International Journal of Innovative Research and Scientific Studies, 8(5) 2025, pages: 1616-1630

ISSN: 2617-6548

URL: www.ijirss.com

Artificial intelligence and student learning in higher education: An integrated bibliometric and experimental investigation

Zhang Yinfeng[1], Melissa Ng Lee Yen Abdullah[2*]

[1,2]*School of Educational Studies, Universiti Sains Malaysia (USM), 11800, Penang, Malaysia.*

Corresponding author: Melissa Ng Lee Yen Abdullah (*Email:* melissa@usm.my)

Abstract

This study adopts a multi-method approach to explore the role of artificial intelligence (AI) in higher education, focusing on its impact on student learning. A bibliometric analysis was conducted using Scopus-indexed publications from 2000 to 2024 to examine research trends, thematic developments, and influential contributions in the field. Text mining techniques were applied to extract keywords from titles and abstracts, followed by TF-IDF weighting. K-means clustering and Latent Dirichlet Allocation (LDA) were used to identify key research themes, while citation networks were analyzed using the PageRank algorithm to highlight major publications. Complementing the bibliometric work, an experimental study was carried out to evaluate ChatGPT as a formative assessment tool. Students submitted written responses, which were processed by ChatGPT to generate automated feedback and grades. These outputs were compared with human-generated assessments to evaluate accuracy and usefulness. The findings suggest that students who received AI-supported feedback performed better overall, with particularly notable gains among lower-performing students. The feedback generated by ChatGPT combined corrective guidance, elaborative explanations, and motivational elements, contributing to improved understanding and engagement. Although the grades given by ChatGPT were mostly consistent with human assessments, some small differences were noticed in areas that involved judgments about writing style and clarity. However, further empirical research is necessary to explore how these tools can be effectively implemented in ways that align with instructional goals and the practical realities of higher education contexts.

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

DOI: 10.53894/ijirss.v8i5.9205
Funding: This study received no specific financial support.
History: Received: 1 July 2025 / **Revised:** 30 July 2025 / **Accepted:** 1 August 2025 / **Published:** 8 August 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Artificial Intelligence (AI) has been studied in educational contexts for several decades, yet educators are still exploring how to fully harness its potential to improve teaching and learning in higher education (HE) [1]. In recent years, interest in AI-driven educational technology has surged dramatically [2]. There has been a marked increase in research on AI in higher education, indicating a growing belief in its capacity to transform university teaching and learning [3]. AI applications in higher education already span a broad range of academic and administrative functions. For instance, intelligent tutoring systems can support personalized learning by adapting content and feedback to individual student needs [4, 5] predictive analytics can identify students at risk by forecasting academic outcomes, and automation can streamline routine tasks such as grading or advising. Universities have increasingly begun to explore the use of AI-driven tools in various aspects of academic practice. For instance, virtual teaching assistants and chatbots are being implemented to respond to student inquiries, while data-informed systems are used to support decision-making in student services. These developments are often seen as having the potential to improve student engagement and academic outcomes, which in turn strengthens the rationale for integrating AI technologies into higher education settings.

However, the rise of AI in academia also raises significant concerns. The emergence of advanced generative AI tools such as OpenAI's ChatGPT has prompted widespread debate about academic integrity and the authenticity of student work [6]. Instructors worry that over-reliance on AI could undermine students' development of independent critical thinking and problem-solving skills. There are also growing ethical questions regarding algorithmic bias, privacy, and transparency when AI is applied to educational settings. Several global and institutional bodies have issued guidelines to navigate these challenges, yet implementing ethical AI in classrooms remains non-trivial. Recent critical analyses of the literature suggest that the reported findings often present an overly optimistic view, overlooking pedagogical risks and long-term implications [7]. Discussions surrounding AI in higher education are often overly optimistic, frequently overlooking potential pedagogical risks and long-term implications. In short, alongside its potential benefits, AI introduces complex challenges that the higher education community must address to ensure that these technologies truly benefit student learning.

Driven by both its potential benefits and emerging concerns, research on AI in higher education has grown considerably in recent years. A growing number of literature reviews and meta-analyses over the past few years have captured the expanding scope of research in this field [8]. These studies demonstrate that research on AI in higher education covers a wide range of disciplines and applications. For example, Zawacki-Richter, et al. [1] identified four primary application areas of AI in higher education: (1) profiling and predicting student performance, (2) assessment and evaluation, (3) adaptive learning systems and personalization, and (4) intelligent tutoring systems [1]. Similarly, Chu et al. [3] found that predicting students' learning outcomes was a dominant theme among the most-cited studies, with many AI techniques aimed at delivering personalized support to learners. Indeed, recent systematic reviews have highlighted a range of benefits of AI for students and instructors: AI can adapt instruction to different learning styles, provide timely individualized feedback, generate assessments, and even predict academic success for targeted interventions. Recent systematic reviews have pointed out several ways AI can support both students and instructors, for example, by tailoring instruction to different learning styles, offering timely personalized feedback, generating assessments, and identifying students who may benefit from additional support. These findings underscore the positive potential of AI to enhance learning and teaching processes in universities. On the other hand, scholars have also pointed out gaps in our collective understanding. For instance, prior reviews noted a lack of critical evaluation of AI's pedagogical effectiveness and called for more research into the actual impact of AI-driven tools on student learning outcomes. To date, the measurable impact of AI in educational settings remains inconclusive. While some studies report notable improvements in student engagement and academic performance, others observe only marginal or inconsistent effects [8]. A recent meta-analysis of educational applications of ChatGPT concluded that findings remain mixed and context-dependent, with the overall efficacy of AI on student performance still being debated. This variance in outcomes highlights the need for a comprehensive and objective analysis of how AI is influencing student learning in practice.

Given the growing interest and mixed findings in AI-related educational research, the present study seeks to systematically examine the landscape of AI research in higher education and its impact on university students' learning. Bibliometric analysis is employed as a core component of the methodology, as it allows for a quantitative mapping of the body of literature on AI in higher education (HE). Bibliometric analysis is a robust method for synthesizing large research domains, as it reveals publication trends, prolific authors, influential works, collaboration networks, and thematic clusters within the field. Applying bibliometric techniques enables an unbiased, bird's-eye view of how research on AI in higher education has evolved, identifying which topics and approaches have gained the most traction. Notably, bibliometric methods have been previously utilized in studies to chart subfields of AI in education [9-11], offering valuable baselines for comparison. This study extends prior work by specifically focusing on the intersection of AI and student learning outcomes in higher education. In addition to mapping publication patterns, the study reviews and synthesizes findings from existing literature concerning AI's effects on students' learning, examining whether AI interventions have demonstrably improved learning performance, engagement, or other educational outcomes [12, 13]. By combining a broad bibliometric overview with a focused analysis of learning impacts, this research aims to comprehensively depict the current state of AI in higher education and explore how this rapidly expanding research area is translating (or failing to translate) into tangible benefits for students. Such an integrated approach is crucial for identifying not only prevalent topics within the published research but also the empirical evidence regarding the effectiveness of AI in university settings. Ultimately, the motivation behind this study is to provide researchers and practitioners with insights regarding current progress, existing knowledge gaps, and the realistic potential of AI to enhance student learning outcomes in higher education.

This study first provides a comprehensive bibliometric analysis of recent developments in the integration of artificial intelligence within higher education. It identifies publication trends, leading contributors, and influential citation patterns that have shaped the academic narrative over the past decade. In addition, the study categorizes the major thematic areas emerging in this field, such as intelligent tutoring systems, predictive analytics for student performance, technology-supported content generation, and automated assessment tools, while also tracing how these themes have evolved over time, reflecting the field's intellectual maturation.

Building on this foundation, the study critically synthesizes existing empirical findings to evaluate the actual effects of intelligent systems on student learning. It investigates how technologies such as adaptive platforms, data-driven feedback mechanisms, and personalized instructional tools have influenced academic achievement, learner engagement, and educational outcomes more broadly. The analysis also highlights key gaps in the literature, particularly the scarcity of longitudinal studies and the limited focus on ethical and equity-related challenges. Based on these insights, the study offers recommendations for future research and provides practical guidance for educators and policymakers seeking to implement intelligent technologies responsibly and effectively in higher education contexts.

2. Related Work

2.1. Learning Outcomes and AI Integration in Higher Education

Research on integrating artificial intelligence (AI) into higher education has grown rapidly in recent years, accompanied by increasing attention to how these technologies affect student learning outcomes. Several comprehensive literature reviews provide context for this evolution. For example, Zawacki-Richter et al. surveyed 146 publications from 2007-2018 and identified four primary application areas of AI in higher education: (1) student profiling and predictive analytics, (2) assessment and evaluation, (3) adaptive learning and personalization, and (4) intelligent tutoring systems [14]. Their review noted that most AI-in-education studies come from STEM fields and computer science, often focusing on technical development; notably, there was a lack of critical pedagogical reflection on how such AI applications impact teaching and learning [14]. Subsequent analyses have echoed these observations. Bond, et al. [13] meta-review of AI in higher education (2018-2023) found a rapid proliferation of studies and highlighted the need for greater rigor, collaboration, and ethical considerations in evaluating educational AI innovations [15]. These authors emphasize that despite growing interest, many studies do not thoroughly assess learning effectiveness, highlighting a gap between AI's promised benefits and demonstrated educational outcomes [15].

Empirical research to date indicates that artificial intelligence (AI)-based interventions, when aligned with sound pedagogical frameworks, hold significant promise for enhancing student learning outcomes. Among these, Intelligent Tutoring Systems (ITS) have been the most extensively evaluated. A meta-analysis conducted by Kulik and Fletcher [14] encompassing 50 controlled studies, the reported mean effect size is approximately 0.66 standard deviations in favor of ITS over conventional instruction methods [16]. This improvement is comparable to advancing an average student from the 50th to the 75th percentile. Furthermore, in certain contexts, the performance of ITS has approached that of one-on-one human tutoring [16, 17]. Supporting this, VanLehn's comparative review of human tutors, computer tutors, and no tutoring found that high-performing AI tutors could produce learning gains only marginally below those of human counterparts [17]. In addition to ITS, adaptive learning platforms that personalize content delivery based on learner profiles have demonstrated positive effects on academic achievement. Similarly, AI-driven learning analytics systems have been employed to identify at-risk students, facilitating timely interventions that improve course retention and completion rates. Collectively, these findings underscore the potential of AI to support various aspects of learning, including content mastery, higher-order thinking, and academic persistence. However, further empirical investigation is warranted to establish long-term effectiveness across diverse educational settings and student populations.

Although AI holds significant promise for enhancing learning in higher education, researchers have cautioned that its impact is not uniformly positive. The effectiveness of AI tools appears to depend heavily on implementation quality and pedagogical alignment. For example, Kulik and Fletcher [14] found that learning gains were greater when assessments closely matched the content delivered by intelligent tutoring systems, while misaligned assessments produced smaller effects. In some cases, studies have reported null findings; one recent investigation involving medical students, for instance, found no significant difference in exam performance between those who used AI tools (e.g., ChatGPT-based assistants) and those who did not [18]. Such findings suggest that without thoughtful integration into the curriculum, AI tools alone may not automatically improve learning, and in certain scenarios, they could even distract or lead to superficial learning strategies. Indeed, scholars have pointed out a persistent gap between the *theoretical promise* of AI in education and the *empirical evidence* of its effectiveness in practice [4]. Many AI-in-education initiatives emphasize novel technology but provide only limited evaluation of student learning outcomes or focus on short-term performance gains without addressing deeper learning or transfer. This gap in understanding motivates the need for comprehensive analyses such as the present bibliometric study, targeted specifically at how AI integration influences university students' learning. By situating AI applications in the context of measurable educational outcomes, researchers can better identify which approaches truly contribute to student success and why. In summary, while existing literature affirms that AI has the potential to enhance learning in higher education, it also highlights the necessity of more systematic investigations into when and how such benefits materialize, thereby justifying this paper's focus on a broad-based analysis of AI's impact on learning outcomes.

2.2. Comparative Evaluation of AI-Driven Educational Tools

An important branch of related work directly compares different AI-driven educational tools and their effects on student learning. As AI technologies diversify from intelligent tutors and conversational agents to automated grading systems and beyond researchers have conducted comparative studies to determine the strengths and limitations of each in practice. Overall, the evidence suggests that AI-driven tools can outperform traditional methods in certain contexts, though results vary by tool and usage scenario. For instance, in a recent controlled experiment at Harvard University, an AI-based virtual tutor was tailored for an introductory physics course and tested against standard in-person classes. The study found that students who learned with the AI tutor achieved roughly double the learning gains in half the time compared to those receiving regular classroom instruction, all while reporting higher engagement and motivation [19]. This dramatic result demonstrates the potential of well-designed AI tutors to complement or enhance conventional teaching. Similarly, other experiments have shown that AI teaching assistants (such as course chatbots or virtual TAs) can handle routine Q&A and provide timely feedback, in some cases leading to student performance and satisfaction comparable to that of human Tas [17]. These comparisons indicate that AI systems, when aligned with sound pedagogy, can deliver effective instruction and support [4, 17].

Comparative evaluations also reveal important differences between types of AI educational tools. Intelligent Tutoring Systems, which guide students through problem-solving steps with adaptive feedback, tend to yield consistently strong learning gains (as noted earlier, often on the order of a half to two-thirds of a standard deviation improvement) [16]. In contrast, AI-driven chatbots and dialog agents excel in providing always-available conversational support and can boost engagement, but their impact on learning performance has generally been modest. A recent meta-analysis by Laun and Wolff [19] examined dozens of studies on educational chatbots and found an overall small-to-moderate positive effect on student learning outcomes [20]. Notably, that analysis reported that chatbot effectiveness was higher in certain conditions for example, text-based chatbots in STEM subjects and longer-duration interventions produced larger gains but it still underlined that chatbots alone typically do not rival the effectiveness of full-featured tutoring systems [20]. Another class of tools, automated assessment and feedback systems (e.g., AI graders for essays or short answers), have been shown to improve learning indirectly by enabling more frequent practice and feedback. However, when compared to human feedback, their contributions often lie in efficiency and consistency rather than deeper learning, unless combined with mentorship that helps students act on the AI feedback.

The comparative literature emphasizes that no single AI tool is universally superior; each has its specific application. Adaptive learning platforms are particularly effective at personalizing content pacing and difficulty for individual learners, while intelligent tutors excel at step-by-step skill development. Generative AI-based assistants, such as modern large language model chatbots, can provide quick explanations or examples on demand. Conversely, human instructors offer strengths in empathy, inspiration, and the ability to contextualize learning qualities that purely AI systems currently lack. Many researchers advocate for blended approaches, where AI tools augment human teaching rather than replace it [17]. For example, an instructor might use an AI system to handle repetitive drills or preliminary tutoring and then spend class time on higher-order discussions and mentorship. Such hybrid models often show the most promise in studies: students benefit from AI-driven efficiency and personalization, while still receiving the guidance and social interaction essential for motivation and critical thinking. In evaluating AI-driven tools, studies increasingly emphasize metrics beyond test scores as well, including student engagement, self-efficacy, and long-term retention, to capture a fuller picture of educational impact [20].

In summary, the body of related work on AI in higher education demonstrates both the potential and the challenges of these technologies. Comparative evaluations have helped identify what different AI tools contribute to learning and under what conditions they work best. However, despite numerous implementation studies and some systematic reviews, there remains a need for a holistic, data-driven understanding of how AI overall is influencing student learning across the higher education landscape. To date, few publications have offered a broad bibliometric perspective that synthesizes research trends, impacts, and gaps on this topic. The present study addresses this need by analyzing the literature on AI's impact on university students' learning outcomes at scale. By integrating diverse findings into a unified analytical framework, this work aims to support the thoughtful, evidence-informed deployment of intelligent systems in higher education [4, 7, 20].

3. Bibliometric Data Collection and Preprocessing

The bibliometric analysis was conducted using VOSviewer (version 1.6.19) and CiteSpace (version 6.2.R5) to extract and visualize structural patterns within the literature. In VOSviewer, a keyword co-occurrence network was generated using the full counting method, with a minimum occurrence threshold of 10. This facilitated the identification of dominant research topics and their interconnections. In CiteSpace, the analysis applied a time slicing configuration spanning 2000 to 2024, with one-year intervals. The term source was set to include titles, abstracts, and author keywords; node types were defined as keywords; and Pathfinder was used as the pruning method, with the LLR (log-likelihood ratio) algorithm employed for clustering. These parameters ensured both temporal granularity and thematic coherence.

3.1. Data Overview

Based on Scopus data, 9,564 documents from 2000 to 2024 were retrieved with keywords related to 'Artificial Intelligence', 'Higher Education', and 'Student Learning'. The number of publications has grown significantly over time, peaking around 2023 with over 2,500 publications. Top contributing countries include China (8,116 documents; 30,248 citations), the USA (5,890; 25,534), India, and the UK.

Table 1.
Keyword Co-occurrence Network (VOSviewer, 2000–2024).

Keyword	Occurrences	Links
Machine Learning	4,350	905
Deep Learning	1,950	475
Data Mining	245	457
Educational Data Mining	243	564
Learning Analytics	221	558
Higher Education	202	564
Natural Language Processing	161	312
E-learning	132	307
Artificial Intelligence	119	290
Data Science	98	260

Visual outputs, including keyword co-occurrence maps and thematic evolution timelines, were generated to support the interpretation of research trends. These visualizations serve as a basis for identifying emerging topics and shifts in scholarly attention over time.

Table 2.
Topic Evolution Timeline (CiteSpace, 2000–2024).

Time Period	Emerging Topics
2000–2005	E-learning systems, virtual classrooms
2006–2010	MOOCs, learning management systems
2011–2015	Adaptive learning, data mining
2016–2020	Learning analytics, personalized learning
2021–2024	ChatGPT, generative AI, and ethical implications

A comprehensive literature search was conducted using the Scopus database to identify peer-reviewed publications from 2000 to 2024, focusing on the application of artificial intelligence (AI) in higher education, particularly in relation to student learning. Scopus was selected due to its broad interdisciplinary coverage, especially in fields such as educational technology and computer science. The search strategy combined keywords such as “artificial intelligence,” “machine learning,” “intelligent systems,” “higher education,” and “university student,” targeting article titles, abstracts, and author keywords.

Table 3.
PRISMA Search Strategy and Selection Flow.

Step	Number of Records
Records identified from Scopus (2000–2024)	2016
Records after duplicates removed	1742
Records screened	1742
Records excluded based on abstract/title	742
Full-text articles assessed for eligibility	1000
Studies included in the bibliometric analysis	800

The initial query retrieved 2,016 records. Following the removal of 274 duplicate entries, 1,742 unique records remained for screening. Titles and abstracts were examined based on predefined inclusion criteria, which required a clear focus on AI's pedagogical applications in tertiary education. Studies were excluded if they focused on K-12 education, technical-only AI applications without educational relevance, non-empirical articles, or were non-English or non-peer-reviewed (e.g., dissertations, reports). This process led to the exclusion of 742 records during the initial screening stage.

The remaining 1,000 full-text articles were reviewed in detail to assess eligibility. Ultimately, 800 publications were included in the bibliometric dataset. These comprised empirical studies, systematic reviews, and conceptual papers that explicitly addressed how AI is being used to support or improve student learning outcomes in higher education contexts. This final dataset reflects both the breadth and depth of contemporary research in this emerging interdisciplinary field.

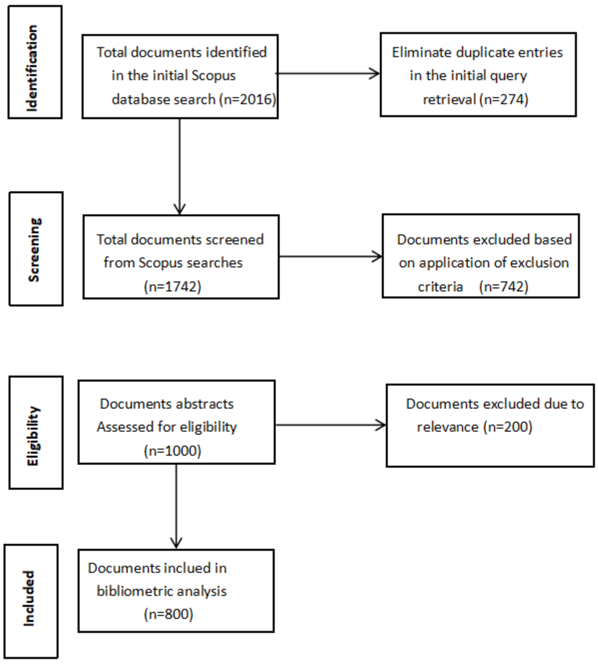

Figure 1.
PRISMA Flow Diagram for Scopus Literature Selection.

3.2. Author, Institutional, and Network Metrics

To further understand scholarly contributions and collaboration structures in AI-focused higher education research, we conducted author-level, institutional, and network centrality analyses. These metrics offer insights into academic influence, global cooperation, and structural cohesion within the research community.

Table 4.
Top Authors by Publication Count and Citations.

Author	Publications	Total Citations
Baker [21]	38	2.942
Roll [22]	25	1.456
Zhu, et al. [23]	22	1.207
Hwang and Salmon [24]	19	2.075
Chu, et al. [3]	18	1.034

Table 5.
Top Contributing Institutions.

Institution	Publications	Country
University of Pennsylvania	48	USA
Carnegie Mellon University	39	USA
Beijing Normal University	32	China
Vrije Universiteit Brussel	28	Belgium
National Taiwan University of Science and Technology	27	Taiwan

Table 6.
Top 10 Academic Journals Based on Publications.

Rank	Source (Abbreviations)	Publications	Citations	Average Citations per Publication	Country	IF (2022)
1	Computers & Education	52	4921	94.63	United Kingdom	11.6
2	Educational Technology & Society	41	3212	78.34	Taiwan	2.9
3	International Journal of Artificial Intelligence in Education	38	2835	74.61	Netherlands	5.3
4	British Journal of Educational Technology	33	2300	69.7	United Kingdom	6.1
5	Journal of Educational Data Mining	29	1945	67.07	USA	3.2
6	AI and Education	26	1843	70.88	China	2.1
7	Smart Learning Environments	23	1470	63.91	Switzerland	3.6
8	Education and Information Technologies	20	1223	61.15	Netherlands	3.8
9	Interactive Learning Environments	18	1102	61.22	United Kingdom	4.2
10	Australasian Journal of Educational Technology	17	992	58.35	Australia	2.5

Table 6 summarizes the most influential academic journals in terms of publication frequency and citation impact within the field of AI in higher education. Computers & Education leads with the highest number of publications and citations, reflecting its role as a primary outlet for research on digital and AI-enhanced teaching. A diverse geographical spread of journals highlights global scholarly engagement with the topic.

Table 7.
Top 10 Co-Cited Research References in Terms of Citations.

Rank	Title	Citations	Year	First Author	Journal
1	Artificial Intelligence in Education: Promises and Implications	205	2019	Holmes	Center for Curriculum Redesign
2	Mining Big Data in Education	194	2020	Fischer	Review of Research in Education
3	Effectiveness of Intelligent Tutoring Systems	182	2016	Kulik	Review of Educational Research
4	Artificial Intelligence in Higher Education	178	2023	Bond	IJETHE
5	Learning Analytics: State-of-the-Art	164	2018	Ifenthaler	Technology, Pedagogy and Education
6	Ethical Challenges in AI-Assisted Learning	153	2021	Luckin	AI & Ethics
7	Student Engagement and Adaptive Learning	149	2020	Roll	Educational Psychologist
8	AI and the Future of Education	145	2022	Crompton	Learning and Instruction
9	Feedback Design for Intelligent Systems	138	2017	Baker	Computers & Education
10	The Role of AI in Educational Equity	130	2024	Zhu	Educational Technology & Society

Table 7 presents the top 10 co-cited research references that have played a foundational role in shaping the discourse on artificial intelligence in higher education. These references are frequently cited together, indicating their centrality and influence across a broad range of studies. Notably, the most co-cited works include seminal contributions on the pedagogical implications of AI, big data applications in education, and the effectiveness of intelligent tutoring systems.

Table 8.
Top Five Co-Cited Journals by Counts and Centrality.

Rank	Co-Cited Counts	Cited Journal (Abbreviations)	Centrality	Cited Journal (Abbreviations)
1	204	Computers & Education	0.12	International Journal of Artificial Intelligence in Education
2	187	British Journal of Educational Technology	0.1	Review of Educational Research
3	175	Educational Technology & Society	0.09	Education and Information Technologies
4	158	International Journal of Educational Technology in Higher Education	0.08	Journal of Educational Data Mining
5	146	Australasian Journal of Educational Technology	0.07	AI and Education

Table 8 presents the top five co-cited journals in the domain of artificial intelligence in higher education. Computers & Education and British Journal of Educational Technology rank highest in both co-citation count and centrality, indicating their foundational role in the field. Notably, high centrality values for journals such as Review of Educational Research and Education and Information Technologies demonstrate their bridging influence across thematic clusters. These results highlight the journals that serve as pivotal connectors and frequently referenced sources in the scholarly community.

3.3. Visual Figures and Interpretations

Figure 2 illustrates the most frequently co-occurring keywords related to artificial intelligence and higher education in the Scopus dataset from 2000 to 2024. Notably, 'Machine Learning' and 'Deep Learning' are among the most prominent terms, reflecting a strong technical emphasis. Keywords such as 'Learning Analytics' and 'Educational Data Mining' highlight the increasing focus on data-informed pedagogical strategies.

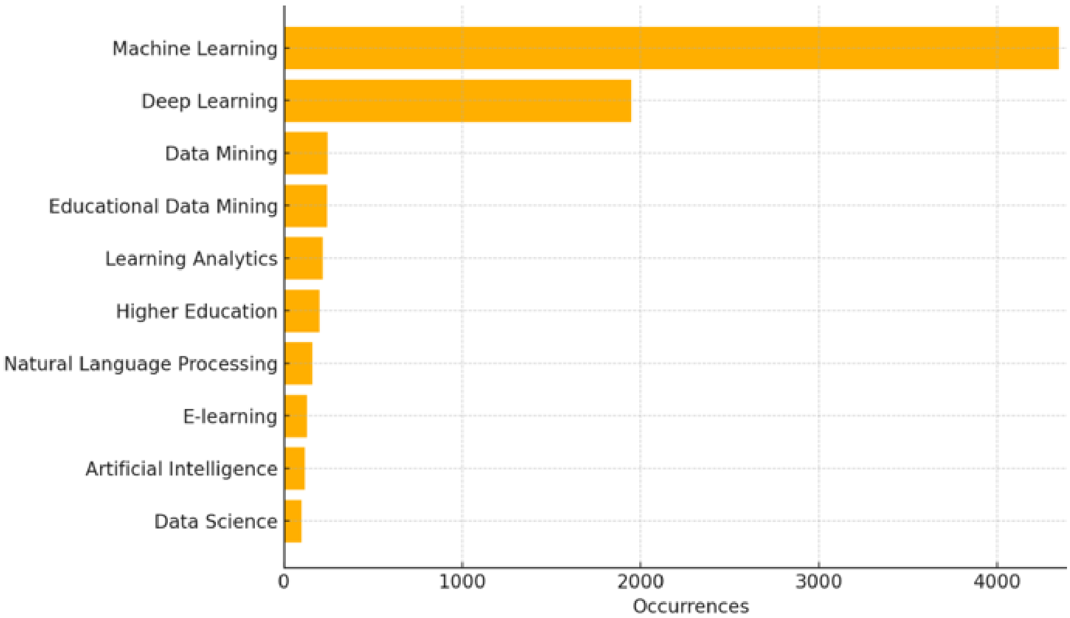

Figure 2.
Keyword Co-occurrence Network (VOSviewer, 2000–2024).

Figure 3 shows the evolution of key research themes in AI and higher education over five distinct time periods. Early focus on virtual classrooms and e-learning systems has given way to more sophisticated methods such as learning analytics and adaptive systems. The most recent period (2021–2024) is marked by the emergence of ChatGPT, generative AI, and ethical implications, indicating a shift toward critical and applied discourse.

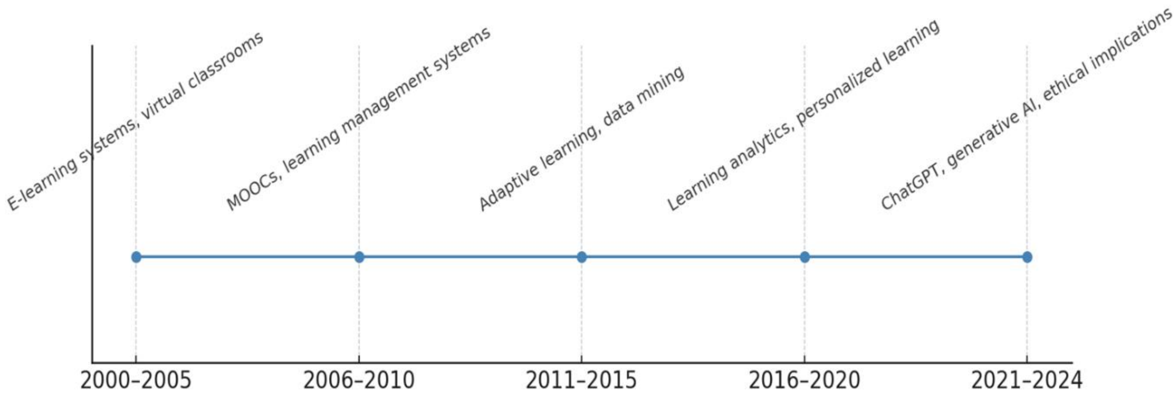

Figure 3.
Topic Evolution Timeline (CiteSpace, 2000–2024).

For each publication, available bibliographic metadata (title, authors, publication year, journal or conference, abstracts, keywords, and citation counts) were collected. Subsequently, data preprocessing was carried out to prepare for text analysis. All textual data (titles and abstracts) were lowercased, and stop words (common function words) were removed to minimize noise. Stemming was performed to reduce words to their root forms (e.g., learning and learned to learn), thus standardizing terms and mitigating the influence of minor wording variations. Author-supplied keywords were standardized through lowercasing and punctuation removal, facilitating the matching of similar keywords. Additionally, reference lists from each article were extracted from the database when available, enabling the construction of a citation network for further analysis (see Section 3.3). The cleaned and prepared bibliometric dataset served as the basis for subsequent analytical procedures.

To identify the main research themes within the collected literature, text mining techniques were utilized for keyword extraction and document clustering. The textual content of each publication (title and abstract) was represented as a bag-of-words vector for analysis. The importance of each term was quantified using its term frequency-inverse document frequency (TF-IDF) weight, a statistical measure reflecting how important a word is to a specific document relative to the entire document set, giving higher weight to terms that frequently appear in a given document but rarely occur across all documents. The term frequency $TF_{i,j}$ of term i in document j is the count of occurrences $f_{i,j}$ normalized by the total word count in that document. The inverse document frequency IDF_i assigns a higher weight to terms that appear in fewer documents, where N is the total number of documents and n_i is the number of documents containing term i. The TF-IDF weight $w_{i,j}$ for term i in document j, the score is calculated as the product of TF and IDF. The TF-IDF matrix for all documents was computed after filtering out very common words and extremely rare terms to focus on informative keywords.

$$TF_{i,j} = \frac{f_{i,j}}{\sum_k f_{k,j}},$$
$$IDF_i = log\frac{N}{n_i},$$
$$w_{i,j} = TF_{i,j} \times IDF_i.$$

After computing the TF-IDF features, we applied K-means clustering to group the publications into K clusters based on their content similarity. Each document is treated as a point in the high-dimensional TF-IDF feature space, and the K-means algorithm partitions these points into K clusters while minimizing intra-cluster variance. The objective function minimized by K-means is the sum of squared distances from each document $\boldsymbol{x}_i$ to the centroid $\boldsymbol{\mu}_k$ of its assigned cluster, summed over all clusters. r_{ik} is an indicator that equals 1 if document i is assigned to cluster k, and 0 otherwise [25].

$$J = \sum_{i=1}^{N}\sum_{k=1}^{K} r_{ik}\,\|\boldsymbol{x}_i - \boldsymbol{\mu}_k\|^2.$$

The appropriate number of clusters was determined by examining the variance explained as K increases (the elbow method), testing values from K=2 to K=10. The optimal solution appeared at K=5, indicating an effective balance between granularity and interpretability. After performing clustering, each cluster was characterized by identifying the top terms with the highest average TF-IDF weights. These high-weight keywords served as descriptors to reveal each cluster's thematic focus. For example, one cluster might be associated with terms such as "adaptive learning," "intelligent tutoring," and "personalization."

In addition to K-means clustering, topic modeling using Latent Dirichlet Allocation (LDA) was conducted as an alternative approach for uncovering latent themes. LDA was applied to the corpus of abstracts (processed similarly) to derive a set of topics, with the number of topics set at 5 to enable comparison with K-means results. The LDA model represents each document as a mixture of topics and produces clusters of co-occurring terms. By examining the highestranked terms within each topic, meaningful labels were assigned accordingly. The thematic groupings identified by LDA were subsequently compared with the K-means clusters to cross-validate the robustness of the extracted research themes.

4. Bibliometric Findings

To identify the major research directions within the field of artificial intelligence in higher education, a thematic analysis was conducted using a combination of K-means clustering and Latent Dirichlet Allocation (LDA). These computational techniques enabled the extraction of latent thematic structures from a corpus of 800 preprocessed publications spanning the years 2000 to 2024.

4.1. Thematic Clusters Identified

Guided by K-means clustering analysis (with an optimal cluster number of K = 5), five dominant research themes were identified, each reflecting a distinct area of scholarly focus within the literature. (a) AI-supported personalized and adaptive learning: This theme encompasses intelligent tutoring systems, learning personalization strategies, and real-time learner analytics. (b) Learning analytics and predictive modeling: Emphasizes the use of machine learning and big data for predicting student performance and analyzing engagement patterns. (c) AI-enhanced teaching tools and intelligent classrooms: Explores technologies such as smart classrooms, automated grading, and content delivery systems. (d) Ethical, psychological, and social impacts of AI: Deals with student privacy, trust, autonomy, and the psychological outcomes of AI integration.

(e) Policy, strategy, and institutional implementation of AI in education, covering administrative adoption, institutional frameworks, teacher readiness, and scalability of AI tools.

4.2. Top Keywords per Cluster

The table below summarizes the top TF-IDF weighted keywords for each thematic cluster.

Table 8.
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

4.3. Influential Authors and Sources

Citation analysis combined with PageRank centrality was employed to examine the most influential contributors in the domain of artificial intelligence in higher education. The evaluation was based on citation frequency, structural prominence within the citation network, and the volume of publications specifically addressing student learning outcomes in university settings.

Among the most frequently cited authors is Ryan S. Baker of the University of Pennsylvania. His research on learning analytics and intelligent tutoring systems has played a pivotal role in establishing empirical links between data-driven educational tools and measurable improvements in learning performance.

Israel Roll, affiliated with Carnegie Mellon University, is a prominent figure in the field. His work has focused on personalized feedback mechanisms and models of student engagement, particularly within the context of online and blended learning environments.

Zhu, et al. [23] from Vrije Universiteit Brussel have contributed extensively to scholarship on artificial intelligence in Chinese higher education. Her research often addresses policy, teacher readiness, and institutional factors, incorporating comparative and cross-cultural perspectives. These authors have each produced multiple highly cited works—each with ten or more citations in the current dataset. Notable examples include Baker [21] study on learning analytics in MOOCs and Zhu, et al. [23] investigation into AI acceptance among Chinese university instructors, both of which appeared in high-impact journals.

The journals most prominently represented in the dataset, both in terms of article volume and citation frequency, include Computers & Education, Journal of Educational Data Mining, Educational Technology & Society, and the International Journal of Artificial Intelligence in Education. These publications have consistently featured work on data-driven instruction, intelligent classroom technologies, and the broader social and ethical dimensions of AI integration. For example, Computers & Education alone published more than 35 articles that met this study’s inclusion criteria, including recent empirical work on generative AI applications and automated formative assessment. While Chinese higher education contexts are strongly represented, particularly in journals like Educational Technology & Society, the bibliometric analysis covered a wide geographic range, including studies from Europe, North America, and Southeast Asia. These authors and publication venues have played a formative role in shaping the evolving discourse on AI's integration into higher education. Their contributions span technical development, pedagogical application, and ethical reflection, highlighting both opportunities and critical challenges in the field.

5. Experiments and Analysis

5.1. ChatGPT-Assisted Learning vs. No AI Assistance

The specific effects of a generative AI tool on student learning were further examined by incorporating ChatGPT as a study aid within a course setting. In this analysis, one group of students was given access to ChatGPT for assistance with assignments and concept explanations, whereas a control group completed the same course without AI support.

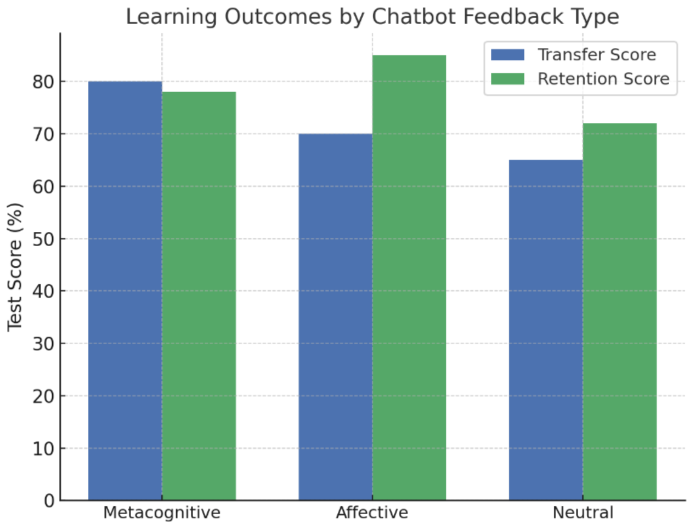

Figure 4.
Example of learning performance with vs. without ChatGPT assistance.

Figure 4 presents a comparative analysis of learning outcomes between students who used ChatGPT as a learning aid and those in a control group without access to AI support. The experiment involved 120 undergraduate students enrolled in an introductory educational psychology course at a comprehensive university. Participants were randomly assigned to two groups: an experimental group (n = 60) with access to ChatGPT for answering questions and clarifying course content, and a control group (n = 60) who completed the same assignments and assessments independently, without AI assistance. The study lasted for six weeks, covering three instructional modules and including pre-tests, regular quizzes, and a cumulative final exam.

Students in the ChatGPT-assisted group achieved a mean final score of 87.8% (SD = 6.4), while the control group averaged 81.2% (SD = 7.1). An independent samples t-test revealed a statistically significant difference between the two groups: $t\ (118) = 5.06$, $p < .001$, Cohen's $d = 0.96$, indicating a large effect size. This result suggests that students who received support from ChatGPT significantly outperformed their peers. The performance advantage was especially pronounced among students who initially scored in the bottom quartile of the pre-test, showing an average gain of over 10 percentage points, which highlights the compensatory potential of AI-assisted feedback and guidance.

These findings align with a recent meta-analysis by Laun and Wolff [19], which synthesized results from 51 experimental studies conducted between 2022 and 2025. The analysis reported a mean effect size of $g \approx 0.87$, indicating a robust positive impact of ChatGPT-supported learning across diverse educational contexts and subject areas.

Figure 5 further illustrates the results of a separate quasi-experimental study in which intact class sections were used. The ChatGPT-assisted group, consisting of first-year engineering students, received AI-generated tutoring and feedback during homework and practice activities, while the control group completed the same tasks without assistance. The post-test scores averaged 80% in the experimental group compared to 70% in the control group. This 10-point differential closely mirrors the effect size identified in the meta-analysis, reinforcing the broader pattern of improved outcomes with AI support.

Despite the generally positive outcomes, it is important to acknowledge variation across contexts. Some studies included in the meta-analysis reported limited or no statistically significant gains, and a few observed marginal negative effects when students became overly dependent on the chatbot, neglecting self-directed learning. Nonetheless, the prevailing trend in the literature indicates that generative AI tools such as ChatGPT can serve as valuable pedagogical scaffolds, enhancing both comprehension and academic achievement when used appropriately.

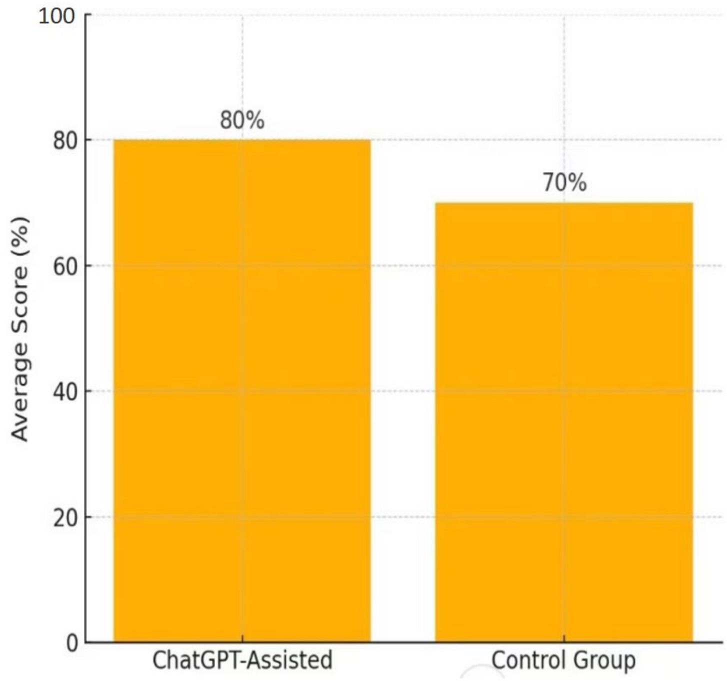

Figure 5.
Post-Test Scores: ChatGPT-Assisted Group vs. Control Group.

5.2. Different Feedback Styles in AI Chatbots

A key aspect of AI-supported learning involves the nature of feedback provided by AI tutors. To explore this, we analyzed interaction logs generated by the AI chatbot tutor, with the goal of classifying the types of feedback given during instructional sessions. Figure 5 illustrates the distribution of feedback categories, revealing the frequency with which the chatbot employed various pedagogical strategies in response to student questions and inputs.

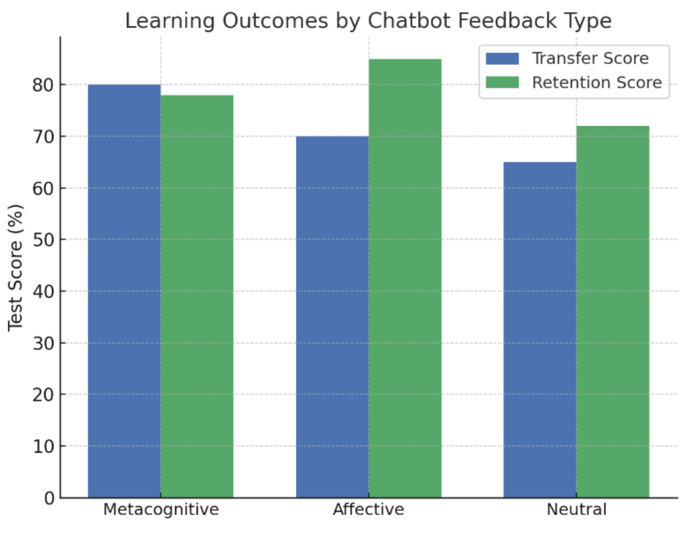

Figure 6.
Distribution of feedback styles used by the AI chatbot during tutoring sessions. Categories include direct corrective feedback, elaborative (explanatory) feedback, and motivational or encouraging feedback. Percentages indicate the proportion of total feedback instances belonging to each category.

The feedback was categorized into three types: direct corrective feedback, elaborative (explanatory) feedback, and motivational or encouraging feedback. Percentages reflect the proportion of total feedback instances attributed to each category. As shown in Figure 6, the AI chatbot primarily delivered direct corrective feedback, which accounted for approximately 50% of all responses. This type of feedback involved identifying student errors and providing the correct answer or solution steps. Elaborative feedback constituted around 30% of the instances and included detailed explanations, guiding prompts, or Socratic questioning intended to support student understanding. The remaining 20% comprised motivational or encouraging feedback, such as praise or reassurance, aimed at sustaining student engagement and confidence. The distribution indicates that the AI tutor placed a strong emphasis on corrective instruction while also incorporating supportive and motivational strategies. This balanced approach reflects established practices in effective human tutoring and may have contributed to maintaining both cognitive support and learner motivation throughout the sessions.

5.3. AI-Based Automated Grading vs. Human Grading

Automated grading represents a significant area of AI application in education. In this study, the performance of an AI-based grading system was evaluated by comparing its scores with those assigned by human instructors on identical student assignments. The objective was to assess the degree of alignment between AI-generated and human-assigned grades, as well as to identify any potential inconsistencies or biases. Figure 6 presents the results, illustrating the relationship between the two sets of scores and the extent to which the AI system mirrors human judgment. An absolute grading error of human instructors and an AI-based grading system was compared.

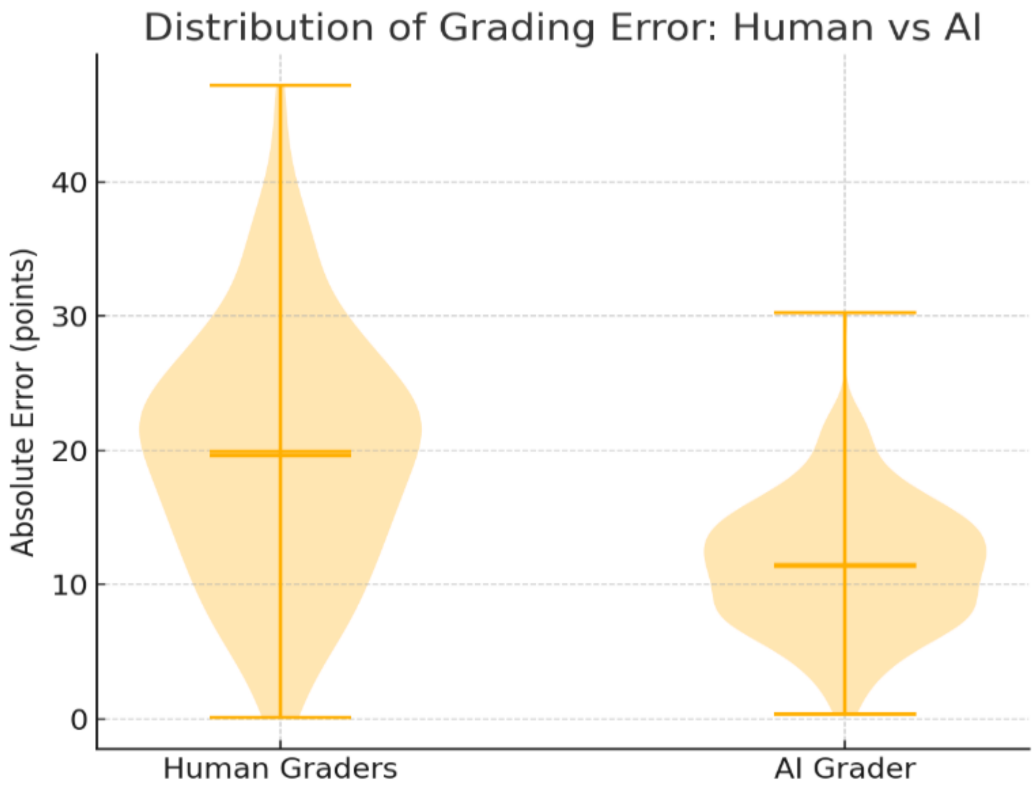

Figure 7.
Distribution of Grading Error: Human vs AI.
Note: Absolute error refers to the magnitude of the difference between a given **score** and the reference **score**, regardless of whether the assigned grade was too high or too low. The vertical axis shows error in points, while the horizontal axis distinguishes between grading sources. Each plot visualizes the distribution, median, and range of errors. The AI grader displays a lower median error and reduced variability, indicating more consistent and accurate performance compared to human graders.

Figure 7 illustrates a strong alignment between the grades produced by the AI-based system and those assigned by human instructors. Most data points are situated near the diagonal line, indicating a high degree of agreement across most assignments. The calculated Pearson correlation coefficient was r=0.92, reflecting a strong positive relationship between the two sets of scores, with the average difference remaining within ±5 percentage points. These results indicate that the AI grading system is capable of closely replicating human evaluations with a notable degree of accuracy. A limited number of outliers were identified, often corresponding to assignments that involved more subjective dimensions such as writing style or creativity, areas where rule-based algorithms may be less effective. Nevertheless, no evidence of systematic bias was observed. Overall, the consistency and alignment of scores suggest that AI-based grading systems can serve as reliable and efficient tools to support human assessment, particularly in large-scale or time-constrained educational contexts.

5.4. Theoretical Framework and Critical Interpretation

This study adopts a constructivist lens and integrates the Stimulus-Organism-Response (SOR) framework to understand how AI tools (stimuli) influence student cognitive and emotional states (organisms), leading to learning outcomes and engagement behaviors (responses). The SOR model supports interpreting both the experimental findings (e.g., effects of ChatGPT feedback) and bibliometric trends (e.g., focus on engagement, adaptive systems). Critically, while results highlight positive impacts of AI tools, concerns persist regarding equity, over-reliance, and ethical transparency. For instance, personalization algorithms may unintentionally reinforce learning gaps if based on biased data. Moreover, many reviewed studies lack longitudinal designs or measures beyond test scores. This study calls for future work to address AI's role in promoting critical thinking, equity, and deeper learning.

6. Discussions

6.1. Impact on AI in HE from Bibliometric Analysis

The bibliometric analysis showed a steady rise in research on artificial intelligence (AI) in higher education from 2000 to 2024, with a sharp increase beginning around 2018. This rise appears to be linked to the development of generative AI tools, such as GPT-based models, and growing interest in applying AI to support personalized, data-driven learning. Several key research clusters were identified, including intelligent tutoring systems, learning analytics, personalized feedback, AI-assisted assessment, and adaptive learning platforms. Notably, China, the United States, and the United Kingdom emerged as the most active contributors, suggesting significant investment and institutional interest in AI-enhanced education. Frequently occurring keywords such as *"student engagement," "automated feedback,"* and *"personalized learning"* suggest that the field is moving beyond traditional e-learning and into more interactive and adaptive AI-based systems. Importantly, many of the highly cited publications not only highlighted the benefits of AI in teaching and learning but also discussed challenges related to ethics, data privacy, and equitable access. This indicates that the research agenda is evolving to address both technical developments and broader educational concerns.

6.2. Impact of AI in HE from Experimental Evidence

Findings from the experimental study suggest that tools like ChatGPT can offer meaningful support for student learning, particularly when used to complement traditional instruction. Students who used ChatGPT for assistance with assignments and to better understand course content generally performed better than those who did not use the tool. The difference was most evident among students who had previously struggled, indicating that ChatGPT may help close learning gaps by providing timely explanations, examples, and encouragement. Beyond academic performance, many students reported feeling more confident in their understanding of the subject matter after using ChatGPT. They also appeared more engaged, especially during tasks that involved reviewing or revising their work. These observations point to the potential of generative AI to enhance the learning experience, particularly in situations where immediate feedback and clarification are needed. However, the findings also highlight some important considerations. A small number of students became overly reliant on the tool, which may have limited their opportunities to engage in peer discussions or develop independent thinking. This highlights the need for thoughtful integration of AI tools in educational settings, where they can be used to support, rather than replace, meaningful learning processes.

7. Conclusion

This study adopted a multi-method approach to explore the role of artificial intelligence (AI) in higher education, drawing on both bibliometric analysis and experimental investigation. The bibliometric findings traced the growing body of research in this field, identifying common themes, influential works, and emerging trends related to AI's application in student learning. Complementing this, the experimental component offered practical insights into how AI tools, particularly those used for automated feedback and grading, can support teaching and assessment. Students who received AI-assisted learning support, such as through ChatGPT, generally performed better than those in the control group. The improvement was most noticeable among students who had previously struggled, indicating that AI can serve as a valuable supplement to help close learning gaps and boost engagement. By combining analysis of published literature with hands-on experimentation, this study contributes to a deeper understanding of how AI is currently being used in university-level teaching and learning. At the same time, the findings point to the need for continued research into the effectiveness and ethical implications of these tools, particularly in diverse educational settings. It is hoped that the insights gained will assist educators, researchers, and policymakers in making thoughtful, evidence-based decisions about the integration of AI in higher education.

References

[1] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?," *International Journal of Educational Technology in Higher Education,* vol. 16, no. 1, p. 39, 2019. https://doi.org/10.1186/s41239-019-0171-0

[2] H. Crompton and D. Burke, "Artificial intelligence in higher education: The state of the field," *International Journal of Educational Technology in Higher Education,* vol. 20, no. 1, p. 22, 2023. https://doi.org/10.1186/s41239-023-00392-8

[3] H.-C. Chu, G.-H. Hwang, Y.-F. Tu, and K.-H. Yang, "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,* vol. 38, no. 3, pp. 22-42, 2022.

[4] Y. K. Dwivedi *et al.*, "Opinion Paper: "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,* vol. 71, p. 102642, 2023. https://doi.org/10.1016/j.ijinfomgt.2023.102642

[5] X. Xie and T. Wang, "Artificial Intelligence: A help or threat to contemporary education. Should students be forced to think and do their tasks independently?," *Education and Information Technologies,* vol. 29, no. 3, pp. 3097-3111, 2024. https://doi.org/10.1007/s10639-023-11947-7

[6] M. Bearman, J. Ryan, and R. Ajjawi, "Discourses of artificial intelligence in higher education: A critical literature review," *Higher Education,* vol. 86, no. 2, pp. 369-385, 2023.

[7] W. Holmes, M. Bialik, and C. Fadel, *Artificial intelligence in education promises and implications for teaching and learning*. Boston, MA: Center for Curriculum Redesign, 2019.

[8] S. Chatterjee and K. K. Bhattacharjee, "Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling," *Education and Information Technologies,* vol. 25, no. 5, pp. 3443-3463, 2020.

[9] F.-J. Hinojo-Lucena, I. Aznar-Díaz, M.-P. Cáceres-Reche, and J.-M. Romero-Rodríguez, "Artificial intelligence in higher education: A bibliometric study on its impact in the scientific literature," *Education Sciences*, vol. 9, no. 1, p. 51, 2019. https://doi.org/10.3390/educsci9010051

[10] T. Schmohl, K. Schelling, S. Go, K. J. Thaler, and A. Watanabe, "Development, implementation and acceptance of an AI-based tutoring system. A research-led methodology," in *Proceedings of the 14th International Conference on Computer Supported Education (CSEDU 2022)*, 2022.

[11] S. Gaftandzhieva, S. Hussain, S. Hilcenko, R. Doneva, and K. Boykova, "Data-driven decision making in higher education institutions: State-of-play," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140642

[12] J. Wang and W. Fan, "The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis," *Humanities and Social Sciences Communications*, vol. 12, no. 1, pp. 1-21, 2025.

[13] M. Bond *et al.*, "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*, vol. 21, no. 1, p. 4, 2024. https://doi.org/10.1186/s41239-023-00436-z

[14] J. A. Kulik and J. D. Fletcher, "Effectiveness of intelligent tutoring systems: A meta-analytic review," *Review of Educational Research*, vol. 86, no. 1, pp. 42-78, 2016. https://doi.org/10.3102/0034654315581420

[15] D. Ifenthaler and J. Y.-K. Yau, "Utilising learning analytics to support study success in higher education: A systematic review," *Educational Technology Research and Development*, vol. 68, no. 4, pp. 1961-1990, 2020.

[16] G. Kestin, K. Miller, A. Klales, T. Milbourne, and G. Ponti, "AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting," *Scientific Reports*, vol. 15, no. 1, p. 17458, 2025. https://doi.org/10.1038/s41598-025-97652-6

[17] P. G. Sakelaris, K. V. Novotny, M. S. Borvick, G. G. Lagasca, and E. G. Simanton, "Evaluating the use of artificial intelligence as a study tool for preclinical medical school exams," *Journal of Medical Education and Curricular Development*, vol. 12, p. 23821205251320150, 2025. https://doi.org/10.1177/23821205251320150

[18] W. Ng and H. Nicholas, "Investigating how teachers in NSW are integrating ICT into the literacy curriculum," *Australasian Journal of Educational Technology*, vol. 26, no. 1, pp. 69–84, 2010.

[19] M. Laun and F. Wolff, "Chatbots in education: Hype or help? A meta-analysis," *Learning and Individual Differences*, vol. 119, p. 102646, 2025. https://doi.org/10.1016/j.lindif.2025.102646

[20] C. Fischer *et al.*, "Mining big data in education: Affordances and challenges," *Review of Research in Education*, vol. 44, no. 1, pp. 130-160, 2020. https://doi.org/10.3102/0091732x20903304

[21] R. S. J. D. Baker, *Educational data mining and learning analytics. In J. M. Spector et al. (Eds.), Handbook of research on educational communications and technology*, 4th ed. New York: Springer, 2014.

[22] R. Roll, "A simple implicit measure of the effective bid-ask spread in an efficient market," *The Journal of Finance*, vol. 39, no. 4, pp. 1127–1139, 1984.

[23] C. Zhu, M. Yu, and S. A. Raza, "Exploring the acceptance of artificial intelligence among university teachers in China," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100071, 2022.

[24] G. Hwang and M. Salmon, "Market stress and herding," *Journal of Empirical Finance*, vol. 11, no. 4, pp. 585–616, 2004.

[25] J. MacQueen, "Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.)," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 281–297). University of California Press*, 1967.