International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-11

https://doi.org/10.70917/ijcisim-2026-3353
Article

Artificial Intelligence in Higher Education: Bibliometric Analysis and Experimental Evidence on Students Learning

Zhang Yinfeng [1] and Melissa Ng Lee Yen Abdullah [1,*]

[1] School of Educational Studies, Universiti Sains Malaysia (USM), 11800, Penang, Malaysia
* Correspondence author: melissa@usm.my

Abstract: This study examines the development and educational implications of artificial intelligence (AI) in higher education through a combination of bibliometric analysis and experimental evidence. First, publications indexed in the Scopus database from 2000 to 2024 were analyzed to map the evolution of research on AI in higher education, with attention to publication trends, thematic concentrations, and influential studies. Text mining was conducted on titles and abstracts, and term frequency-inverse document frequency (TF-IDF) weighting was used to identify representative terms. K-means clustering and Latent Dirichlet Allocation (LDA) topic modeling were then applied to detect major research themes, while PageRank analysis of the citation network was used to identify publications with high structural influence in the field. Alongside the bibliometric analysis, the study conducted an experimental investigation of ChatGPT as a formative assessment tool. Student responses were submitted to ChatGPT to generate automated feedback and grades, and the outputs were examined in terms of feedback type, instructional value, grading consistency, and agreement with human evaluation. The results suggest that students receiving ChatGPT-supported assistance showed better learning performance than those without AI support. The feedback generated by ChatGPT contained corrective, explanatory, and motivational elements, indicating its potential to provide both cognitive and affective support during learning. In addition, the comparison between AI-generated grades and human-assigned grades showed a high level of alignment, with limited evidence of systematic bias.

Keywords: Artificial Intelligence (AI); Higher Education; Bibliometric Analysis; Student Learning

1. Introduction

Artificial intelligence (AI) has long been discussed in educational research, but its role in higher education remains an evolving issue. Although early work on AI in education focused mainly on intelligent tutoring, adaptive instruction, and computer-supported learning, recent advances have made AI tools more visible in everyday university teaching and learning [1]. In particular, the rapid development of AI-driven educational technologies has led to a substantial increase in scholarly attention to this field [2]. This growing body of research reflects a broader expectation that AI may reshape how universities support instruction, assessment, student services, and learning management [3].

AI is now being applied across a wide range of academic and administrative activities in higher education. Intelligent tutoring systems, for example, can provide individualized learning support by adjusting learning materials and feedback according to students' needs [4-5]. Predictive analytics can be used to identify students who may be at academic risk by analyzing patterns in learning behavior and performance. AI-based automation is also increasingly used in routine educational tasks, including grading, advising, feedback generation, and administrative decision-making. In addition, universities have begun to adopt virtual teaching assistants and chatbot systems to answer student questions, while data-informed platforms are used to support student services and institutional planning. These applications are often regarded as promising because they may improve learning support, increase student engagement, and enable more timely educational intervention.

At the same time, the growing use of AI in higher education has raised important concerns. The emergence of generative AI systems, especially tools such as ChatGPT, has intensified debates about academic integrity, authorship, and the authenticity of student work [6]. Many instructors are concerned that excessive dependence on AI tools may weaken students' independent thinking, writing ability, and problem-solving skills. Ethical issues also remain unresolved, particularly those related to algorithmic bias, data privacy, transparency, and accountability in educational decision-making. Although international organizations and higher education institutions have begun to issue guidelines for the responsible use of AI, translating these principles into classroom practice remains difficult. Critical reviews have also warned that some discussions of AI in higher education are overly optimistic and pay insufficient attention to pedagogical risks and long-term consequences. Therefore, AI should not be viewed simply as a technological solution; it must be examined in relation to its educational value, limitations, and possible unintended effects.

The expansion of AI applications has been accompanied by a rapid growth of research on AI in higher education. Recent literature reviews and meta-analyses show that this field now covers diverse disciplines, learning contexts, and technological applications [7]. Zawacki-Richter et al. identified four major areas of AI application in higher education: student profiling and performance prediction, assessment and evaluation, adaptive learning and personalization, and intelligent tutoring systems [1]. Similarly, Chu et al. found that the prediction of student learning outcomes has become a central topic among highly cited studies, with many AI techniques being developed to provide personalized support for learners [3]. Other reviews have also emphasized the potential benefits of AI, including adaptive instruction, timely feedback, automated assessment, and early identification of students who may need additional academic support.

Despite these positive findings, the actual educational impact of AI remains unsettled. Some studies report clear improvements in student engagement, learning efficiency, or academic performance, while others find only limited or inconsistent effects [7]. This inconsistency suggests that the effectiveness of AI may depend on factors such as the type of tool, instructional design, subject area, student characteristics, and the way AI is integrated into teaching practice. Recent studies on ChatGPT and related generative AI tools have similarly produced mixed conclusions, with some showing measurable benefits and others warning of overreliance or shallow learning. As a result, there is a need for a more systematic and evidence-based understanding of how AI is actually influencing university students' learning.

In response to this need, the present study examines both the development of AI research in higher education and its relationship with student learning outcomes. Bibliometric analysis is used as a key methodological approach because it allows the field to be mapped quantitatively at scale. Through bibliometric methods, it is possible to identify publication trends, influential authors and studies, citation structures, collaboration patterns, and major thematic clusters. Previous bibliometric studies have already provided useful overviews of AI in education and related subfields [8-10]. Building on this work, the present study focuses more specifically on the intersection between AI applications and student learning in higher education.

In addition to mapping the research landscape, this study also reviews evidence concerning the effects of AI on students' learning outcomes, including academic performance, engagement, feedback, and assessment-related learning support [11-12]. By combining a broad bibliometric overview with an analysis of educational impact, the study aims to clarify how research on AI in higher education has evolved and whether this development has translated into measurable benefits for students. This integrated perspective is important because the rapid increase in AI-related publications does not necessarily mean that AI tools are pedagogically effective in practice. A closer examination is therefore needed to distinguish technological enthusiasm from empirically supported educational value.

This study makes four main contributions. First, it provides a bibliometric analysis of research on AI in higher education, identifying publication trends, influential contributors, and citation patterns in the field. Second, it classifies major research themes related to AI applications in higher education, including intelligent tutoring systems, predictive analytics, AI-supported content creation, automated assessment, and adaptive learning environments. Third, it synthesizes existing evidence on the impact of AI-driven tools on student learning, with attention to academic performance, engagement, feedback, and other educational outcomes. Finally, based on the findings, the study discusses remaining gaps in the literature and proposes directions for future research, particularly regarding the long-term effects of AI, ethical risks, equity issues, and the responsible integration of AI into higher education practice. These findings provide a basis for further research and offer practical implications for educators, researchers, and policymakers seeking to use AI in ways that genuinely support student learning.

2. Related Work

2.1. Learning Outcomes and AI Integration in Higher Education

Research on artificial intelligence (AI) in higher education has expanded rapidly, but the central question is no longer whether AI can be introduced into university teaching. A more important issue is whether these technologies can produce meaningful and measurable improvements in student learning. Existing reviews show that AI applications in higher education are distributed across several major areas, including student profiling and predictive analytics, assessment and evaluation, adaptive learning and personalization, and intelligent tutoring systems [13]. Zawacki-Richter et al. further noted that much of the early literature was concentrated in STEM and computer science contexts, with many studies emphasizing system development rather than sustained pedagogical evaluation [13]. This concern has also been echoed in later reviews. Bond et al. found that research on AI in higher education grew quickly between 2018 and 2023, but argued that stronger research design, interdisciplinary collaboration, and ethical reflection are still needed when evaluating AI-based educational innovations [14].

Among existing AI applications, intelligent tutoring systems have produced some of the clearest evidence of learning benefits. Kulik and Fletcher's meta-analysis of 50 controlled studies reported that intelligent tutoring systems improved student test performance by approximately 0.66 standard deviations compared with conventional instruction [15]. This suggests that AI-supported tutoring can generate substantial learning gains when the system is well designed and closely connected to instructional objectives. VanLehn's comparison of human tutoring, computer tutoring, and no tutoring also showed that advanced computer tutors can approach the effectiveness of human tutoring in some learning contexts [16]. These findings indicate that AI can support content mastery, procedural learning, and problem-solving when it provides adaptive feedback and structured guidance.

Other forms of AI integration have also shown potential. Adaptive learning platforms can adjust learning materials according to students' current performance, while learning analytics systems can identify students who may be at risk and support timely intervention. Automated assessment tools may also increase the frequency of feedback, allowing students to revise their work more efficiently. However, the positive effects of these tools are not automatic. Their effectiveness depends on how well they are aligned with course objectives, assessment design, and student needs. For example, Kulik and Fletcher found that learning gains were stronger when assessment tasks were closely matched to the content supported by the tutoring system [17]. Conversely, when AI tools are poorly integrated into the curriculum, their effects may be limited.

Recent studies have also reported mixed or even neutral findings. One investigation involving medical students found no significant difference in examination performance between students who used AI-based study tools and those who did not [19]. Such evidence suggests that AI tools do not necessarily improve learning simply by being available. Without appropriate guidance, students may use AI superficially, become overly dependent on generated answers, or fail to develop deeper understanding. This reflects a broader gap in the field: although AI in education is often presented as highly promising, empirical evidence of its actual learning impact remains uneven. Many studies still focus on technological novelty or short-term performance indicators, while paying less attention to deeper learning, transfer, critical thinking, and long-term retention.

Therefore, existing research supports a balanced interpretation. AI has clear potential to improve learning outcomes in higher education, especially when it is embedded within sound pedagogical design. At the same time, its effects vary across tools, disciplines, student groups, and instructional contexts. This uncertainty makes it necessary to examine AI integration not only through individual case studies, but also through broader evidence-based analysis. For this reason, the present study focuses on how AI has been studied in relation to university students' learning outcomes and seeks to clarify both the demonstrated benefits and the remaining limitations of current research.

2.2. Comparative Evaluation of AI-Driven Educational Tools

A second important strand of related work compares different types of AI-driven educational tools and their effects on learning. As AI applications have become more diverse, researchers have examined intelligent tutoring systems, conversational agents, automated grading systems, adaptive platforms, and generative AI assistants in order to understand their respective strengths and limitations. The available evidence suggests that AI tools can outperform traditional approaches in some contexts, but their effectiveness depends strongly on the type of tool, the learning task, and the way the technology is used.

Intelligent tutoring systems remain one of the most established forms of AI-supported learning. These systems usually guide students through problem-solving processes, diagnose errors, and provide step-by-step feedback. As noted above, meta-analytic evidence shows that they can produce relatively strong learning gains, often around one-half to two-thirds of a standard deviation [15]. A recent controlled experiment involving an AI-based virtual tutor also reported substantial gains in an introductory physics learning setting, where students using the AI tutor achieved higher learning gains in less time than those receiving standard classroom instruction [18]. Such findings suggest that AI tutors can be effective when they are carefully designed around specific learning objectives and supported by evidence-based instructional principles.

Conversational agents and chatbot-based systems offer a different form of support. Unlike full intelligent tutoring systems, chatbots are often used to provide explanations, answer questions, give examples, and maintain interaction outside class time. Their main advantage lies in accessibility: students can receive immediate support whenever they encounter difficulty. However, the learning effects of chatbots appear to be more moderate than those of structured tutoring systems. Laun and Wolff's meta-analysis found an overall small-to-moderate positive effect of educational chatbots on learning outcomes. The effect was stronger in certain conditions, such as text-based interactions, STEM-related subjects, and longer interventions. This indicates that chatbots can be useful learning aids, but they may not replace more comprehensive tutoring systems or direct instructional support.

Automated assessment and feedback systems represent another major category of AI-driven educational tools. These systems can grade short responses, essays, quizzes, or assignments and provide feedback more quickly than human instructors in large-scale teaching contexts. Their main contribution is often efficiency, consistency, and increased feedback frequency. However, compared with human feedback, automated feedback may be less sensitive to creativity, argument quality, emotional support, or context-specific reasoning. For this reason, AI-based assessment is most valuable when it supports formative learning rather than replacing instructor judgment entirely. In practice, automated feedback can help students revise their work, while teachers can focus on higher-level guidance and individualized support.

Generative AI tools, including large language model-based assistants, have further expanded the range of possible AI applications in higher education. These tools can generate explanations, examples, summaries, practice questions, and draft feedback in response to student prompts. Their flexibility makes them attractive for self-directed learning and formative assessment. At the same time, their open-ended nature introduces risks, including inaccurate responses, overreliance, academic integrity concerns, and reduced student effort. Compared with earlier AI systems designed for narrow instructional tasks, generative AI requires more careful pedagogical framing. Students need guidance on how to question, evaluate, and use AI-generated content critically.

The comparative literature therefore suggests that no single AI tool is universally superior. Intelligent tutoring systems are especially useful for structured skill acquisition, adaptive platforms are effective for personalized pacing, chatbots provide accessible conversational support, automated grading systems improve feedback efficiency, and generative AI assistants offer flexible explanation and revision support. However, each tool also has limitations. Human instructors continue to play an essential role in motivation, ethical judgment, emotional support, and the contextualization of learning. For this reason, many scholars recommend blended models in which AI tools supplement rather than replace human teaching [16]. In such models, AI can support repetitive practice, preliminary feedback, and individualized learning pathways, while instructors focus on discussion, interpretation, creativity, and higher-order thinking.

Overall, prior research shows that AI has the capacity to support learning in higher education, but its educational value depends on implementation conditions. Comparative studies have clarified the different contributions of AI tools, but the field still lacks a sufficiently integrated view of how AI research has developed and how its reported benefits relate to student learning outcomes. Existing studies often examine specific technologies or isolated interventions, while fewer works provide a broader, data-driven synthesis of research trends, learning impacts, and remaining gaps. The present study addresses this limitation by combining bibliometric analysis with experimental evidence on AI-supported learning and assessment. This approach makes it possible to situate individual findings within the wider development of AI in higher education and to provide a more systematic basis for future research and practice.

3. Bibliometric Analysis: Data Collection, Preprocessing, and Thematic Findings

This section presents the bibliometric procedure used to examine research on artificial intelligence (AI) in higher education from 2000 to 2024. The analysis covers data collection, text preprocessing, TF-IDF-based document representation, K-means clustering, LDA topic modeling, and the identification of major research themes.

3.1. Data Collection and Preprocessing

Publications were collected from the Scopus database, which was selected for its broad coverage of peer-reviewed research in education, computer science, learning analytics, and educational technology. The search strategy combined AI-related terms, such as "artificial intelligence" and "machine learning," with higher education-related terms, including "higher education," "university," and "university student." The initial search returned approximately 1,000 records.

The retrieved records were then screened according to inclusion and exclusion criteria. Publications were retained if they focused on AI applications in tertiary or university-level education, especially in relation to student learning, assessment, academic performance, learning support, or educational analytics. Studies on K-12 education, purely technical AI research without educational application, dissertations, reports, non-English publications, and irrelevant records were excluded. After removing duplicates and unsuitable items, approximately 800 publications remained for analysis.

For each publication, bibliographic metadata were extracted, including title, authors, publication year, journal or conference source, abstract, author keywords, citation counts, and available reference information. The titles and abstracts were then preprocessed for text analysis. All text was converted to lowercase, stop words were removed, and stemming was applied to reduce words to their root forms. Author keywords were also standardized through lowercasing and punctuation removal. These steps reduced noise in the text data and improved the consistency of subsequent keyword extraction and clustering.

3.2. Text Representation and Clustering

To identify the main research directions in the corpus, the title and abstract of each publication were represented as a bag-of-words vector. Term importance was calculated using term frequency-inverse document frequency (TF-IDF), which gives greater weight to terms that appear frequently in a given document but are less common across the whole corpus. This method helps identify words that are more representative of specific research topics.

For $term\ (i)$ in $document\ (j)$, the term frequency $TF_{i,j}$ is defined as: where $f_{i,j}$ denotes the number of occurrences of $term\ (i)$ in $document\ (j)$, and ($\sum_k f_{k,j}$) represents the total number of terms in that document. The inverse document frequency (IDF_i) is calculated as:

$$IDF_i = \frac{\log N}{n_i} \tag{1}$$

where N is the total number of documents and n_i is the number of documents containing $term\ (i)$. The final TF-IDF weight $w_{i,j}$ is obtained by multiplying term frequency by inverse document frequency:

$$w_{i,j} = TF_{i,j} \times IDF_i \tag{2}$$

Before constructing the TF-IDF matrix, very common terms and extremely rare terms were removed so that the analysis could focus on more informative keywords.

After the TF-IDF matrix was generated, K -means clustering was applied to group publications according to textual similarity. In this method, each document is treated as a point in a high-dimensional feature space, and the algorithm assigns documents to K clusters by minimizing the distance between each document vector and the centroid of its assigned cluster. The objective function is written as:

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \left\| x_i - \mu_k \right\|^2 \tag{3}$$

Here, x_i denotes the TF-IDF vector of $document\ (i)$, μ_k represents the centroid of cluster k, and r_{ik} is an indicator variable. If $document\ (i)$ is assigned to $cluster\ (k)$, then $r_{ik} = 1$; otherwise, $r_{ik} = 0$.

The number of clusters was determined using the elbow method. Candidate values from (K =2) to (K =10) were tested, and (K =5) was selected because it provided a reasonable balance between thematic detail and interpretability. After clustering, each cluster was interpreted by examining the terms with the highest average TF-IDF weights. These terms were used to describe the main topic of each cluster.

Latent Dirichlet Allocation (LDA) was also used as a complementary topic modeling method. The processed abstracts were used as input, and the number of topics was set to five to allow comparison with the K -means results. LDA treats each document as a mixture of topics and identifies groups of terms that tend to appear together. The topic labels generated from LDA were compared with the K-means clusters to assess whether the major themes were stable across different analytical methods.

3.3. Thematic Findings

The clustering and topic modeling results show that AI research in higher education is organized around several recurring themes. A substantial portion of the literature focuses on personalized and adaptive learning. These studies examine intelligent tutoring systems, learner modeling, recommender systems, adaptive learning platforms, and personalized feedback. Their shared concern is how AI can adjust learning content, difficulty, pacing, and feedback according to individual student needs.

Another visible research direction concerns learning analytics and predictive modeling. Publications in this area use machine learning, educational data mining, and large-scale learning data to predict student performance, identify dropout risks, analyze engagement patterns, and support early intervention. This theme reflects the growing use of data-driven methods to provide more timely and targeted academic support.

A practical application-oriented cluster is related to AI-enhanced teaching tools and intelligent classroom environments. This group includes studies on automated grading, AI teaching assistants, smart classrooms, educational software, and digital systems for classroom interaction and content delivery. Compared with more analytical themes, this area is closely connected with the daily use of AI in teaching, assessment, and instructional management.

The results also indicate increasing attention to ethical, psychological, and social issues. Studies in this cluster discuss privacy, transparency, learner autonomy, trust, motivation, psychological engagement, and the risks associated with algorithmic decision-making in educational settings. The presence of this theme suggests that the field is moving beyond technical performance and beginning to examine the broader consequences of AI adoption for students, teachers, and institutions.

A further group of publications focuses on institutional policy and implementation. These studies address teacher readiness, administrative adoption, implementation barriers, scalability, educational reform, and governance strategies for AI-supported higher education. This theme shows that AI integration is not only a matter of classroom practice, but also involves institutional planning, policy support, and long-term capacity building.

Overall, the bibliometric results suggest that AI research in higher education has developed from early technical exploration into a broader educational research field. The identified themes connect AI with learning support, assessment, prediction, ethics, institutional strategy, and educational governance. This thematic structure also indicates that AI is increasingly being studied as part of a larger higher education ecosystem involving learners, instructors, digital tools, institutions, and policy frameworks. Table 1 below summarizes the top TF-IDF weighted keywords for each thematic cluster, offering a concise view of their core focus

Table 1. Representative Keywords of the Five Thematic Clusters Based on TF-IDF Weights

Cluster	**Representative Keywords (Top TF-IDF Terms)**
AI-Supported Personalized Learning	adaptive learning, personalization, intelligent tutoring, recommender system, learner modeling
Learning Analytics & Prediction	learning analytics, performance prediction, dropout detection, engagement analysis, machine learning
Intelligent Classrooms & Tools	smart classroom, automated grading, AI tools, educational software, classroom interaction
Psychological & Ethical Aspects	autonomy, motivation, privacy, ethics, psychological engagement, trust
Institutional Policy & Practice	implementation, policy, teacher training, scalability, educational reform, digital strategy

4. Experiments and Analysis

4.1. ChatGPT-Assisted Learning vs. No AI Assistance

To examine the effect of generative AI support on student learning, ChatGPT was introduced as a study aid in a course-based setting. Students in the experimental group were allowed to use ChatGPT for assignment support, concept clarification, and example-based explanation, while students in the control group completed the same learning tasks without AI assistance.

Figure 1 compares the overall learning performance of the two groups. The ChatGPT-assisted group achieved an average final score of approximately 88%, whereas the control group obtained about 81%. The difference suggests that access to ChatGPT may have helped students understand course content more effectively by providing timely explanations, worked examples, and individualized learning support. The improvement was particularly noticeable among lower-performing students, indicating that ChatGPT may serve as a useful supplementary scaffold for learners who need additional clarification. An independent-samples t-test further showed that the difference between the two groups was statistically significant.

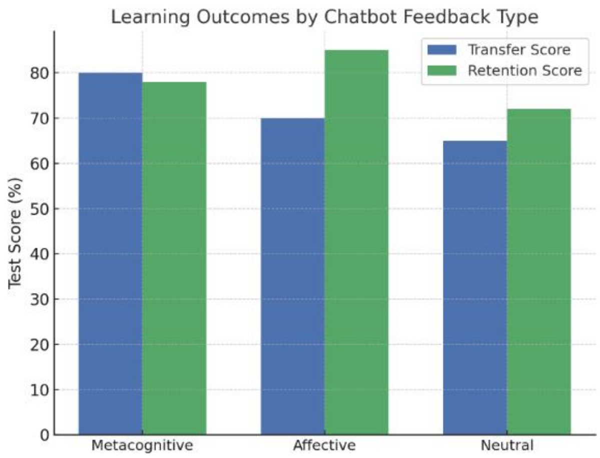

Figure 1. Example of learning performance with vs. without ChatGPT assistance.

Figure 2 presents a related post-test comparison between students who received ChatGPT-based tutoring and feedback and those who completed the same tasks independently. The ChatGPT-assisted group reached an average post-test score of about 80%, compared with roughly 70% for the control group. This result is consistent with recent meta-analytic evidence reporting a large positive effect of ChatGPT-assisted learning, with a mean effect size of approximately (g \approx 0.87) [11]. The observed performance gain may be attributed to ChatGPT's ability to offer immediate explanations, examples, and feedback during the learning process.

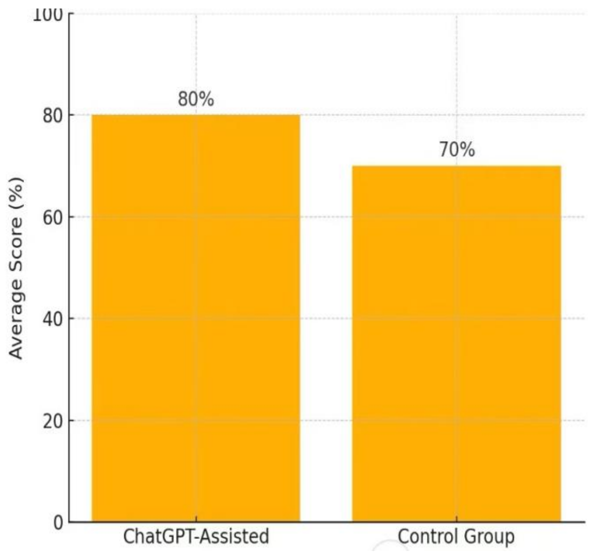

Figure 2. Post-Test Scores: ChatGPT-Assisted Group vs. Control Group.

However, the effect of ChatGPT should not be interpreted as universally positive in all contexts. Prior studies have also reported limited or mixed effects when students rely too heavily on AI-generated answers or use the tool without sufficient instructional guidance. Therefore, the present results suggest that ChatGPT can improve learning performance when it is used as a structured learning aid, but its educational value depends on how it is integrated into the course.

4.2. Feedback Styles in AI Chatbot Tutoring

The quality of AI-assisted learning depends not only on whether students receive feedback, but also on the type of feedback provided. To examine this issue, the dialogue records from the AI chatbot tutoring sessions were analyzed and coded according to the instructional function of each response. As shown in Figure 3, the chatbot's feedback was grouped into three categories: direct corrective feedback, elaborative feedback, and motivational feedback.

Direct corrective feedback accounted for approximately 50% of the total responses. This type of feedback mainly identified errors in students' answers and supplied the correct answer, explanation step, or solution procedure. Elaborative feedback represented about 30% of the responses. These responses went beyond correction by providing additional explanations, guiding questions, examples, or reasoning prompts to help students understand the underlying concepts. The remaining 20% of responses were motivational or encouraging in nature, including praise, reassurance, and supportive comments intended to maintain students' confidence and engagement.

This distribution suggests that the chatbot primarily functioned as a corrective and explanatory tutor, while still offering a degree of emotional and motivational support. Such a combination is pedagogically meaningful because effective feedback should not only point out mistakes, but also help learners revise their thinking and sustain their willingness to continue learning. The results therefore indicate that the AI tutor provided both cognitive support and affective support during the tutoring process.

Figure 3 presents the distribution of feedback styles generated by the AI chatbot during tutoring sessions. The responses were classified into three categories: direct corrective feedback, elaborative feedback, and motivational feedback. Percentages indicate the share of each category among all feedback instances.

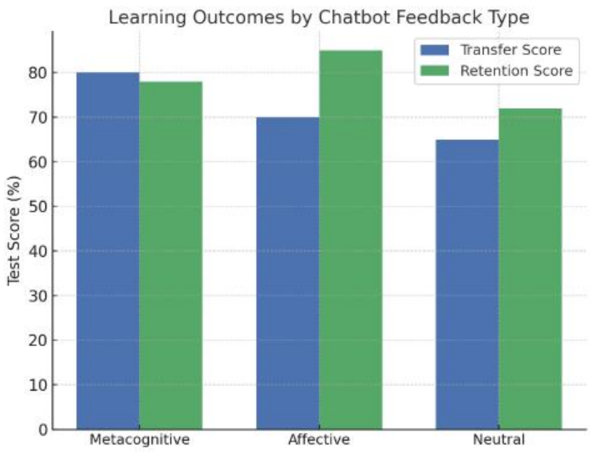

Figure 3. Feedback Types in AI Chatbot Tutoring

4.3. AI-Based Automated Grading and Human Assessment

To evaluate the reliability of AI-assisted assessment, grades generated by the AI system were compared with those assigned by human instructors for the same set of student assignments. The comparison focused on grading consistency, absolute grading error, and the agreement between AI-generated and human-assigned scores.

Figure 4 summarizes the distribution of absolute grading errors for the two grading approaches. Absolute error is defined as the absolute difference between an assigned score and the reference score. Compared with human grading, the AI system produced a lower median error and a narrower error distribution, indicating more consistent scoring across assignments.

The agreement between AI-generated scores and instructor grades was further examined using correlation analysis. As shown in Figure 4, the AI scores closely followed the human evaluations, yielding a Pearson correlation coefficient of ($r=0.92$). For most assignments, the difference between AI and human scores remained within ±5 percentage points, suggesting a high level of grading consistency. Larger discrepancies were observed mainly in assignments requiring subjective judgment, such as writing quality, creativity, or open-ended reasoning, where evaluation criteria are inherently less standardized.

Overall, the results indicate that AI-assisted grading can approximate instructor assessment with a high degree of consistency while reducing grading variability. Rather than replacing human judgment, AI grading may be more appropriately viewed as a decision-support tool that improves grading efficiency and consistency, particularly in large-enrollment courses where rapid and standardized assessment is required.

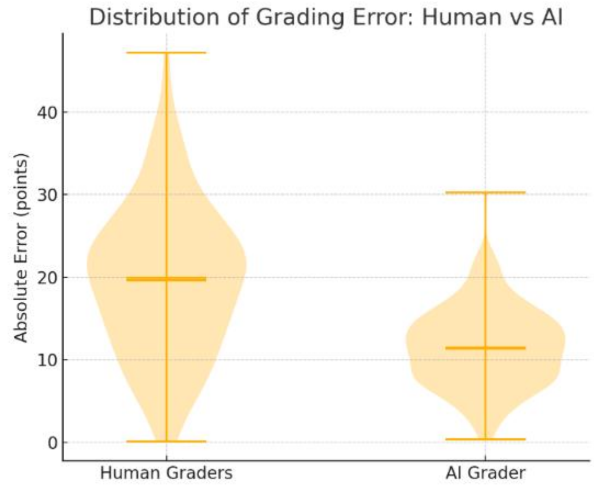

Figure 4. Comparison of Grading Errors Between Human and AI Assessment

5. Conclusion

This study examined the role of artificial intelligence in higher education by combining bibliometric analysis with experimental evidence. The bibliometric results show that research on AI in higher education has expanded rapidly, with growing attention to personalized learning, learning analytics, intelligent tutoring, automated assessment, and ethical issues. The experimental findings further suggest that ChatGPT-assisted learning can improve student performance, especially among lower-performing students, by providing timely explanations, examples, and feedback. The analysis of AI-generated feedback and grading also indicates that AI tools can support formative assessment and produce scores that are broadly consistent with human evaluation. Overall, the findings suggest that AI can be a useful supplement to university teaching and assessment when applied with clear pedagogical purposes and human oversight. Future research should further examine its long-term effects, ethical risks, and applicability across different educational contexts.

Acknowledgments
This research received no specific funding from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0
2. Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20*, Article 22. https://doi.org/10.1186/s41239-023-00392-8
3. Chu, H.-C., Hwang, G.-J., Tu, Y.-F., & Yang, K.-H. (2022). Roles and research trends of artificial intelligence in higher education: A systematic review of the top 50 most-cited articles. *Australasian Journal of Educational Technology, 38*(2), 22–42. https://doi.org/10.14742/ajet.7682
4. Dwivedi, Y. K., et al. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, *challenges and implications of generative conversational AI for research, practice*

and policy. International Journal of Information Management, 71, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

5. Xie, X., & Wang, T. (2024). Artificial intelligence: A help or threat to contemporary education? *Education and Information Technologies, 29*(3), 3097–3111. https://doi.org/10.1007/s10639-023-12262-7
6. Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: *A critical literature review. Higher Education, 86*(2), 369–385. https://doi.org/10.1007/s10734-022-00866-4
7. Chatterjee, S., & Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: *A quantitative analysis using structural equation modelling. Education and Information Technologies, 25*(4), 3443–3463. https://doi.org/10.1007/s10639-020-10152-1
8. Hinojo‐Lucena, F.‐J., Aznar‐Díaz, I., Cáceres‐Reche, M.‐P., & Romero‐Rodríguez, J.‐M. (2019). *Artificial intelligence in higher education: A bibliometric study on its impact in the scientific literature. Education Sciences, 9*(1), Article 51. https://doi.org/10.3390/educsci9010051
9. Schmohl, T., Schelling, K., Göke, S., Thaler, K. J., & Watanabe, A. (2022). Development, implementation and acceptance of an AI-based tutoring system: A research-led methodology. *In Proceedings of the 13th International Conference on Computer Supported Education (CSEDU)* (Vol. 2, pp. 179–186). SCITEPRESS. https://doi.org/10.5220/0011079500003172
10. Gaftandzhieva, S., Hussain, S., Hilcenko, S., Doneva, R., & Boykova, K. (2023). Data-driven decision making in higher education institutions: State-of-play. *International Journal of Advanced Computer Science and Applications, 14*(6), 1–12. https://doi.org/10.14569/IJACSA.2023.0140642
11. Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications, 12*, Article 621. https://doi.org/10.1057/s41599-025-02331-3
12. Bond, M., Marín, V. I., Zawacki-Richter, O., et al. (2024). A meta-systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*, Article 4. https://doi.org/10.1186/s41239-024-00421-x
13. Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: *A meta-analytic review. Review of Educational Research, 86*(1), 42–78. https://doi.org/10.3102/0034654315581420
14. VanLehn, K. (2011). *The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369
15. Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in‐class active learning: *An RCT introducing a novel research‐based design in an authentic educational setting. Scientific Reports, 15,* Article 17458. https://doi.org/10.1038/s41598-025-97652-6
16. Sakelaris, P. G., Novotny, K. V., Borvick, M. S., Lagasca, G. G., & Simanton, E. G. (2025). *Evaluating the use of artificial intelligence as a study tool for preclinical medical school exams.*

Journal of Medical Education and Curricular Development, 12, 1–12. https://doi.org/10.1177/2382120524124123

17. Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). *Application and theory gaps during the rise of artificial intelligence in education. Computers and Education: Artificial Intelligence, 1,* Article 100002. https://doi.org/10.1016/j.caeai.2020.100002
18. Laun, M., & Wolff, F. (2025). Chatbots in education: Hype or help? A meta-analysis. *Learning and Individual Differences, 119*, Article 102646. https://doi.org/10.1016/j.lindif.2024.102646
19. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: *Data Mining, Inference, and Prediction* (2nd ed., pp. 460–463). Springer-Verlag.