BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

REVIEW

Exploring ChatGPT's Role in Language Learning Assessment: A Two-Year Bibliometric Analysis

Gulnoza Sabirova [1*] **, *Lola Shasaidova*** [2] **, *Madina Nazarova*** [3] **, *Nodira Isamukhamedova*** [4]

[1] *Teaching Theory and Methodology Department, Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, National Research University, Tashkent 101000, Uzbekistan*

[2] *Chinese Philology Department, The Uzbek State University of World Languages, Tashkent 100170, Uzbekistan*

[3] *English Language Department, University of World Economy and Diplomacy, Tashkent 100007, Uzbekistan*

[4] *Global Education Department, Westminster International University in Tashkent, Tashkent 100047, Uzbekistan*

ABSTRACT

ChatGPT is changing the way we test language more and more by using advanced deep learning models that mimic how people talk to each other. These models, especially large language models (LLMs), can understand, create, and evaluate natural language. This makes them very useful for educational evaluation. This report looks at how ChatGPT is changing the way we evaluate things by looking at academic articles from 2023 to 2024. We used the Web of Science database to carefully choose 350 peer-reviewed articles that looked at worldwide research trends. Bibliometric technologies like VOSviewer and Biblioshiny were used to find important changes in the field. These technologies made it possible to see how publication patterns, co-authorship networks, and the development of thematic trends relating to the function of ChatGPT in education have changed over time. The results showed a number of important things. First, the United States, China, and the United Kingdom were the top contributors to this sector. This shows that institutions and governments in these countries are very interested in using AI language models in education. Second, Elyoseph Z. and Levkovich I. were two of the most numerous writers who contributed to the discussion around AI-driven assessment. Some of the most well-known journals that published this research are JMIR Medical Education, Journal of Medical Internet Research, JMIR

*CORRESPONDING AUTHOR:

Gulnoza Sabirova, Teaching Theory and Methodology Department, Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, National Research University, Tashkent 101000, Uzbekistan; Email: gulnoza1978@mail.ru

ARTICLE INFO

Received: 4 June 2025 | Revised: 19 June 2025 | Accepted: 25 June 2025 | Published Online: 25 August 2025
DOI: https://doi.org/10.30564/fls.v7i8.10340

CITATION

Sabirova, G., Shasaidova, L., Nazarova, M., et al., 2025. Exploring ChatGPT's Role in Language Learning Assessment: A Two-Year Bibliometric Analysis. Forum for Linguistic Studies. 7(8): 1151–1165. DOI: https://doi.org/10.30564/fls.v7i8.10340

Formative Research, and Education Sciences. The keyword co-occurrence analysis showed that words like "Chatgpt," "assessment," and "artificial intelligence" are used a lot, which strengthens the connection between computational linguistics and educational evaluation.

Keywords: ChatGPT; Bibliometric Analysis; Review; Artificial Intelligence in Education; Language Assessment; Formative and Summative Assessment; Multilingual Learning

1. Introduction

Since its launch in late 2022, ChatGPT has rapidly become a transformative tool in educational contexts, particularly in the domain of language acquisition. Its integration into classroom and remote learning environments has altered traditional teaching and learning methodologies, offering innovative ways to deliver instruction, engage learners, and assess linguistic performance. Among the various technological developments that have emerged over the past decade, ChatGPT's impact on the teaching and evaluation of speech and writing skills stands out as particularly significant, introducing new dynamics to how language is taught, practiced, and assessed [1].

One of the key areas where ChatGPT has proven influential is in language assessment, which includes both formative and summative evaluation of learners' productive skills—namely speaking and writing. Educators can now use ChatGPT to create and customize a wide array of text-based assignments that align with specific learning objectives and language proficiency levels. By leveraging its natural language processing capabilities, ChatGPT can generate authentic prompts for speaking or writing practice, simulate dialogues, and provide contextualized vocabulary and grammar support. This enables language teachers to design rich linguistic tasks that go beyond rote memorization or decontextualized drills, encouraging more meaningful and communicative use of language [2,3].

Moreover, ChatGPT facilitates a shift toward more personalized and student-centered approaches in assessment. With access to instant AI-generated feedback, students are able to engage in self-directed learning and iterative improvement. For example, learners can input a written response or spoken transcript and receive near-immediate commentary on grammatical accuracy, lexical range, coherence, and task achievement. This real-time feedback loop is especially useful in language learning, where continuous input and correction are crucial for skill development. Such features not only enhance student motivation but also make language learning more accessible outside of the classroom, allowing for flexible, autonomous practice [4].

From the perspective of instructional efficiency, ChatGPT offers time-saving benefits for teachers, particularly in grading and assessment preparation. Instructors can use the tool to evaluate student responses according to established rubrics or linguistic criteria, such as vocabulary usage, syntactic complexity, or genre-specific conventions. While these automated assessments may not yet match the nuance of human judgment in all cases, especially in evaluating creativity or pragmatic appropriateness, they provide a useful baseline that can be supplemented by teacher feedback [3]. In blended assessment models, AI serves as a preliminary evaluator, while the teacher applies deeper pedagogical insight to refine and contextualize feedback [5].

However, the widespread adoption of ChatGPT in language education also presents several challenges, particularly in ensuring the authenticity and integrity of student work. One of the most pressing concerns is the potential for academic dishonesty, as students may be tempted to submit AI-generated responses as their own. This issue is particularly acute in writing tasks, where polished, grammatically correct essays can easily be produced with minimal human input [5]. The limitations of current AI detection tools, which struggle to reliably distinguish between human- and machine-generated text, make it difficult for educators to verify the originality of student submissions [6].

The challenge of maintaining assessment validity is further compounded by equity concerns. Not all students have equal access to AI tools like ChatGPT, either due to technological infrastructure, financial constraints, or institutional policies. This digital divide can lead to discrepancies in performance and unfair advantages for students who can routinely use AI to draft, revise, or rehearse their language production. In such scenarios, AI access becomes a hidden

variable that affects test outcomes, potentially undermining the fairness and reliability of language assessments [2].

To address these concerns, researchers such as Cotton et al. have proposed several pedagogical and administrative strategies. Their study emphasizes the importance of designing assessments that are less susceptible to automation and more reflective of authentic cognitive and linguistic engagement [7]. For instance, tasks that require learners to draw on personal experiences, respond to localized or classroom-based content, or reflect critically on AI-generated input are less likely to be effectively completed by ChatGPT alone. Such reflective assignments promote higher-order thinking and require students to demonstrate both linguistic competence and metacognitive awareness [8].

Additionally, Cotton et al. advocate for the development and application of transparent rubrics that explicitly outline assessment criteria, including elements such as voice, originality, coherence, and audience awareness. When students understand the expectations for each linguistic component, they are more likely to approach tasks thoughtfully rather than relying on AI shortcuts [7]. Blending automated assessment with teacher-led review is also encouraged, allowing instructors to spot inconsistencies, gauge student progress over time, and provide qualitative feedback that AI systems may miss.

The potential of ChatGPT for formative assessment is particularly notable in the realm of speaking skills. Oral language production—often neglected in technology-enhanced language learning—can be scaffolded through ChatGPT's simulated dialogues and interactive prompts. For instance, students can engage in role-plays, interviews, or debates with the AI, receiving immediate feedback on fluency, pronunciation (with integrated tools), and grammatical control. While the AI may lack the full pragmatics and non-verbal cues of human interlocutors, it nonetheless offers a scalable solution for practicing real-time interaction, especially in contexts where access to native speakers or language labs is limited [9].

Furthermore, ChatGPT's multilingual capabilities enable cross-linguistic comparisons and code-switching exercises that deepen learners' understanding of grammatical structures, false friends, or idiomatic expressions. Teachers can exploit this feature to raise students' language awareness and support the development of metalinguistic knowledge [10]. This is particularly beneficial for advanced learners preparing for standardized language proficiency exams (e.g., TOEFL, IELTS, CEFR-based tests), where nuanced command of vocabulary, syntax, and discourse is essential Bibliometric research into AI-driven language education, as highlighted by Cotton et al. [7], also reveals growing academic interest in the ethical, technical, and pedagogical implications of tools like ChatGPT. Emerging trends include studies on AI-assisted peer feedback, generative AI for vocabulary expansion, and the integration of ChatGPT into CLIL (Content and Language Integrated Learning) environments. However, gaps remain, particularly regarding the long-term impact of AI tools on learner autonomy, linguistic creativity, and language identity formation. These dimensions call for longitudinal research and interdisciplinary dialogue among educators, linguists, technologists, and policymakers. ChatGPT's integration into language education offers a paradigm shift in how language is taught, practiced, and assessed. It brings numerous benefits: enhancing feedback, increasing access to individualized practice, and supporting efficient assessment design. However, these advantages must be carefully balanced with critical concerns around academic integrity, digital equity, and the evolving role of the language teacher. As the field moves forward, it is essential to establish ethical guidelines, refine assessment methodologies, and foster digital literacy so that AI can be harnessed to genuinely support, rather than supplant, human language learning and instruction.

There have been many recent studies on how ChatGPT can be used in schools. However, most of the existing research is limited in scope, focusing on specific situations, small groups of people, or single-country settings. To get a more global and complete view, it's important to integrate results from a wide range of schools and levels of learning. Recent work by European and worldwide scholars has shed light on how ChatGPT and AI tools are being used in different cultural and educational settings, which helps to reduce this gap in representation.

For example, Lavidas et. al. [11] looked into how likely humanities and social sciences students were to employ AI technologies in their schoolwork. Their study showed that students are typically willing to adopt AI for learning activities like organising and creating content. However, they still need sufficient ethical advice, digital literacy, and institutional support. This study helps us understand how students in higher education, especially in Europe, think and fits with

the trend of more people accepting AI around the world. In contrast, Aravantinos et al.[12] looked at how AI may be used in primary school education by doing a systematic review of research that was indexed in Scopus. Their results show that ChatGPT is becoming more important for personalised learning and classroom engagement. However, they also point out problems such as teachers not getting enough training, infrastructural shortages, and restricted policy frameworks. This review adds to the body of research on higher education by looking at early education, where AI is still being tested but has a lot of potential. Uğraş et al.[13] conducted two studies on ChatGPT's role in early childhood and primary school settings, which further supports this cross-level exploration. One study focused on how the technology might help with sustainable teaching by making repetitive chores easier, helping with lesson planning, and giving feedback in real time in the classroom. The other study, which was based on interviews with teachers, showed that ChatGPT can help students be more creative and think critically in early STEM education. It also said that it should be used with care and guidance to make sure it is appropriate for the age group.

These studies, when taken together, give us a better and more complete picture of ChatGPT's function in education, from the point of view of both students and teachers, and from early childhood to college. Importantly, they fill a vacuum in prior literature reviews and bibliometric studies, which frequently don't cover a wide range of geographic and educational areas. By adding these results to the current study, we can have a more well-rounded and topically rich discussion of ChatGPT's effects on education. This integration increases the basis of our bibliometric analysis by bringing it in line with current empirical and review-based research and by bringing attention to areas that haven't been studied enough, such as sustainability, ethics, early education, and institutional preparation. In the end, this bigger picture makes it easier to see how this study serves to map ChatGPT's function as both a new technology and a catalyst for change in many different educational institutions throughout the world.

This study uses bibliometric analysis to look at current trends, find the most important researchers, and point out gaps in the literature about using ChatGPT for educational assessment. Researchers can use bibliometric analysis to look at big datasets and get a full picture of the publication landscapes in different fields, such as language teaching. This study's goal is to find the most important countries, journals, keywords, and publications in the context of ChatGPT-enhanced assessment. By looking at these things, this study gives useful information that can be used in future work on using AI-powered tools to learn and test language skills.

According to these objectives, we formulated the following research questions:

1. What are the current trends in utilising ChatGPT in educational assessment?
2. What are the leading countries, journals, and authors contributing to the research field?
3. What are this field's key topics, concepts, and keywords?

2. Materials and Methods

The review aimed to analyse the scientific articles on using ChatGPT in Assessment. The articles for bibliometric analysis were obtained from the WOS database for the period of 2023 and 2024. The reason for analyzing articles in the timespan 2023 and 2024 is that ChatGPT emerged in November 2022, and most articles began to be published from 2023. 150 articles were selected due to the inclusion criteria and reviewed in the study. The keywords for searching the relevant articles were "ChatGPT" and "Assessment" language learning," "assessment," "AI in education," and "academic writing". The data was obtained in December 2024. Data visualisation was employed using VOSviewer and Biblioshiny software. The quantitative research method was used in analysing the articles.

2.1. Inclusion Criteria

The articles were selected based on the PRISMA protocol[14] with the keywords "ChatGPT" and "Assessment" language learning," "assessment," "AI in education," and "academic writing". Articles published between 2023–2024 were selected. The English language was another criterion for selecting articles for review. **Figure 1** illustrates the flowchart of selected articles for bibliometric analysis.

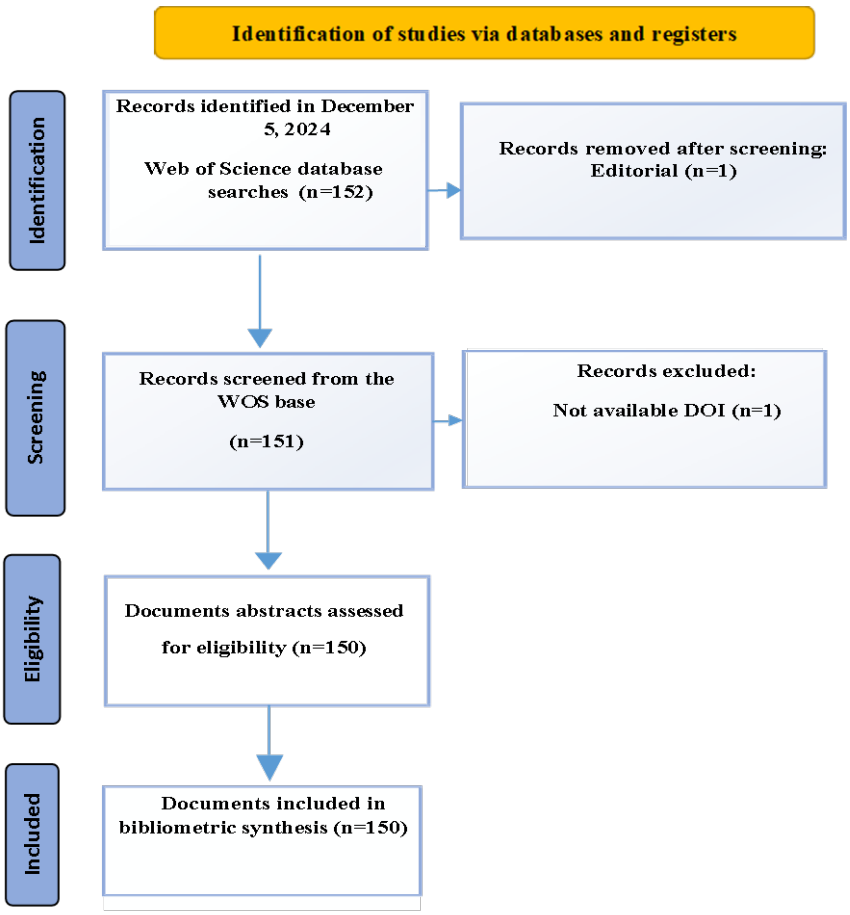

Figure 1. Methodology flowchart for the research[14].

2.2. Bibliometric Analysis

The data for bibliometric analysis was obtained in CVS, BibTex, Excel format and uploaded to Excel, Biblioshiny, and VOSviewer software. The articles were thoroughly checked for relevance and analysed based on a number of categories such as top authors, top publishing countries, keywords, co-authors, top journals, and co-occurrence of keywords. It is the goal of bibliometric reviews to provide a comprehensive overview of the literature rather than to assess the quality of research or to define terminology[15].

3. Results

3.1. Trends in Publication on ChatGPT in the Assessment

In the period from 2023 to 2024, there were 150 English publications indexed in the WOS online database. These documents encompass 116 articles, 16 proceeding papers, 8 review articles, 6 letters, and 4 editorial materials (**Figure 2**).

3.2. Top Authors

Figure 3 illustrates the authors' productivity on the research topic ChatGPT in assessment over the last two years. 533 authors from 47 countries published 150 articles. **Figure 3** demonstrates the top 10 authors who have published more than 1 article on the related theme. Elyoseph Z.[16–19] emerges as the top author with 4 articles, followed by Levkovich I.[17–19] with 3 articles published. Furthermore, Berger C.[20], Cabrero-Daniel B.[20], Nouri A.[21], Sivencrona H.[20], Surovková J.[22,23], Thurza A.[22,23], Urban R.[23,24], and Strunga M.[22,23] are among the authors with 2 publications each.

Figure 4 shows data on the distribution of articles based on the corresponding authors' countries. It represents the

total number of articles with a single corresponding author (SCP) and the number of articles with multiple corresponding authors (MCP). The data reveals that the country with the most publications on the topic is the United States, with 24 articles, followed by China, which has 17 articles. Most articles have corresponding authors from the United States. The data indicates that the USA has the highest proportion of single and multiple corresponding authors.

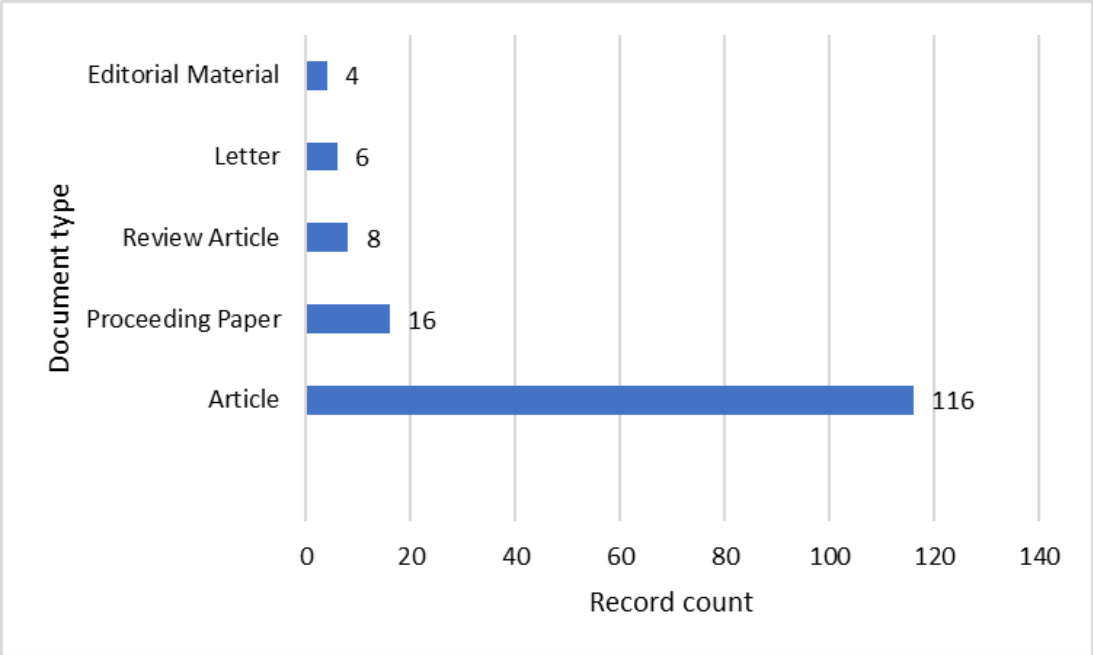

Figure 2. Document types.

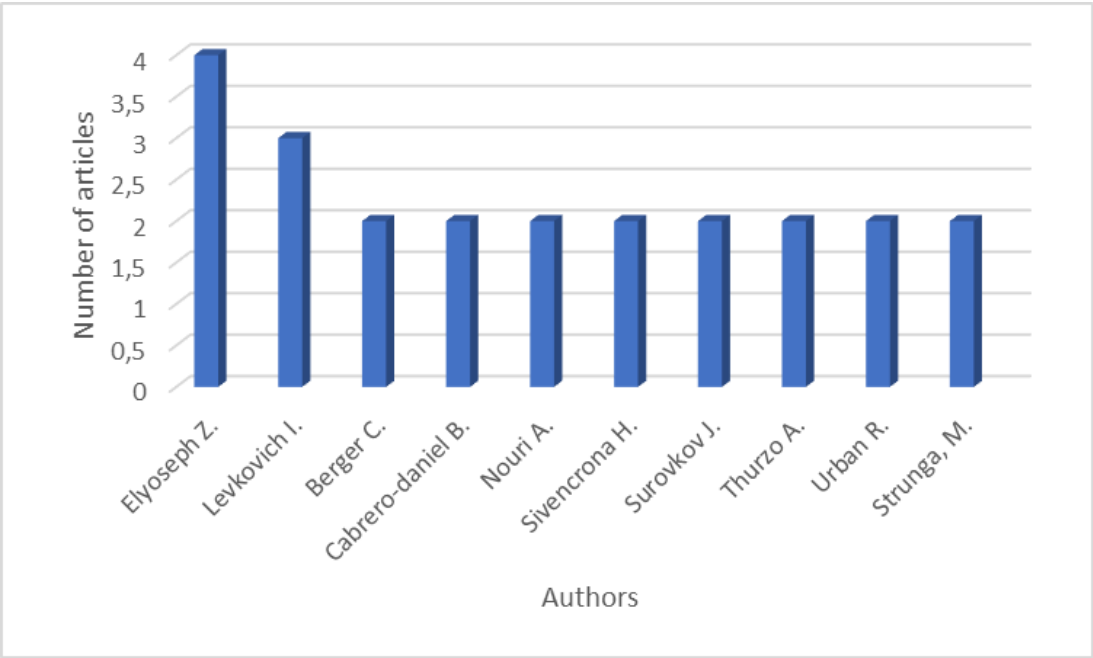

Figure 3. List of top authors published on ChatGPT in the assessment.

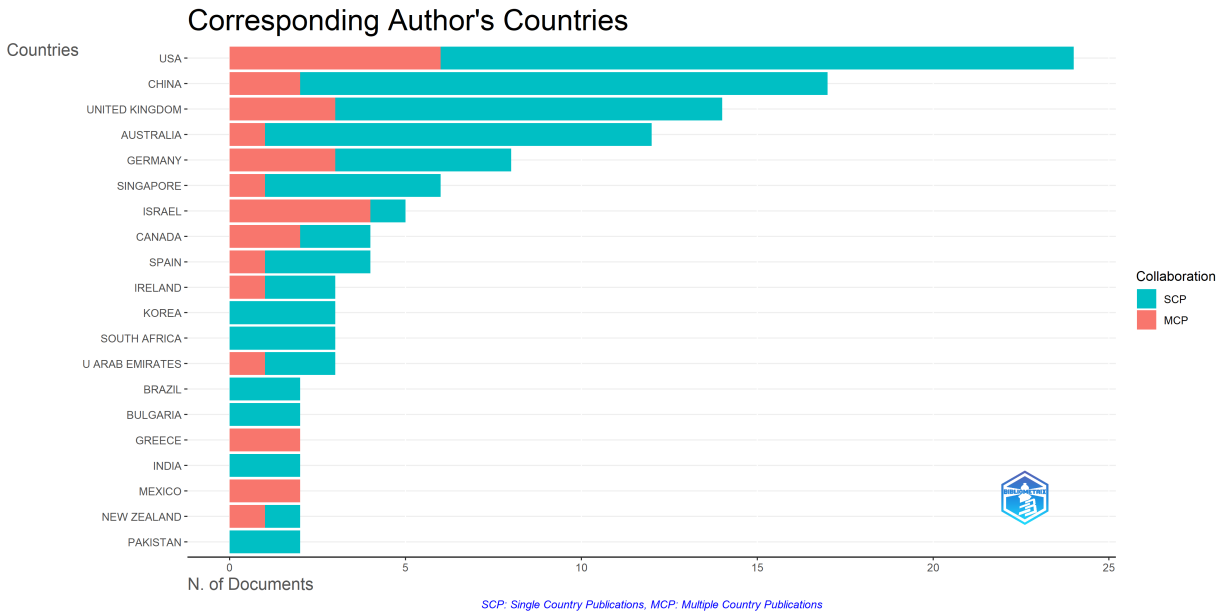

Figure 4. Corresponding authors' countries.

3.3. Co-Occurrence of Keyword

A total of 706 keywords appear on ChatGPT in the assessment, and meet 30 of the five author-keyword co-occurrences threshold, and are divided into 5 clusters. The author's keyword co-occurrence is presented in **Figure 5**. The authors' keywords distributed in different clusters are given below:

- Cluster 1—red colour consists of 10 keywords such as "ai", "chatbot", "chatbots", "examination", "large language models", "machine learning", "medical education", "natural language processing", "nlp", "quality assessment".
- Cluster 2—green colour consists of 9 keywords such as "academic integrity", "artificial intelligence in edu", "assessment", "assessment design", "authentic assessment", "chatgpt", "education", "generative artificial intelligence", "higher education".
- Cluster 3—blue colour consists of 7 keywords such as "artificial intelligence", "artificial-intelligence", "assessment", "diagnosis", "gpt-4", "medicine", "risk assessment".
- Cluster 4—yellow colour consists of 3 keywords such as "large language model", "llm", "prompt engineering".
- Cluster 5—purple colour consists of 1 keyword, such as "generative AI".

Figure 6 represents the frequency of authors' keywords, which can be visualized in a word cloud representation. The most frequent term is "artificial intelligence", followed by "education". Other terms include "impact", "tests", "performance", "tool", "diagnosis", "rubrics", "interpersonal-psychological theory", and "system". They provide insights into the most frequently discussed topics, which can be useful for further research.

3.4. Three Field Plots on ChatGPT in Education

The Sankey diagram illustrates the interconnectedness between key research descriptors, authors, and their countries of affiliation in the field of AI-driven education, particularly focusing on ChatGPT and assessment. Among the descriptors, "ChatGPT" emerges as the most dominant topic, frequently linked with others such as "assessment," "artificial intelligence," "LLM (Large Language Model)," and "education," indicating a strong research interest at the intersection of AI technologies and educational practices. The most active contributors in this domain include Elyoseph Z and Levkovich I, followed by authors like Griewatz J, Festl-Wietek T, Friederike F, and Berger C, who are engaged with multiple themes and contribute significantly to the discourse on AI in education. Geographically, Germany and Israel are the most represented countries, reflecting robust research activity, followed by notable contributions

from the United Kingdom, the United States, and China. The diagram highlights a growing international and interdisciplinary focus on applying large language models like ChatGPT for educational assessment, showcasing a trend towards integrating AI tools into teaching and evaluation methodologies (**Figure 7**).

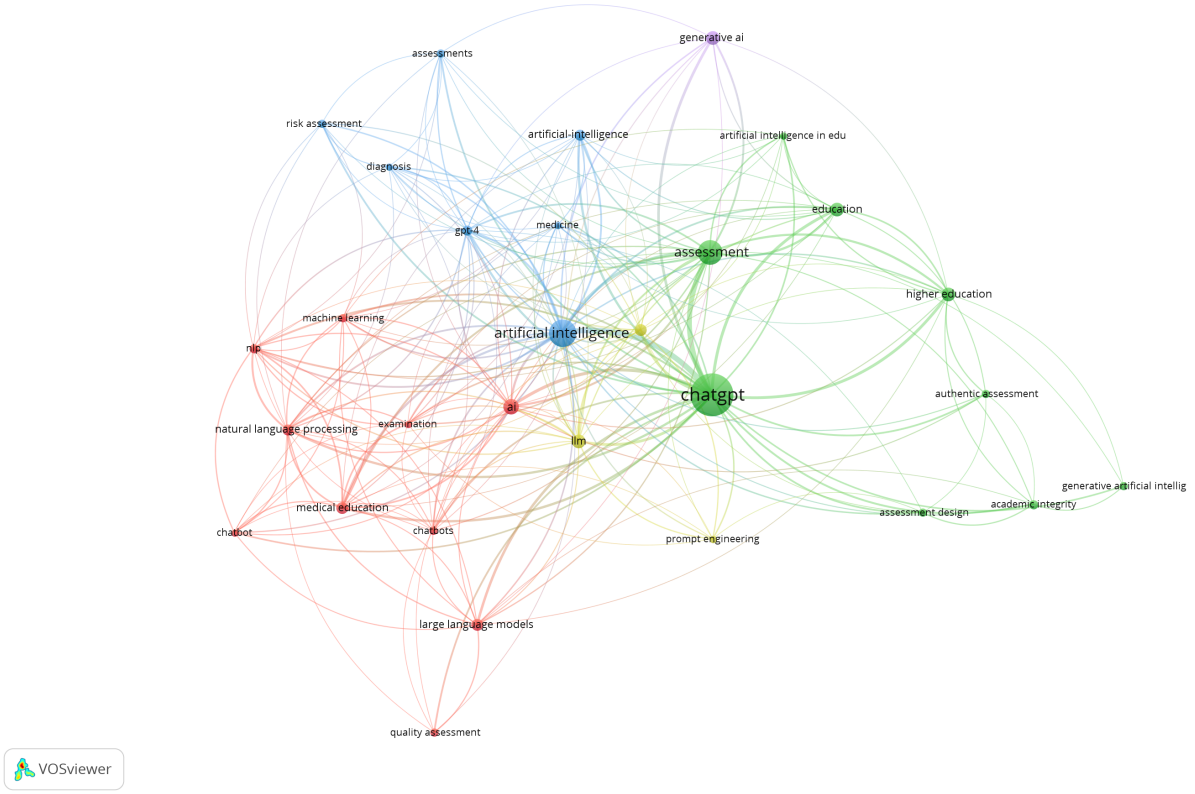

Figure 5. Author keyword co-occurrence network.

Figure 6. Author keyword cloud.

Figure 8 visualizes the connections among Authors (AU), Descriptors (DE), and Cited References (CR), illustrating the intellectual structure and citation flow within research on AI and educational technologies, particularly focusing on ChatGPT and large language models. The analysis reveals that "ChatGPT" remains the most dominant descriptor, closely followed by "LLM", "assessment", and "artificial intelligence", indicating that these concepts form the core thematic focus of the analyzed research. These descriptors are widely cited across various high-impact studies, such as Rudolph et al.[24], Cotton et al.[7], Kung et al.[25], and Kasneci et al.[26], reflecting their academic relevance. Authors like Jiang Y, Elyoseph Z, and Levkowich I are highly engaged with multiple descriptors, contributing significantly

to research on AI-assisted education and its implications in assessment and higher education. Notably, the authors' contributions often bridge multiple themes, such as combining "ChatGPT" with "assessment" and "higher education," showing the interdisciplinary nature of the field. The cited references span journals such as *J. Applied Learning Teaching*, *Innovative Education and Teaching International*, *PLOS Digital Health*, and *JMIR Medical Education*, indicating a broad and cross-disciplinary interest, especially in the intersection of AI, education, and digital health. The diagram also shows increasing references to medical and healthcare-focused publications (e.g., Sallam[27], Alkaissi and McFarlane[28]), highlighting a growing trend of leveraging generative AI tools like ChatGPT in medical education and clinical training environments. Overall, this Sankey diagram demonstrates that current research is not only highly collaborative across authors but also grounded in a robust set of citations that reflect the rapidly evolving academic discourse on AI in education. The centrality of descriptors like "ChatGPT" and "assessment" and their widespread citation underscore the field's growing interest in practical applications of generative AI in formal learning environments.

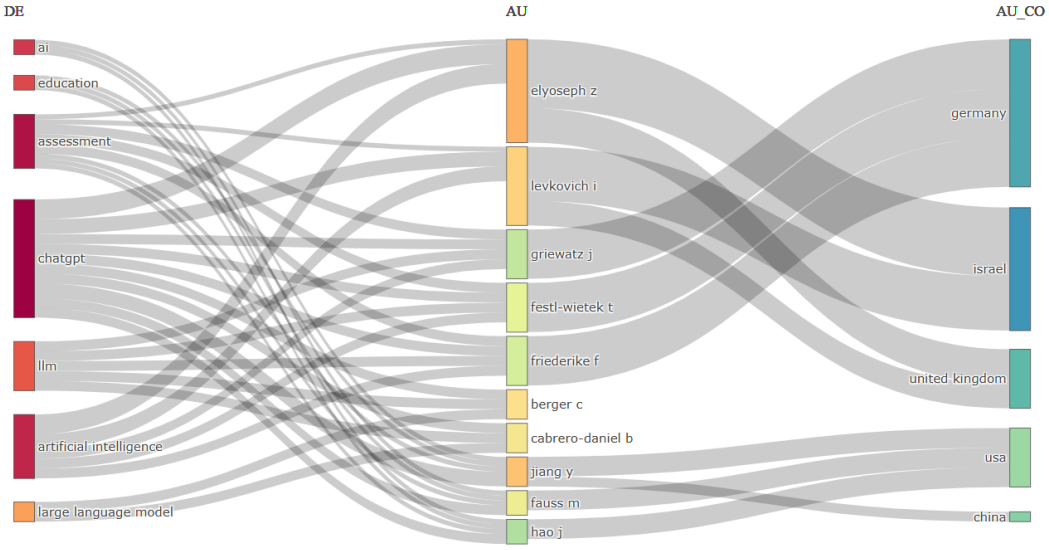

Figure 7. Three field plot author (middle), author keyword (right), and country (left).

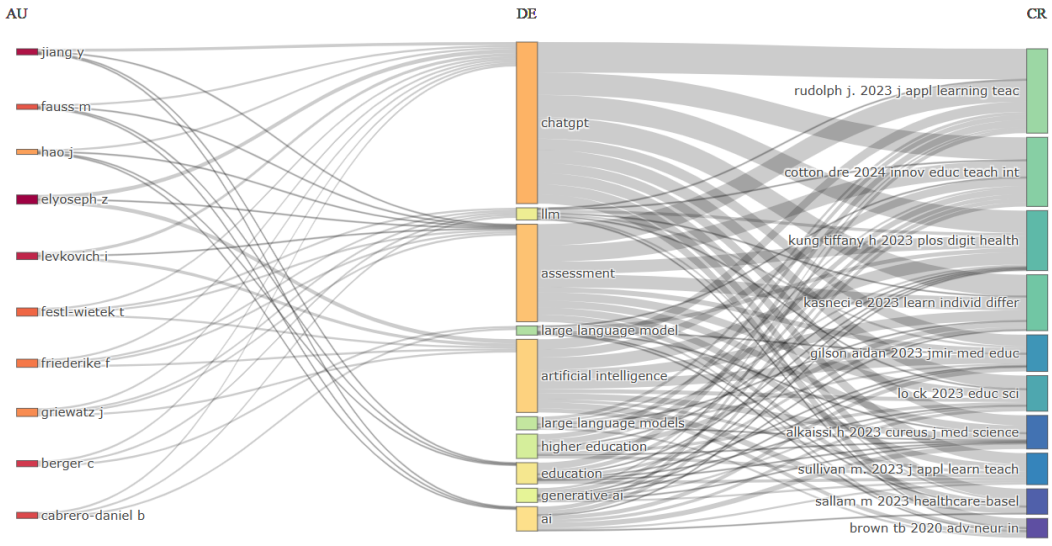

Figure 8. Three field plot author keywords (middle), author (right), and cited reference (left).

4. Discussion

The fast growth of artificial intelligence (AI), especially the creation of huge language models like ChatGPT, has had a big impact on many areas, such as education, healthcare, and psychological testing. Researchers are becoming more and more interested in using ChatGPT for testing as part of this technological change. The main goal of this bibliometric evaluation is to look at scholarly papers from the last two years (2023–2024) and see how ChatGPT is being used in assessments. The review looks for the most important contributors, the most popular study areas, the most prolific countries and publications, and gaps that show where future research might go. The bibliometric study shows that Z. Elyoseph is the most important person in this topic, with four papers about how to use ChatGPT in assessments. I. Levkovich is next, with three works on the subject. A lot of their work looks at how people and AI interact, especially how ChatGPT can be used in high-stakes psychological situations. Their research has looked at ChatGPT's involvement in assessing the risk of suicide and measuring emotional awareness in particular. This research represents a big step towards using AI tools in sensitive areas of mental health, where it's important to be able to understand things in a nuanced way and handle personal information in an ethical way. The analysis looks at the geographical distribution of corresponding authors, as well as highlighting the best authors. The United States comes in first place for scholarly work in this area, followed by China. There are probably a number of reasons why this is so well-known. First, both countries are known for giving a lot of money and support to AI research and development. Second, the fact that digital infrastructure and expertise are widely available makes it easier to use and evaluate new technologies like ChatGPT in schools and hospitals. Because of this, these countries have become global centres for AI research and innovation.We did a keyword co-occurrence analysis to find common themes and study topics in the chosen body of literature. This is a keyword co-occurrence network made with VOSviewer. The clustering in the picture shows groups of keywords that frequently appear together in writing about ChatGPT and AI in education. The green cluster, which has "ChatGPT" as its main node, is tightly linked to words like "assessment," "prompt engineering," "higher education," and "automated assessment." This indicates a growing interest in the application of generative AI technologies in learning assessment procedures. The red cluster focuses on "large language models," "medical education," and "natural language processing," which shows that the medical and clinical education fields are very interested in these topics. The blue cluster has the words "machine learning," "diagnosis," and "risk assessment," which suggests that there are some similarities between AI applications in decision-making and evaluation systems. The purple and light blue clusters, which include terms like "artificial intelligence in education" and "ethics," indicate bigger conversations on policy and theory. These clusters, however, show more than just common themes. They also show fundamental changes in teaching and institutions. For example, the importance of "assessment" and "automated feedback" shows that ChatGPT is not only being explored as a new technology, but it is also starting to change how teachers plan and give tests. The fact that "prompt engineering" and "higher education" are in the same cluster shows that researchers are starting to look into how students use AI systems and how the curriculum may need to change. This means that schools will have to change how they teach, moving from traditional tests to AI-assisted, formative assessment models. **Figure 2** is a word cloud that shows how often certain words appear. Words like "artificial intelligence," "education," "impact," and "tests" stand out the most in the picture. The fact that these phrases are so big shows how common they are in the literature. This image is more descriptive than analytical, but it backs up image 1's conclusions by showing how important AI's effect is on educational outcomes, performance evaluation, and behavioural learning systems. Words like "chatgpt," "health behaviour," and "rubrics" show that research may be done in many different sectors, especially in public health and instructional design. These charts and graphs show more than just numbers; they show how AI technologies like ChatGPT are becoming more common in both the theoretical and practical parts of schooling. The trends found show that AI is changing the way schools make decisions, offer content, and grade students. This thematic grouping is important because it fits with recent requests in the literature for more focus on ethics, accessibility, and teacher preparation, especially as AI technology becomes more common in schools.Some recent studies have looked at how well ChatGPT works in real-life classrooms. For example, Eltobgy et al. looked into how it may be used in tourist education and discovered

that ChatGPT made a big difference in how well students did on all levels of Bloom's Taxonomy, from remembering and understanding to applying, analyzing, evaluating, and producing. The study found that ChatGPT could explain up to 54% of the differences in higher-order cognitive skills, showing that it is a very useful tool for learning [29]. But the scientists warned that this effectiveness relied on careful integration, since using it on the surface may lead to passive or shallow learning. In the same way, Cain and Rajan looked into using ChatGPT as a virtual tutor for pharmacy students in their third year. They added that 99% of students thought the tool helped them grasp the material better, and more than 80% felt it made them feel more confident about their exams. It just took the teacher two hours to set up the tutor, which was speedy. Even though these results were good, the authors stressed that teachers should check ChatGPT's output for possible factual errors [30]. Al Shloul et al. used ChatGPT in activity-based learning settings in another study. Their research indicated that students who used ChatGPT were more engaged, more motivated, better at critical thinking, and did better in school than kids who were in traditional settings. The authors, on the other hand, warned that using AI without structure could make pupils too reliant on it, which could make it harder for them to learn on their own unless they are guided properly [31]. Lastly, Samala et al.'s publication in the International Journal of Evaluation and Research in Education gave a full bibliometric and thematic evaluation of ChatGPT's uses in education. It showed how the tool could help with creativity, individualized learning, and making teaching more productive. At the same time, it stressed ethical hazards like false information (hallucinations), possible plagiarism, and the lack of rules from schools on how to utilize AI responsibly in education [32]. Some surveys and systematic reviews have looked into the function of ChatGPT and other large language models (LLMs) in education in the last few years. They have mostly focused on how these models can be used for teaching, learning, testing, and writing in school. For instance, Lavidas et al. did a thorough survey of humanities and social sciences students' plans to use AI tools in school settings. They found that most students had positive views, but also pointed out that students need to be more digitally literate and have clearer teaching frameworks to help them use the tools [11]. Aravantinos et al. also did a thorough literature analysis of AI applications in primary education. They focused on both the educational potential of tools like ChatGPT and the growing worries about how to use them ethically and how equipped teachers are to utilize them [12]. Uğraş et al. did a study of teachers' experiences using ChatGPT-supported learning in early STEM education and found that students were more motivated and curious, but were also worried about AI dependence and content correctness [13]. These studies offer a lot to help us understand how ChatGPT was first used in education. However, they all share one major problem: they are too focused on their specific contexts. Most polls only look at one country, age group, or topic of study, which makes it hard to apply the results to larger, global trends in education. Also, these surveys often involve qualitative analysis or self-reported user impressions, which are useful but don't give a whole picture of how ChatGPT is being talked about and written about in the academic literature over time, across disciplines, and geographic boundaries. They also do not often provide quantitative mapping of co-authorship networks, topic progression, or citation trends, which are all important for showing how this research subject is changing and increasing quickly. Even when PRISMA protocols are used, like in the reviews by Aravantinos et al. [12] and Lavidas et al. [11], the research that comes out of them typically does not have bibliometric visualization or data-driven mapping of scholarly communication. So, even though previous surveys give us some idea of what teachers and students think about ChatGPT, they do not tell us much about how the academic research community is using it, who the leading researchers are, what the most important themes are, or what intellectual gaps still exist. This work fills in the gap by doing a full bibliometric and visual analysis of 150 peer-reviewed articles published between 2023 and 2024, which were found in the Web of Science database. This study looks at publishing trends, co-authorship relationships, institutional and country-level productivity, and keyword co-occurrence patterns using bibliometric tools like VOSviewer and Biblioshiny to give a big-picture view of ChatGPT's significance in educational research. It does this by finding not only the most important topic clusters, including AI-assisted assessment, language acquisition, and feedback automation, but also areas that haven't been studied enough, like ethical implications, cross-cultural adaptation, and policy creation. This study is different from past survey-based assessments

since it gives objective, data-driven information about how the field is changing, who the main contributors are, and where future research might go. This study adds to prior perception-based studies and addresses a major research gap by showing how ChatGPT is changing modern educational discourse in a way that is based on evidence. Recent real-world research from Western countries has greatly helped us understand how ChatGPT and other large language models are being used in real classrooms, showing both their strengths and weaknesses. For instance, Von Garrel and Mayer's [33] large-scale survey in Germany included over 6,300 university students and found that almost two-thirds of those who answered said they used ChatGPT or GPT-4 for schoolwork. Students in the humanities, engineering, math, and natural sciences used it the most. Students mostly used ChatGPT to help them understand difficult academic ideas and to help them learn on their own, especially in areas that required a lot of readingn. The study also discovered that master's students utilised ChatGPT more often than PhD students. This shows that the use of AI varies by level of education. The fact that so many people use ChatGPT on their own shows how useful it is seen as a personal learning tool in higher education, especially in Germany, where AI tools are quickly becoming a part of everyday academic life. Borges et al. [34] carried out an empirical test at EPFL in Switzerland to see if ChatGPT could solve exam questions from 50 engineering courses. This added to the findings. Their results showed that ChatGPT answered about 66% of the questions correctly. This is a very good level of performance, especially given how sophisticated and specialised engineering information is. This result shows that ChatGPT could be used as an extra learning aid to help students remember what they learnt in school. The scientists also warned that this kind of performance would make standard assessment formats less reliable and advocated for a re-evaluation of academic testing methodologies in AI-supported learning environments. This study brings up some key considerations about teaching: if AI tools can answer common problems, then tests should focus more on evaluating students' ability to think critically, be creative, and make moral decisions. In North America, Ganjavi et al. [35] asked medical students about their knowledge of, use of, and opinions on ChatGPT. Their research showed that 96% of medical students knew about ChatGPT and 52% had used it for schoolwork, like making study notes, summarising course materials, and getting ready for tests. Many students liked how easy and flexible ChatGPT was to use, but they were also worried about how accurate the information was, how biased it could be, and the fact that there were no official rules for how to use it ethically. These worries are especially important in medical education, where being accurate and following professional norms are very important. The report says that schools and colleges need to make explicit rules and training programs to ensure that AI is used responsibly in health-related education. These three studies, which took place in Germany, Switzerland, and the United States, give us a lot of information on how students are using ChatGPT in real-life educational settings. They show that ChatGPT works well and is widely accepted, but they also point out the ethical, instructional, and institutional problems that arise with using it. The current study's bibliometric analysis is placed in a more worldwide, discipline-diverse, and empirically grounded context by adding these findings to the current discussion. This makes its conclusions more relevant and rigorous.

5. Conclusions

In conclusion, this work adds a lot to the expanding corpus of academic research that looks into how ChatGPT could be used in language education assessment. The study looks at two years of academic work in this new field and gives us useful information on how artificial intelligence, especially large language models (LLMs) like ChatGPT, are being seen as game-changing tools for testing language skills. The bibliometric research shows that more and more people around the world are interested in using AI technologies in language testing, especially those that help students improve their speaking, writing, and overall communication skills in a second or foreign language. The results show that the research landscape is broad and always changing, thanks to the work of leading countries, renowned academics, high-impact publications, and large academic institutions. This worldwide scholarly engagement shows that people around the world agree on the educational potential of AI and are becoming more committed to critically and constructively studying new technologies. The study shows how conversations about ChatGPT and other AI tools are changing in the fields of applied linguistics and educational evaluation

by visually mapping co-authorship, keyword co-occurrence, and publication patterns. One of the most important things this evaluation does is point out research themes that keep coming up. Some of the most important issues include the roles of AI-assisted feedback, the creation of automated evaluation systems, and ethical issues connected to academic integrity. These topics not only show what people are interested in right now, but they also point to crucial areas for future research. For example, ChatGPT is often praised for being able to tailor lessons to each student, give instant feedback, and promote learner autonomy. However, there is still a lack of empirical research that shows how well it works in different linguistic, cultural, and institutional settings. This difference between what is possible in theory and what is possible in practice shows that we need more detailed and context-sensitive research. There is a lot more to learn about how AI tools work in different classrooms, how students react to and use AI-generated feedback, and how teachers use these tools in their teaching. In addition, concerns of access and fairness need to be looked at, because if AI resources are not available to everyone equally, it can unintentionally make achievement inequalities bigger and make language assessment processes less fair. We also need to pay more attention to ethical issues. Concerns about plagiarism, the originality of student work, and the trustworthiness of AI in high-stakes testing situations show how hard it is to use these kinds of technology responsibly. These worries show how important it is to have clear evaluation frameworks and teacher training that connect new technologies with good educational practices. In the end, this bibliometric evaluation will be a key resource for future and continuing research into language instruction that uses AI. It not only shows where the current research stands, but it also encourages more in-depth study of the ethical, technological, and pedagogical aspects of employing ChatGPT in language learning settings. We need to keep changing the way we make sure that AI use leads to meaningful learning, fair access, and accurate assessment results as it continues to grow.

5.1. Limitations

The study has a number of limitations. First, the articles analysed were retrieved only from the WOS online database, which is one of the largest databases with a large number of articles, which meet quality requirements. Second, articles only in the English language were reviewed, while reviewing articles in other languages could provide different results.

Author Contributions

Conceptualization, G.S.; methodology, L.S.; validation, M.N.; formal analysis, N.I.; investigation, G.S.; resources, G.S.; data curation, N.I.; writing—original draft preparation, L.S.; writing—review and editing, M.N.; visualization, L.S.; supervision, N.I.; project administration, G.S.; funding acquisition, M.N. All authors have read and agreed to the published version of the manuscript.

Funding

This research has not yet received external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon request.

Acknowledgments

The authors express their appreciation to the “Bibliometric research” course, implemented by the rector of «Tashkent Institute of Irrigation and Agricultural Mechanization Engineers» National Research University, Bakhadir Mirzaev, for providing the direction and inspiration to carry out this research. Special thanks to Dr. Mukhiddin Juliev and Dr. Zulfiya Kannazarova for their guidance and supervision during the “Bibliometric Research” course. Additionally, the authors are grateful to the anonymous reviewers of earlier drafts for their constructive comments, which significantly contributed to the improvement of the manuscript.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

[1] Mavropoulou, E., Arvanitis, P., 2024. Leveraging Artificial Intelligence to Enhance Oral Communicative Skills in French Language Education. EQR. 7(4). Available from: https://www.asianinstituteofresearch.org/EQRarchives/leveraging-artificial-intelligence-to-enhance-oral-communicative-skills-in-french-language-education

[2] Owan, V.J., Abang, K.B., Idika, D.O., et al., 2023. Exploring the potential of artificial intelligence tools in educational measurement and assessment. EURASIA Journal of Mathematics, Science and Technology Education. 19(8), em2307. DOI: https://doi.org/10.29333/ejmste/13428

[3] Yang, H., Gao, C., Shen, H.Z., 2024. Learner interaction with, and response to, AI-programmed automated writing evaluation feedback in EFL writing: An exploratory study. Education and Information Technologies. 29(4), 3837–3858. DOI: https://doi.org/10.1007/s10639-023-11991-3

[4] Lachini, K., 2025. Rethinking Language Assessment in the Age of AI: A Call for a Paradigm Shift. Available from: https://www.ssrn.com/abstract=5043405 (cited 9 June 2025).

[5] Neupane, B.P., Paudel, P., Dahal, N., et al., 2025. English Language Teaching in the Age of Artificial Intelligence: Tools, Techniques, and Methodologies. Advances in Mobile Learning Educational Research. 5(1), 1356–1369. DOI: https://doi.org/10.25082/AMLER.2025.01.011

[6] Zhai, C., Wibowo, S., Li, L.D., 2024. Evaluating the AI dialogue System's intercultural, humorous, and empathetic dimensions in English language learning: A case study. Computers and Education: Artificial Intelligence. 7, 100262. DOI: https://doi.org/10.1016/j.caeai.2024.100262

[7] Cotton, D.R.E., Cotton, P.A., Shipway, J.R., 2024. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International. 61(2), 228–239. DOI: https://doi.org/10.1080/14703297.2023.2190148

[8] Zhao, Y., 2025. Artificial Intelligence and Education: End the Grammar of Schooling. ECNU Review of Education. 8(1), 3–20. DOI: https://doi.org/10.1177/20965311241265124

[9] Khan, M.A., Kurbonova, O., Abdullaev, D., et al., 2024. Is AI-assisted assessment liable to evaluate young learners? Parents support, teacher support, immunity, and resilience are in focus in testing vocabulary learning. Language Testing in Asia. 14(1), 48. DOI: https://doi.org/10.1186/s40468-024-00324-x

[10] Tian, Z., Wang, C., 2025. Rethinking Language Education in the Age of Generative AI, 1st ed. Routledge: New York, NY, USA. Available from: https://www.taylorfrancis.com/books/9781003426929

[11] Lavidas, K., Voulgari, I., Papadakis, S., et al., 2024. Determinants of Humanities and Social Sciences Students' Intentions to Use Artificial Intelligence Applications for Academic Purposes. Information. 15(6), 314. DOI: https://doi.org/10.3390/info15060314

[12] Aravantinos, S., Lavidas, K., Voulgari, I., et al., 2024. Educational Approaches with AI in Primary School Settings: A Systematic Review of the Literature Available in Scopus. Education Sciences. 14(7), 744. DOI: https://doi.org/10.3390/educsci14070744

[13] Uğraş, H., Uğraş, M., Papadakis, S., et al., 2025. Innovative Early Childhood STEM Education with ChatGPT: Teacher Perspectives. Tech Know Learn. 30(2), 809–831. DOI: https://doi.org/10.1007/s10758-024-09804-8

[14] Page, M.J., McKenzie, J.E., Bossuyt, P.M., et al., 2021. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. BMJ. 372, n71. DOI: https://doi.org/10.1136/bmj.n71

[15] Mardieva, D., Bakiyeva, G., Kannazarova, Z., et al., 2024. A Bibliometric Review: Interventions for Enhancing Speaking Skills in Non-English-Speaking Contexts. XL. 17(4), 195–224. DOI: https://doi.org/10.30564/fls.v7i7.10284

[16] Elyoseph, Z., Hadar-Shoval, D., Asraf, K., et al., 2023. ChatGPT Outperforms Humans in Emotional Awareness Evaluations. Front Psychol. 14, 1199058. DOI: https://doi.org/10.3389/fpsyg.2023.1199058

[17] Elyoseph, Z., Levkovich, I., 2023. Beyond Human Expertise: The Promise and Limitations of ChatGPT in Suicide Risk Assessment. Front Psychiatry. 14, 1213141. DOI: https://doi.org/10.3389/fpsyt.2023.1213141

[18] Levkovich, I., Elyoseph, Z., 2023. Suicide Risk Assessments Through the Eyes of ChatGPT-3.5 Versus ChatGPT-4: Vignette Study. JMIR Ment Health. 10, e51232. DOI: https://doi.org/10.2196/51232

[19] Shinan-Altman, S., Elyoseph, Z., Levkovich, I., 2024. The Impact of History of Depression and Access to Weapons on Suicide Risk Assessment: A Comparison of ChatGPT-3.5 and ChatGPT-4. PeerJ. 12, e17468. DOI: https://doi.org/10.7717/peerj.17468

[20] Nouri, A., Cabrero-Daniel, B., Torner, F., et al., 2024. Welcome Your New AI Teammate: On Safety Analysis by Leashing Large Language Models. In Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI, Lisbon, Portugal, 15–19 April 2024; pp. 172–177. DOI:

https://doi.org/10.1145/3644815.3644953

[21] Nouri, A., Cabrero-Daniel, B., Törner, F., et al., 2024. Engineering Safety Requirements for Autonomous Driving with Large Language Models. In Proceedings of the 2024 IEEE 32nd International Requirements Engineering Conference (RE), Reykjavik, Iceland, 12–16 August 2024; pp. 218–228. Available from: https://ieeexplore.ieee.org/document/10628478/

[22] Strunga, M., Urban, R., Surovková, J., et al., 2023. Artificial Intelligence Systems Assisting in the Assessment of the Course and Retention of Orthodontic Treatment. Healthcare. 11(5), 683. DOI: https://doi.org/10.3390/healthcare11050683

[23] Surovková, J., Haluzová, S., Strunga, M., et al., 2023. The New Role of the Dental Assistant and Nurse in the Age of Advanced Artificial Intelligence in Telehealth Orthodontic Care with Dental Monitoring: Preliminary Report. Applied Sciences. 13(8), 5212. DOI: https://doi.org/10.3390/app13085212

[24] Rudolph, J., Tan, S., Tan, S., 2023. ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education? Journal of Applied Learning and Teaching. 6, 342–363. DOI: https://doi.org/10.37074/jalt.2023.6.1.9

[25] Kung, T.H., Cheatham, M., Medenilla, A., et al., 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLoS Digit Health. 2(2), e0000198. DOI: https://doi.org/10.1371/journal.pdig.0000198

[26] Kasneci, E., Sessler, K., Küchemann, S., et al., 2023. ChatGPT for good? On opportunities and challenges of large language models for education, Learning and Individual Differences. 103, 102274. DOI: https://doi.org/10.1016/j.lindif.2023.102274

[27] Sallam, M., 2023. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare (Basel). 11(6), 887. DOI: https://doi.org/10.3390/healthcare11060887

[28] Alkaissi, H., McFarlane, S.I., 2023. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. Cureus. 15(2), e35179. DOI: https://doi.org/10.7759/cureus.35179

[29] Eltobgy, A.E., Abaza, W., Abd, Y., 2024. Evaluating the Effectiveness of ChatGPT in Enhancing Cognitive Learning Outcomes in Tourism Higher Education Based on Biggs's 3P Model. Journal of Association of Arab Universities for Tourism and Hospitality. 27(1), 256–276. DOI: https://doi.org/10.21608/jaauth.2024.320773.1608

[30] Cain, J., Rajan, A.S., 2024. Proof of Concept of ChatGPT as a Virtual Tutor. Am J Pharm Educ. 88(12), 101333. DOI: https://doi.org/10.1016/j.ajpe.2024.101333

[31] Al Shloul, T., Mazhar, T., Abbas, Q., et al., 2024. Role of activity-based learning and ChatGPT on students' performance in education. Computers and Education: Artificial Intelligence. 6, 100219. DOI: https://doi.org/10.1016/j.caeai.2024.100219

[32] Samala, A.D., Sokolova, E.V., Grassini, S., et al., 2024. ChatGPT: A Bibliometric Analysis and Visualization of Emerging Educational Trends, Challenges, and Applications. International Journal of Educational Research and Evaluation (IJERE). 13(4), 2374. DOI: http://doi.org/10.11591/ijere.v13i4.28119

[33] Von Garrel, J., Mayer, J., 2023. Artificial Intelligence in Studies—Use of ChatGPT and AI-Based Tools Among Students in Germany. Humanities and Social Sciences Communications. 10(1), 799. DOI: https://doi.org/10.1057/s41599-023-02304-7

[34] Borges, B., Foroutan, N., Bayazit, D., et al., 2024. Could ChatGPT Get an Engineering Degree? Evaluating Higher Education Vulnerability to AI Assistants. Proceedings of the National Academy of Sciences (PNAS). 121(49), e2414955121. DOI: https://doi.org/10.1073/pnas.2414955121

[35] Ganjavi, C., Eppler, M., O'Brien, D., et al., 2024. ChatGPT and Large Language Models (LLMs) Awareness and Use: A Prospective Cross-Sectional Survey of U.S. Medical Students. PLOS Digital Health. 3(9), e0000596. DOI: https://doi.org/10.1371/journal.pdig.0000596