_____

AUTOMATED ESSAY SCORING SYSTEMS: LEVERAGING ARTIFICIAL INTELLIGENCE TO ENHANCE ACADEMIC WRITING ASSESSMENT IN HIGHER EDUCATION

Musayeva Adiba Abdumajidovna

Doctor of Philosophy (PhD) in Pedagogical Sciences, Associate Professor of Uzbekistan National Pedagogical University After named Nizami

Usmonova Diyora Azamatovna

3rd year Student of Uzbekistan National Pedagogical University After named Nizami

Abstract: *The rapid evolution of Artificial Intelligence (AI) has introduced transformative possibilities for educational assessment, particularly through Automated Essay Scoring (AES) systems. This paper examines the integration of AI-driven tools in higher education to evaluate academic writing, addressing the challenges of grader subjectivity and time inefficiency in traditional methods. By utilizing Natural Language Processing (NLP) and machine learning algorithms, AES systems provide instantaneous, objective feedback, allowing for a more scalable approach to writing instruction. The study analyzes the correlation between human scoring and automated systems, highlighting the strengths of AI in linguistic accuracy while identifying current limitations in assessing deep rhetorical nuances. Ultimately, the research proposes a balanced pedagogical framework where AI serves as a formative support tool, enhancing rather than replacing human expertise in the academic writing process.*

Keywords: *Artificial Intelligence (AI), Automated Essay Scoring (AES), Natural Language Processing (NLP), Academic Writing, Educational Assessment, Higher Education, Formative Feedback.*

The landscape of modern higher education is undergoing a paradigm shift, driven by the rapid integration of sophisticated computational tools into pedagogical practices. Among these advancements, Automated Essay Scoring (AES) has emerged as a critical field of study, particularly within the context of academic writing assessment. Writing is a cornerstone of academic success, yet evaluating student discourse remains one of the most resource-intensive tasks for educators. Traditional manual grading, while offering qualitative depth, is frequently constrained by rater fatigue, temporal delays, and inherent subjectivity.

The emergence of Natural Language Processing (NLP) and machine learning has provided a robust framework for simulating human-like judgment in text evaluation. Unlike early

models that relied on superficial metrics such as word count or sentence length, contemporary AES systems utilize deep learning architectures to analyze structural coherence, lexical diversity, and even nuanced argumentative flow. [1] This study seeks to investigate the reliability of these systems in the English as a Second Language (ESL) environment, specifically focusing on how AI can be leveraged to provide formative, real-time feedback that fosters student autonomy.

This research employs a quantitative comparative analysis to evaluate the reliability of Automated Essay Scoring (AES) systems against human expert judgment. The methodology is structured to assess the linguistic precision and evaluative consistency of AI-driven platforms. [2]

Data Collection and Corpus Description

The primary dataset consists of 100 academic essays authored by third-year undergraduate students at the University.

- Topic of Essays: "The Role of Digital Literacy in Modern Education."
- Word Count: 300–500 words.
- Format: Argumentative academic discourse.

Evaluation Framework

To ensure a balanced comparison, the essays were graded by two distinct entities using a standardized 100-point rubric:

1. Human Raters (HR): Three senior lecturers from the English Philology department, using a double-blind grading process.

2. AI System (AES): A custom-integrated model utilizing the GPT-4o API and Natural Language Toolkit (NLTK) for syntactic analysis.

Statistical Metrics

The degree of agreement between AI-generated scores and human-assigned grades is measured using the Pearson Correlation Coefficient (r). This formula determines the linear relationship between the two scoring sets:

$$r = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}}$$

Where:

- n = Number of essays (100).
- x = Scores assigned by human raters.
- y = Scores assigned by the AES system.

Additionally, Mean Absolute Error (MAE) is used to calculate the average magnitude of the errors in the AI's predictions:

_____

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i|$$

Comparative Rubric Criteria

The evaluation was broken down into five core linguistic dimensions as shown in the table below:

Dimension	Description	Human Weight	AI Algorithm Metric
Grammatical Accuracy	Syntax, tense usage, and punctuation.	25%	Dependency Parsing & Error Detection
Lexical Diversity	Vocabulary range and academic tone.	20%	Type-Token Ratio (TTR)
Coherence & Cohesion	Logical flow and use of transitions.	20%	Latent Semantic Analysis (LSA)
Content Relevance	Alignment with the given prompt.	25%	Cosine Similarity Vectors
Mechanical Precision	Spelling and formatting.	10%	Pattern Matching

Discussion: The Limitations of AI in Nuanced Assessment

While the AES (Automated Essay Scoring) system demonstrates remarkable efficiency in quantifying linguistic parameters, its qualitative depth remains a subject of academic debate. The statistical high correlation in technical domains does not necessarily translate into a comprehensive understanding of human communication. [3]

The Challenge of Rhetorical and Creative Nuances

AI models primarily function on pattern recognition and probabilistic distributions of words. Consequently, they often struggle with:

- Sarcasm and Irony: AI frequently misinterprets sarcastic remarks as literal statements or factual errors, failing to recognize the underlying rhetorical intent.

- Creative Divergence: Highly creative or non-conventional arguments may be penalized by the algorithm because they deviate from the "standard" linguistic patterns found in the training data.

- Cultural Context: Automated systems lack the "world knowledge" required to appreciate culturally specific metaphors or idiomatic expressions prevalent in ESL (English as a Second Language) writing.

_____

"Gaming the System" (Prompt Engineering by Students)

A significant concern identified in this study is the potential for students to "game" the AES. [4] Since the AI rewards high lexical density and complex syntax, students may use sophisticated vocabulary inappropriately or artificially lengthen sentences to achieve a higher score, without necessarily improving the quality of the argument. This leads to a "surface-level" improvement that masks deficiencies in critical thinking.

Conclusion: Towards a Hybrid Pedagogical Model

The integration of Artificial Intelligence into writing assessment should not be viewed as a replacement for human intellect but as a strategic augmentation of pedagogical capabilities.

The Proposed Hybrid Model

To maximize the benefits of technology while preserving the integrity of human judgment, we propose a Hybrid Assessment Framework:

1.   AI as a Formative Tool: AI should handle the initial screening phase, providing real-time feedback on grammar, spelling, and structural cohesion. This allows students to refine their technical skills independently. [5]

2.   Human as a Summative Judge: The final evaluative stage must remain under human oversight. Educators should focus on the "soul" of the essay—logic, original thought, and emotional resonance—which AI cannot yet replicate.

Recommendations for Future Implementation

•   Institutional Training: Universities should provide training for both faculty and students on the ethical use of AI tools to prevent plagiarism and maintain academic honesty.

•   Context-Aware Algorithms: Future research should focus on developing "Explainable AI" (XAI) that provides students with clear reasons why a certain score was given, rather than just a numerical output. [6]

In conclusion, the goal of AES is not to dehumanize education, but to liberate the educator from mechanical tasks, allowing them to focus on the more profound aspects of teaching and mentorship. By embracing this synergy, institutions can significantly enhance the quality of linguistic education in the 21st century.

References

1.   Brown, G. W., & Chodorow, M. (2023). Natural Language Processing in Educational Assessment: Evolution of Automated Scoring. Routledge.

2.   Gardner, J., & Holmes, W. (2024). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. UNESCO Publishing.

_____

3. He, Q., & Von Davier, M. (2025). Advanced Methodologies in Automated Essay Scoring. Springer Nature.

4. Karypis, G., & Kumar, V. (2023). Multilevel algorithms for multi-constraint graph partitioning. Journal of Artificial Intelligence Research, 76, 112-145.

5. National Council of Teachers of English (NCTE). (2024). Position Statement on the Role of Artificial Intelligence in Writing Instruction. NCTE Reports.

6. Suleymanov, A. (2024). Digitalization of Higher Education in Uzbekistan: Current Trends and Challenges. Central Asian Journal of Education, 9(2), 45-58.