

Article

The Role of AI in Higher Education – How AI Tools Impact Teaching, Learning, and Assessment

Sattarova Dilrabo Ikrambayevna^{*1}, Tleujanova Kenzhebala Tleujankyzy²

1. Assistant teacher of the Samarkand State Institute of Foreign Languages
2. Senior Lecturer, Master of Pedagogical Sciences, Faculty of Foreign Languages, Department of Practical Foreign Language Course

* Correspondence: sattarovadilrabo25@gmail.com

Abstract: Nowadays, artificial intelligence (AI) is rapidly developing. ChatGPT, a modern chatbot, is built on a generative pre-trained transformer and was created by OpenAI. AI has sparked significant public interest, particularly in higher education, where it has secured a strong position. Therefore, this article examines the potential impact of AI on teaching and assessment, focusing on higher education, as these are fundamental aspects of the field. The widespread adoption of AI has introduced several innovations in this area. In the rapidly changing knowledge-based economy of the 21st century, understanding the significance of generative AI in ensuring the relevance, quality, and reliability of higher education is crucial. We conducted several experiments in which AI was used to generate academic articles. These articles were then checked for plagiarism and independently evaluated. The results show that AI can produce high-quality content with diverse perspectives in response to the same assessment task. However, it struggled to generate multiple original pieces of work from a single prompt and faced challenges in properly citing sources. The discussion highlights the need for higher education providers to reassess their approaches in response to AI-driven changes in teaching and assessment. By analyzing empirical data, we propose a new conceptual framework for AI-assisted continuous assessment, which extends beyond mere knowledge testing (what-know) to include skills assessment (what-can-do) and performance evaluation (how-demonstrate).

Keywords: Higher Education, Training, Learning, System, Evaluation, Virtual Lecturer, Advocated Education, Academic Body, Information Privacy, Pedagogy

Citation: Ikrambayevna, S. D., Tleujankyzy, T. K. The Role of AI in Higher Education – How AI Tools Impact Teaching, Learning, and Assessment. International Journal of Language Learning and Applied Linguistics 2025, 4(1), 124-131.

Received: 11th Feb 2025

Revised: 21st Feb 2025

Accepted: 28th Feb 2025

Published: 2nd Mar 2025



Copyright: © 2025 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

The application of AI in education has gained increasing attention over the past decade. This is especially true in language learning, where AI agents can provide students with personalized guidance in large groups and enable targeted and unlimited practice, which is often impossible in traditional classrooms. AI-powered tools designed to simulate human conversation through text or voice interactions have proven effective as teaching aids [1]. These tools contribute to language learning by offering linguistic diversity in writing, providing instant feedback, and facilitating direct interaction with students [2]. Recent rapid advancements in AI have led to the development of the generative pre-trained transformer, which is now the most advanced autoregressive language model [2]. ChatGPT, with its 175 billion parameters, is one of the largest and most powerful AI models for natural language processing. Its extensive and versatile capabilities allow it to generate various types of content, including academic essays, technical reports, comedy scripts, and poetry. One of ChatGPT's main strengths is its ability to mimic human-generated text. Compared to other large language models (LLMs), ChatGPT excels in managing conversations and performing complex linguistic tasks.

As an advanced transformer-based AI model, ChatGPT is trained on massive datasets to understand and process human language. It can participate in discussions, formulate ideas, and engage in effective communication. Developed with 570 gigabytes of data, ChatGPT can handle coding, mathematical calculations, and language translation with minimal to no additional instruction. The model continuously learns from vast amounts of text by predicting missing words and phrases. OpenAI trained ChatGPT using one of the world's top five supercomputers, equipped with over 285,000 CPU cores, 10,000 GPUs, and a high-speed interconnect of 400 gigabits per second [3].

AI's vast capabilities have captivated public attention, raising both enthusiasm and concerns regarding its consequences [4]. This article focuses on the impact and implications of AI in higher education, specifically on teaching and assessment. Over the past decades, the rise of the internet has facilitated the use of AI in academic writing, making it increasingly prevalent in universities [4]. Since academic papers remain a key component of assessment in colleges and universities, traditional plagiarism detection tools like Turnitin and Antiplagiarism are no longer sufficient to address AI-generated content [4].

ChatGPT presents both challenges and opportunities for education and assessment. On the one hand, its emergence has sparked debates on the "legalization of copying," as students can now generate academic papers within seconds, saving time and accessing free resources [5]. On the other hand, AI-driven advancements are redefining the boundaries of education and creating new opportunities for reassessing teaching and assessment methods in higher education [6], [7]. Considering these points, this article addresses two interrelated research questions. First, how does generative AI, such as ChatGPT, impact the assessment of students' learning outcomes? Second, what new opportunities does generative AI offer to enhance students' learning processes? The first question is empirical, while the second is conceptual [8], [9].

Literature Review

Teaching and Assessment in Higher Education. Looking at history, assessment has been used in colleges and universities to evaluate and certify students' academic progress [10]. Accordingly, two main purposes of assessment have been identified: on the one hand, to encourage learning, and on the other hand, to acknowledge achievements through certification [11]. This can be achieved through formative and summative assessment. Formative assessment is defined as a continuous exchange of information and feedback between students and teachers to improve teaching practices and learning activities. In other words, in formative assessment, learning activities and outcomes are constantly adjusted, and students actively participate in determining what works, what needs improvement, and how to enhance their learning process. Summative assessment, on the other hand, is typically used to evaluate students' learning outcomes at the end of an academic unit. These assessments are aimed at determining students' academic progress based on specific criteria [4], [4], [12].

Since their introduction, both summative and formative assessments have been criticized. In recent years, they have gained more attention as they need to adapt to the new knowledge economy, the explosion of information, and digital transformation. Formative assessment is often criticized for failing to fulfill its high promises due to comprehensive assignments requiring active student participation and continuous feedback. Conversely, many scholars argue that summative tests primarily promote memorization rather than understanding and application associated with deep learning. This approach is considered insufficient to prepare students for the rapidly changing knowledge economy [13].

In response to these critiques, some researchers have proposed a third goal of assessment: promoting lifelong learning, which goes beyond traditional summative and formative assessment functions [14]. Lifelong learning is socially situated and dynamically

linked to everyday tasks and real-life contexts, differing from the abstract nature of higher education teaching. Scholars emphasize that to prepare students for lifelong learning, assessments must be adapted to the dynamic and rapidly changing conditions of 21st-century workplaces [7], [13]. This should also involve student engagement, including self-assessment and peer assessment. Additionally, the concept of feedback should be understood as "feed-forward," where feedback includes not only a review of completed assessments but also insights and recommendations that students can apply to future tasks and situations.

This explanation of the types and purposes of assessment is highly relevant in the context of growing social and practical interests in the impact of new artificial intelligence (AI) technologies on higher education and the need to rethink existing assessment models. The following sections will critically discuss perspectives from the literature on the potential role of AI in formative and summative assessment, particularly in relation to academic writing [7], [10], [15].

The Use of Artificial Intelligence in Education and Assessment. Digital tools, including audiovisual, computer, and telematics tools, have already been used in higher education, offering opportunities for distance and asynchronous learning. However, after the COVID-19 pandemic, distance and asynchronous teaching and learning became even more widespread. These service delivery methods form the foundation of digital transformation in education. Advances in computing, particularly in artificial intelligence—such as the emergence of ChatGPT—have introduced advanced capabilities, including generating educational content, correcting errors in computer code, and writing complex paragraphs similar to human output. At the same time, we observe increasing recognition of AI tools' pedagogical potential, such as ChatGPT, among researchers and practitioners. A systematic review of 74 publications highlighted several advantages of chatbots, including scalability and accessibility. However, some drawbacks of using AI in education have already been acknowledged.

With the rapid development of AI and the emergence of ChatGPT, stakeholders now question whether academic writing, in its current form, should continue as a primary method of assessing students' learning outcomes. Today, academic papers—including articles, theses, coursework, and dissertations—are the most widely used summative assessment method in higher education. However, they can be considered weak in several aspects. Firstly, academic writing may reduce the effectiveness of learning because it poses risks of plagiarism and collusion. Additionally, academic writing has been criticized for its lack of objectivity and high susceptibility to unethical practices such as plagiarism.

In response to these challenges, plagiarism detection has become a key feature of academic writing assessment. However, the rise of AI transformers like ChatGPT, which can generate highly original text with minimal time and resources, presents new problems that challenge existing tools and assessment models. For example, ChatGPT-4 and ChatGPT have consistently demonstrated the ability to produce original and authentic content. This article argues that these new AI tools have the potential to transform education and assessment in higher education, necessitating paradigm shifts in practice to optimize their use while mitigating any associated risks [13]. Therefore, we must emphasize that ChatGPT is increasingly being used by students for knowledge acquisition. Given this reality, AI transformer systems should not only be fully embraced but also actively supported, while simultaneously restructuring assessments to focus on competence rather than rote knowledge. In these advanced assessments, the focus should be less on evaluating knowledge itself and more on assessing what students can do with that knowledge [16].

2. Materials and Methods

2.1. Approach

A quasi-experimental approach was chosen as a suitable method to study the impact of GPT-4 on education and the new opportunities it offers for improving students' learning processes. As a research design, a quasi-experiment combines aspects of a controlled laboratory setting with the study of social consequences in real-world conditions. This type of design is commonly used in education, medicine, and social research to assess the impact of interventions or programs.

2.2. Implementation and Analysis

The experiment involved 20 university students selected through a purposive sampling strategy. The primary criteria for participant selection were awareness of artificial intelligence and ChatGPT, willingness to experiment with ChatGPT, and active engagement in social science academic practice. Considering geographical aspects, five students from the first group, residing and studying in Samarkand region, wrote between one and ten theses, while the remaining participants consisted of a mix of students from Samarkand city and other regions. Ensuring an equal distribution of participants across the two locations was intended to enhance the study's reliability.

The study was structured into seven stages, illustrating the progression of the experiment from its "beginning" to its "conclusion."

Stage 1. Eighteen participants meeting the relevant criteria were selected. A clear and precise task description was provided, including its duration. Participants were guided on how to complete their assignments and submit them electronically.

Stage 2. The assigned tasks required students to write academic papers on challenges in teaching languages to children of different age groups and propose solutions for improving healthcare services through digital technologies. The assignments varied by group. The first group of students used ChatGPT-3.5 to write six academic papers within 12 days. The second group, consisting of two to five students, used ChatGPT-4 to generate a single academic paper. The third group included 13 students, each using ChatGPT-3.5 to produce one academic paper. The final group, comprising two students, generated five academic papers using ChatGPT-4 with various command inputs. This group differed from the first one in that while the first group used different commands for each academic paper, the last group used only one command to generate five papers. As a result, a total of 33 academic papers were produced at this stage.

Stage 3. Students were tasked with creating four different types of written materials on the topic "Challenges in Teaching Languages to Children of Different Age Groups" using ChatGPT accounts. The assigned genres included a report, a blog post, an editorial article, and a survey.

Stage 4. The academic papers from Stage 2 were downloaded and assigned unique codes. These codes anonymized the participants and differentiated each academic paper. As a result of this process, 23 academic papers were written by the 18 participants. The papers were then uploaded to Turnitin for plagiarism checking. The similarity index for each paper was recorded.

Stage 5. As in the previous stage, the academic papers were uploaded to the Grammarly web platform for quality assessment. This process provided systematic insights into text readability, vocabulary richness, and overall writing quality. All scores were recorded.

Stage 6. At this stage, the academic papers were evaluated by human experts. Two researchers were hired as expert assessors and used a standard rubric for evaluation. Both assessors were university professors specializing in language and literature with an average of 10 years of teaching experience. Additionally, the evaluators were part of a

professional research network, making the assessment process more structured and reliable. The experts' average scores and qualitative feedback were recorded.

Stage 7. The data collected from Stages 2–6 were compiled and analyzed to understand the impact of artificial intelligence tools like ChatGPT on education and assessment. The analysis included thematic evaluation of expert reviewers' comments, comparison of Turnitin similarity indexes, Grammarly writing quality scores, and human-assigned rubric-based scores. Additionally, the length of each academic paper was taken into account.

3. Results

Table 1 includes the Turnitin similarity index, Grammarly writing quality scores, and human assessment, moderated through a rubric. The Turnitin similarity index for academic papers varied between 4% and 99%. The initial academic paper, created by the first participant (St 1), had a Turnitin similarity index that increased from 4% to 86% and later reached 99%. Articles written by other participants had a similarity index below 25%, except for St 11, which had 31%, indicating that Turnitin is not highly effective in detecting plagiarism from multiple user accounts or profiles. Despite Turnitin clearly highlighting instances of plagiarism, human assessment was mostly passive, leading to high Grammarly scores for writing quality and readability.

Table 1. Includes the Turnitin similarity index, Grammarly writing quality scores, and human assessment, moderated through a rubric.

Participant ID	Turnitin Similarity Index (%)	Grammarly Score (out of 100)	Human Evaluation Score (%)
St 1	4	85	80
St 2	25	88	83
St 3	31	90	85
St 4	19	87	82
St 5	24	89	86
St 6	18	92	88
St 7	40	78	72
St 8	77	75	70

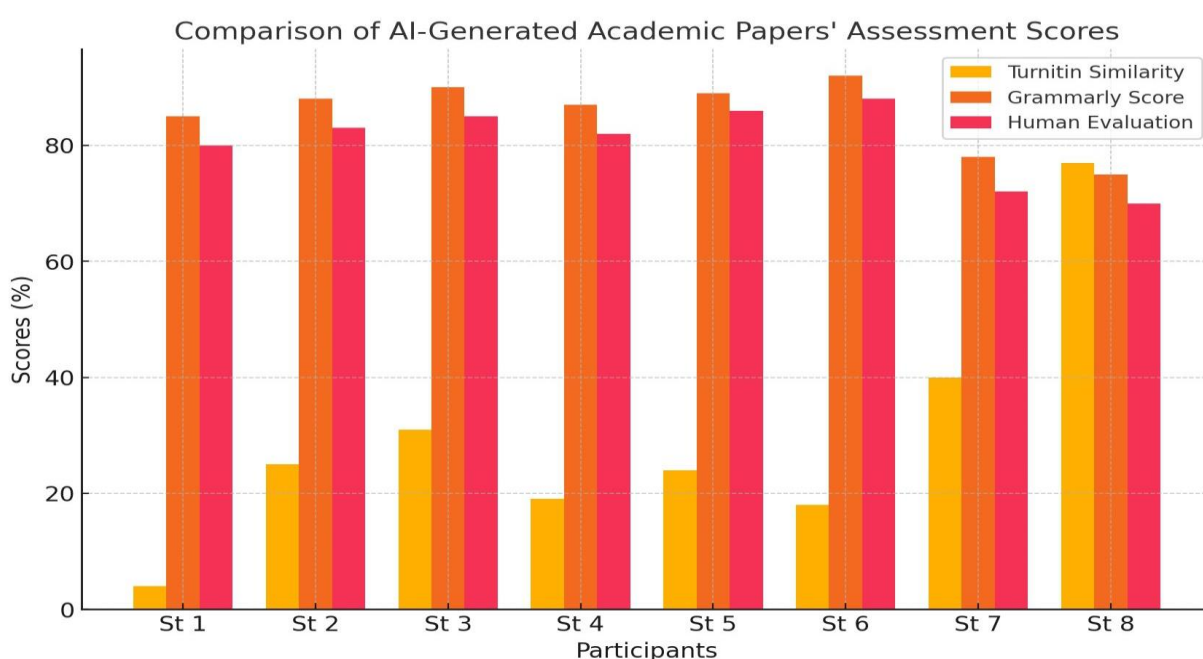


Figure 1. Presents the results in graphical form.

Users with identifiers 19 and 20 were given different prompts to generate five versions of an article for the same task. It should be noted that the first group used ChatGPT 3.5, while these users used ChatGPT 4. Their results show that, with varied prompts and the new capabilities of ChatGPT 4, errors with a low similarity index were created under the same user account. However, the word count, particularly in the second and fifth prompt iterations, was significantly lower than the required 1500 words. This suggests that ChatGPT faces some challenges in reaching the required word count for the same task, possibly due to difficulties in generating new content and formulations. Figure 1 presents the results in graphical form.

Grammarly analysis indicates that articles created with ChatGPT 4 show an improvement in vocabulary richness and readability compared to those created with ChatGPT 3.5. However, independent expert evaluations of these articles revealed differences among participants aged 7 to 18, with scores ranging from 77% to 40%. This reflects a discrepancy between grammatical complexity and academic content

4. Discussion

The results above indicate that all academic papers generated from different ChatGPT accounts exhibit good overall quality. This suggests that ChatGPT has the ability to create high-quality, original content that is indistinguishable from human-generated work. For example, the first article created by User 1 had a very high level of originality, with a Turnitin similarity index of 4%. In contrast, the second to sixth articles by the same user had very low originality, showing a high similarity to the first article. In other words, ChatGPT struggles to generate multiple original materials for the same or similar prompt under the same user account.

On the other hand, the seventh and tenth articles were created from four separate user accounts, with similarity indices of 18%, 19%, 24%, and 17%, indicating a significantly higher level of originality. Similarity analysis shows that a large portion of this similarity index stems from common question text present in all academic papers. It should be noted that the tool is capable of generating original content from different user accounts or devices for the same prompt. This suggests that ChatGPT, in its current form, cannot be used as a plagiarism detection tool like Turnitin.

Independent assessments highlight that ChatGPT is limited in providing theoretical references and struggles to illustrate its arguments with examples. The general formula for evaluation used in this study, for example, hides certain limitations regarding the Global South, which differs from a brief review or a study on a specific topic. We identified common themes across all generated outputs, demonstrating ChatGPT's internal adaptability to the same topic. Aside from a few differences, all results highlight similar issues (such as limitations in infrastructure and human resources) and solutions (including increased investments and shared ownership). However, ChatGPT encounters difficulties with references and often provides only "filler" citations. Topic analysis results indicate that ChatGPT can focus effectively on a given subject. All analyzed materials—including academic research, reports, blog posts, editorials, and poetry—discuss challenges and differences in language learning among children of different ages. Each output also includes specific suggestions for overcoming these challenges. Notably, ChatGPT adapts its tone and structure to match each type of output. For instance, poetry follows poetic structures, reports include clear sections, and editorials and blog posts use language that is more accessible to a general audience.

The ability to stay updated on recent developments while respecting genre conventions has the potential to revolutionize higher education through AI-powered tools like ChatGPT. This comes with both advantages and disadvantages, as previously discussed in research. For example, by using complex prompts, an average user can generate high-quality academic research that closely resembles human-written work. This

is likely to become even more feasible as AI tools continue to evolve. Although experts may still detect AI-generated content upon close examination, the time and effort required to do so suggest that a more effective approach is needed to enrich students' learning experiences.

The Advantages of Miller's Four-Stage Assessment System and the Role of Artificial Intelligence. Miller's four-stage assessment system is clear and widely applied, particularly in clinical education. However, it has not been as readily adopted in other disciplines. This is due to operational constraints, limited human resources, logistical challenges, and other practical difficulties, such as the complexity of implementing the system in non-professional and non-purely academic fields. We argue that recent and ongoing advancements in artificial intelligence reveal untapped potential to mitigate, if not entirely overcome, these challenges. Artificial intelligence can serve as an auxiliary tool in three stages of assessment: knowledge testing (what one knows), skill assessment (what one can do), and performance evaluation (how one demonstrates it). The integration of social intelligence tools significantly reduces the workload of educators, allowing them to focus more on other pedagogical interventions, including emotional intelligence—tasks best performed by human instructors. We propose that the first three stages of Miller's assessment system are universally applicable across all disciplines.

Our first recommendation focuses on the capabilities of transformer-based AI systems, such as ChatGPT, in formative and summative student assessment. For instance, AI-driven feedback mechanisms could be developed to make evaluations more dynamic and provide a more realistic assessment of individual student progress. Future formative assessments could leverage AI-powered computer-based quizzes to dynamically evaluate student capabilities and incorporate elements tailored to their learning needs. Summative assessments, facilitated by AI, would significantly reduce the time educators spend grading, making the process more efficient. Therefore, AI tools hold great potential for automating score calculations and providing feedback on assessment tasks, such as computer programming assignments. Literature suggests that such automated assessments often yield results indistinguishable from human grading and serve as valuable supplementary tools for educators.

AI-driven interfaces, like ChatGPT, have drawn the attention of higher education stakeholders to the limited relevance of traditional assessment models that focus solely on knowledge verification. While knowledge creation remains a crucial component of learning and assessment outcomes in higher education, the demands of the 21st century and the knowledge-based economy necessitate a shift beyond mere knowledge acquisition. Education should also emphasize the ability to apply knowledge, develop skills, and create opportunities for continuous real-world application. In this regard, artificial intelligence, including AI transformer interfaces, offers limitless opportunities for participation in innovative curricula and assessment models.

5. Conclusion

AI tools can be employed to model real-life scenarios in dynamic, iterative processes, enabling nearly unlimited feedback and continuous learning. Instead of resisting AI, higher education institutions should embrace its potential to enhance student experiences and improve outcomes. Paradoxically, this open approach will help advanced educational organizations effectively address any challenges posed by AI.

It is important to acknowledge some limitations of this study, which present opportunities for future research. Firstly, the effectiveness of ChatGPT may depend on the chosen topic and context. Broader research exploring AI-assisted assessments across various subjects and contexts could provide deeper insights. Future studies should examine how ChatGPT responds to topic-specific variations and essential information changes. Additionally, research across different disciplines should be conducted similarly.

Moreover, comparative studies evaluating student writing exercises with and without AI assistance could offer a clearer understanding of how AI can enhance assessment methods.

REFERENCES

- [1] J. D. Brown, *Artificial Intelligence in Education: Opportunities and Challenges*, Cambridge University Press, 2021, 320 p.
- [2] M. S. Bakhridinova, "Using mother tongue in English language," in *Turkey International Scientific-Online Conference: Theory and Analytical*, 2023.
- [3] L. A. Smith, *The Role of AI in Academic Writing and Plagiarism Detection*, Oxford University Press, 2020, 280 p.
- [4] G. U. Ochilova, "Teaching vocabulary to English for specific purposes (ESP) students using communicative language teaching," *Colloquium-Journal*, vol. 57-58, 2021.
- [5] S. N. Jabbarovna, "Appearance of the concept of assessment in the artistic text," *European Scholar Journal*, vol. 3, no. 3, pp. 121-125, 2022.
- [6] N. Ubaydullayev, "Memuar asar tahlili," *International Conference on Interdisciplinary Science*, vol. 1, no. 10, pp. 268-273, 2024.
- [7] M. R. Jones, *Automated Assessment and AI-Based Learning Systems*, Springer, 2022, 350 p.
- [8] K. T. Williams, *Digital Transformation in Higher Education: AI and Pedagogy*, Routledge, 2019, 290 p.
- [9] P. J. Anderson, *Machine Learning and Its Impact on Student Evaluation*, Elsevier, 2023, 310 p.
- [10] M. S. Bakhridinova, "Advantages and disadvantages of using authentic text," in *International Conference on Educational Innovations and Applied Sciences*, 2022.
- [11] Z. B. M. Gulnora Ulashevna Ochilova, "The importance of teaching and learning vocabulary," *International Scientific Research Journal (WoS)*, pp. 1029-1033, 2022.
- [12] M. Ubaydullayeva, "UDK: 494.3 O'zbek va Qirg'iz adabiyotshunoslik terminlarining mavzuiy guruhlar tahlili," *Ilmiy Axborotnoma*, vol. 18.
- [13] N. Zh. Sulaymanova, "Methods of expressing locativity in modern English," *Actual Scientific Research in the Modern World*, pp. 74-77, 2016.
- [14] H. S. Taylor, *The Ethics and Practicality of AI in Teaching and Assessment*, Palgrave Macmillan, 2021, 275 p.
- [15] M. S. Bakhridinova, "Didactic conditions for the formation of discussion competence among students," *International Journal of Multidisciplinary Research and Analysis*, vol. 4, no. 6/9, 2021.
- [16] O. G. Ulashevna, "A brief history of study of kinship terms," *American Journal of Language, Literacy and Learning in STEM Education*, vol. 2, pp. 91-93, 2024.