

269. AUTOMATED ESSAY SCORING SYSTEMS IN HIGHER EDUCATION

Pakutnik D.V.

*Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus*

Pristupa N.N. – PhD in Philology, Associate Professor

This paper examines Automated Essay Scoring systems in higher education, focusing on their capabilities, limitations, and pedagogical impact. It highlights the efficiency and accuracy of modern AI-based approaches while addressing concerns such as bias, lack of transparency, and challenges in evaluating creativity. Hybrid human-AI models are proposed as a balanced solution for effective implementation.

The expansion of higher education and increasing student enrollment have created significant challenges for maintaining assessment quality while ensuring timely feedback delivery. Automated Essay Scoring (AES) systems represent a promising technological solution that leverages natural language processing and machine learning to evaluate written assignments without direct human intervention. This paper examines the current state of AES technology, its pedagogical implications, and the challenges limiting widespread adoption in higher education institutions.

Modern AES systems employ sophisticated algorithms including feature-based approaches, neural networks, and transformer architectures to analyze textual characteristics such as grammar, vocabulary complexity, argument structure, and semantic coherence. According to recent research, these systems achieve scoring accuracy comparable to human raters while reducing evaluation time by up to 80 percent [1]. The integration of AES with Learning Management Systems enables immediate feedback delivery, supporting formative assessment practices that enhance student learning outcomes.

Despite technological advances, AES systems face significant limitations that constrain their applicability. Algorithmic bias remains a critical concern, as training data may reflect historical inequities in educational assessment. Studies demonstrate that machine learning models can perpetuate demographic biases in scoring outcomes [2]. Additionally, the assessment of creative writing, innovative arguments, and interdisciplinary responses poses challenges for systems trained on conventional essay structures. The black-box nature of deep learning models raises transparency concerns, as students and educators cannot fully understand evaluation criteria.

Hybrid assessment models combining automated preliminary evaluation with human oversight offer a promising direction for addressing these limitations. Research on human-in-the-loop systems demonstrates that such approaches maintain assessment efficiency while preserving pedagogical validity for complex evaluation tasks [3]. These models allocate automated scoring to straightforward responses while routing ambiguous or creative submissions to human evaluators, optimizing resource allocation.

Furthermore, the successful adoption of AES systems depends not only on technological performance but also on institutional readiness and stakeholder acceptance. Educators may express concerns about over-reliance on automated evaluation, while students may question the fairness and reliability of machine-generated scores. Therefore, effective implementation requires transparent communication, proper training for instructors, and the development of clear guidelines that define the role of AES within the broader assessment framework. Addressing these human and organizational factors is essential for ensuring trust and long-term sustainability of AES integration in higher education.

This paper demonstrates that Automated Essay Scoring technology provides significant advantages for large-scale assessment in higher education, particularly in the context of standardized writing tasks. However, institutional implementation requires careful consideration of fairness, transparency, and pedagogical alignment. Future research should prioritize developing explainable AI techniques, establishing bias detection frameworks, and designing effective human-machine collaboration protocols. Educational institutions should adopt phased implementation strategies, beginning with low-stakes assessments while gradually expanding to higher-stakes applications as confidence in system reliability grows. In addition, continuous monitoring and evaluation of system performance are necessary to ensure consistent assessment quality. Stakeholder feedback should also be systematically incorporated to support ongoing system improvement.

References:

1. *Automated Essay Scoring Using Natural Language [website]*. – URL: <https://ieeexplore.ieee.org/document/9310845> (date of access: 17.03.2026).
2. *A survey of automated essay scoring: Challenges, methods, and future directions [website]*. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S0925231225015887> (date of access: 18.03.2026).
3. *AI-assisted assessment of higher education quality: A visual analytical approach [website]*. – URL: <https://www.sciencedirect.com/science/article/pii/S2468502X26000021> (date of access: 18.03.2026).