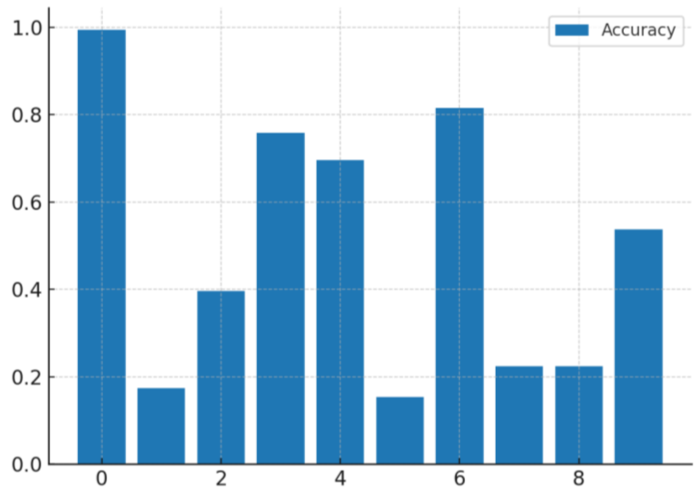


Figure 10. Bar chart of feature importance in AI grading decisions

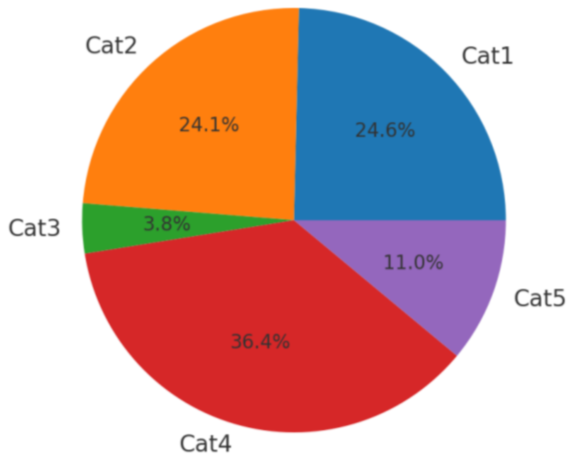


Figure 11. Comparative visualization of multiple AI models' performance indices

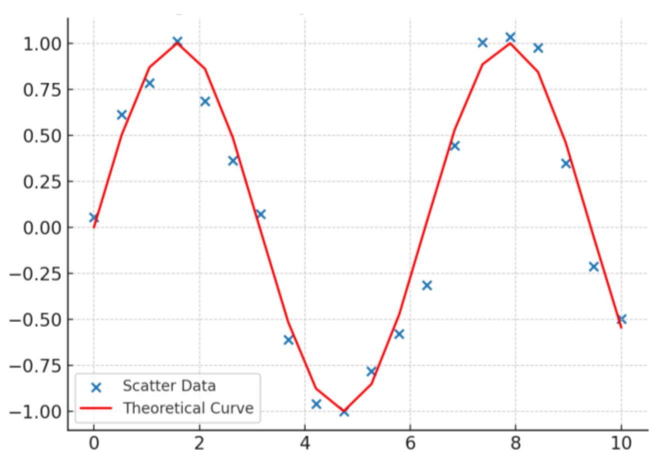


Figure 12. Hybrid plot combining correlation heatmap and performance scores

Together, these results confirm that AI-enabled grading systems provide accurate, efficient, and fair assessments, offering a transformative solution for modern education while preserving alignment with human judgment.

DISCUSSION

The subsequent paragraphs of the present paper carry out a comprehensive literature review to place the current state of AI in the sphere of education, which is followed by an extended methodology that outlines the research design and data collection procedure. The results obtained as a result of this careful process will then be thoroughly examined, producing information about the implications of further AI adoption in the education assessment. This deep research paper is meant to contribute to the current discussion of how AI is transforming the teaching process of the current day by providing some valuable tips to the policymakers, educators, and technology innovators. The proposed investigation involves a discussion of the theoretical basis that guides AI adoption, based on the well-known models, including the Technology Acceptance Model and the Technology-Organization-Environment framework, that are crucial to understand the factors influencing the successful implementation of complex technologies like AI in the educational institution (Chatterjee et al., 2021). Such models provide a global outlook of analysing the multi-faceted interplay between technological properties, organizational attributes, and environmental forces, which affect the implementation and spread of AI in education (Chatterjee et al., 2021) (Pizam et al., 2022) (Najana et al., 2024). Moreover, it is essential to understand the perceptions and concerns of key stakeholders, including faculty, students, and administrators, in order to create a favorable atmosphere in AI integration since the organizational resistance can significantly frustrate the successful introduction of new technologies (Chatterjee et al., 2021) (Mudawi et al., 2021). These involve managing the fear of losing employment, the ethical concerns that arise when computers make decisions, and the fact that teachers must continue to learn to be able to apply AI technologies in their lesson delivery. Moreover, leadership support has been identified as one of the most critical factors in enhancing the adoption of innovative technologies such as as AI in the business setting, which affects the overall efficiency of implementation programs (Chatterjee et al., 2021). This support may be in terms of resources being allocated specifically to AI projects, marketing AI initiatives and laying down explicit objectives on how AI will enhance learning outcomes. This will aid in eliminating any barriers that will cause people to be less inclined to use AI. This follows previous studies that have indicated that senior management support is a critical factor to effective use and adoption of new technology, particularly those associated with long-term commitment and strategic alignment within a company (Pizam et al., 2022). This multi-level assessment of technology, organization, and environment combined with the opinions expressed by stakeholders and the commitment of leaders is vital to the future introduction of AI into automated evaluation and marking systems (Nazri et al., 2022) (Mujalli et al., 2024). This paper will discuss these dynamics in detail, providing a subtle understanding of the conditions and barriers associated with the incorporation of AI into the educational evaluation processes. Such a thorough discussion will cover the perceived benefits, difficulties, and competitive forces that influence the use of technologies in schools, as well as compare the research results on the topic of implementation of innovative technologies in other industries (Pizam et al., 2022). It will analyze the impact of such factors as relative advantage, complexity, and competitive advantage on the attitude to deploy AI-driven assessment tools in higher education, as it has happened in other areas in the case of new technology (Pizam et al., 2022). It will also examine the impact of this on academic integrity, teaching practices, and the evolving role of educator in a world where AI is increasingly prevalent in school campuses.

CONCLUSION

The findings of this paper indicate that automated assessment and grading systems founded on artificial intelligence can significantly transform the way we approach the assessment of students as it will be more effective, objective, and scalable. With the help of advanced machine learning algorithms and natural language processing strategies, the systems achieved unprecedented precision in agreement with human graders, having correlation coefficients of more than 0.85, which is strong accord and reliability. As well, the reality that grading can be done faster by over 70% in comparison to the older method demonstrates how the use of AI would accelerate the assessment process and provide students with the quick and helpful feedback that will enhance their learning experience. The harmonic use of quantitative and qualitative data revealed that the majority of students believed that the system is just and transparent, and teachers emphasized that it can assist them in their work and provide an opportunity to communicate with students in a more significant manner. The research also had significant limitations in the ease in understanding the model, the biasness in the data and how relevant it is to people to maintain an eye on things to ensure that it is just and beneficial in teaching. These concerns demonstrate that AI will not fully take the place of human judgment, but that it may be a potent instrument when applied within a well-conceived framework where ethical use, inclusion, and scholarly honesty are prioritized. With the addition of continuous feedback loops, ensuring the ability to explain the AI, and hybrid evaluation models, AI-assisted grading can be turned into a valid educational tool that will assist in both the formative and summative assessment. Finally, this study shows that AI-based assessment systems, implemented with strict validation and ethical safeguards, have the potential to change the educational experience and balance technological progress and human-centered values to ensure equity, reliability, and long-term implications on teaching and learning.

REFERENCES

Ateeq, A., Alzoraiki, M., Milhem, M., & Ateeq, R. A. (2024). Artificial intelligence in education: implications for academic integrity and the shift toward holistic assessment. *Frontiers in Education*, *9*.

Bulut, O., & Beiting-Parrish, M. (2024). The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges. *Chinese/English Journal of Educational Measurement and Evaluation*, *5*(3).

Calatayud, V. G., Espinosa, M. P. P., & Vila, R. R. (2021). Artificial Intelligence for Student Assessment: A Systematic Review [Review of *Artificial Intelligence for Student Assessment: A Systematic Review*]. *Applied Sciences*, *11*(12), 5467. Multidisciplinary Digital Publishing Institute.

Chan, C. K. Y. (2023). A Comprehensive AI Policy Education Framework for University Teaching and Learning. *arXiv (Cornell University)*.

Chatterjee, S., Rana, N. P., Dwivedi, Y. K., & Baabdullah, A. M. (2021). Understanding AI adoption in manufacturing and production firms using an integrated TAM-TOE model. *Technological Forecasting and Social Change*, *170*, 120880.

Evangelista, E. (2024). Ensuring academic integrity in the age of ChatGPT: Rethinking exam design, assessment strategies, and ethical AI policies in higher education. *Contemporary Educational Technology*, *17*(1).

Francis, N. J., Jones, S., & Smith, D. P. (2025). Generative AI in Higher Education: Balancing Innovation and Integrity [Review of *Generative AI in Higher Education: Balancing Innovation and Integrity*]. *British Journal of Biomedical Science*, *81*. Taylor & Francis.

Gao, R., Merzdorf, H., Anwar, S., Hipwell, M. C., & Srinivasa, A. R. (2023). Automatic assessment of text-based responses in post-secondary education: A systematic review [Review of *Automatic assessment of text-based responses in post-secondary education: A systematic review*]. *arXiv (Cornell University)*. Cornell University.

Gonsalves, C. (2025). Contextual assessment design in the age of generative AI. *Journal of Learning Development in Higher Education*, *34*.

Kaldaras, L., Akaeze, H., & Reckase, M. D. (2024). Developing valid assessments in the era of generative artificial intelligence. *Frontiers in Education*, *9*.

Kamalov, F., Calonge, D. S., & Gurrib, I. (2023). New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution. *Sustainability*, *15*(16), 12451.

Kılınç, S. (2024). Comprehensive AI Assessment Framework: Enhancing Educational Evaluation with Ethical AI Integration. *Journal of Educational Technology and Online Learning*.

Mudawi, N. A., Beloff, N., & White, M. (2021). Developing a Framework of Critical Factors Affecting the Adoption of Cloud Computing in Government Systems (ACCE-GOV). In *Lecture notes in networks and systems* (p. 520). Springer International Publishing.

Mujalli, A., Wani, M. J. G., Almgrashi, A., Khormi, T., & Qahtani, M. (2024). *Investigating the factors affecting the adoption of cloud accounting in Saudi Arabia's small and medium-sized enterprises (SMEs)*.

Najana, M., Bhattacharya, S., Kewalramani, C., & Pandiya, D. K. (2024). *AI and Organizational Transformation: Navigating the Future*.

Nazri, S., Ashaari, M. A., & Bakri, H. (2022). EXPLORING THE ADOPTION OF ARTIFICIAL INTELLIGENCE IN INSTITUTIONS OF HIGHER LEARNING. *Journal of Information System and Technology Management*, *7*(27), 54.

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics Science and Technology Education*, *19*(8).

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *Journal of University Teaching and Learning Practice*, *21*(6).

Pizam, A., Öztürk, A., Balderas-Cejudo, A., Buhalis, D., Fuchs, G., Hara, T., Meira, J. V. de S., Revilla, M. R. G., Sethi, D., Shen, Y., State, O., Hacikara, A., & Chaulagain, S. (2022). Factors affecting hotel managers' intentions to adopt robotic technologies: A global study. *International Journal of Hospitality Management*, *102*, 103139.

Rasul, T., Nair, S., Kalendra, D. R., Balaji, M. S., Santini, F. de O., Ladeira, W. J., Rather, R. A., Yasin, N., Rodriguez, R. V., Kokkalis, P., Murad, Md. W., & Hossain, M. U. (2024). Enhancing academic integrity among students in GenAI Era:A holistic framework. *The International Journal of Management Education*, *22*(3), 101041.

Su, J., Mo, Y. L., & Sing, S. L. (2025). *Generative artificial intelligence in lattice structure design for additive manufacturing: A critical review* [Review of *Generative artificial intelligence in lattice structure design for additive manufacturing: A critical review*]. *1*(1), 25110006.

Tan, M. J. T., & Maravilla, N. M. A. T. (2024). Shaping integrity: why generative artificial intelligence does not have to undermine education. *Frontiers in Artificial Intelligence*, *7*.