IJSER
ISSN: 2149-5939

International Journal of Social Sciences and Education Research
Online, https://dergipark.org.tr/tr/pub/ijsser
Volume: 11(1), 2025

Research trends and the impact of ChatGPT on educational environments

Thoriqi Firdaus[1], Rizqoh Mufidah[2], Rika Nur Hamida[3], R'Maya Inkya Febrianti[4] and Alvira Eka Rahel Guivara[5]

[1]*Corresponding author, Department of Natural Science Education, Universitas Negeri Yogyakarta, Indonesia, E-mail: thoriqifirdaus.2023@student.uny.ac.id, ORCID: https://orcid.org/0009-0005-2340-8468*
[2]*Department of Natural Science Education, Universitas Trunojoyo Madura, Indonesia, ORCID: https://orcid.org/0009-0006-5746-3120*
[3]*Department of Primary Education, Universitas Trunojoyo Madura, Indonesia, ORCID: https://orcid.org/0009-0003-8027-1336*
[4]*Department of Natural Science Education, Universitas Trunojoyo Madura, Indonesia, ORCID: https://orcid.org/0009-0007-7728-2674*
[5]*Department of Natural Science Education, Universitas Trunojoyo Madura, Indonesia, ORCID: https://orcid.org/0009-0008-6990-598X*

Article Info

Research Article

Received: 30 December 2024
Revised: 11 March 2025
Accepted: 12 March 2025

Keywords:
ChatGPT,
Education,
Learning,
Trends,
Impact

Abstract

This study aims to explore research trends and patterns and analyze ChatGPT's impact on education. The methodology employs a mixed-method approach, incorporating bibliometric analysis and a systematic literature review. Research data were sourced from the Scopus database using the keywords "ChatGPT" AND "Education" OR "Learning." The findings indicate that the trend of document publications in the Scopus database related to ChatGPT has seen a notable increase since its introduction in 2022, continuing through 2024. The journal JMIR Medical Education has emerged as the foremost source of citations, making significant contributions. The United States leads the way in article contributions (22.6%), followed by China (9.6%). Countries such as the United Kingdom, Canada, and Italy display high levels of international collaboration, likely enhancing the diversification and quality of research.

1. Introduction

The advancement of technology in the era of Society 5.0 has transformed the educational paradigm, ushering in a shift toward digitalization in the learning process (Firdaus, 2023). The advancement of the digital era has positioned artificial intelligence (AI) technology as a central component in the education sector (Sadiku et al., 2022). One of the latest developments in AI is the introduction of natural language models such as ChatGPT in educational settings (Kamalov et al., 2023). The deployment of ChatGPT in educational environments yields significant impacts (Castillo et al., 2023). ChatGPT introduces a new paradigm for teachers and students through AI's capability to answer various questions and respond in a human-like manner (Dalalah & Dalalah, 2023). The emergence of ChatGPT has revolutionized education in an era characterized by rapid and boundless technological usage (Vargas-Murillo et al., 2023). ChatGPT is unavoidable in demonstrating education's role in addressing the era's advancements.

The effects of ChatGPT on learning compel educators and students to continuously explore its strengths and weaknesses (Gill et al., 2024; Firdaus et al., 2024). The technological era has broadly impacted the use of ChatGPT, making learning more diverse by fostering greater student engagement and understanding (Einarsson & Lund, 2023). ChatGPT's ability to respond to various questions, akin to human intelligence, has led to significant changes in learning (Javaid et al., 2023). Concerns about detecting fake texts generated by ChatGPT highlight teachers' need for proper management as facilitators (Dalalah & Dalalah, 2023). Supervising ChatGPT in addressing various scientific inquiries can reveal potential weaknesses in the teaching and learning process (Karatas et al., 2024). Academics must adapt teaching practices and assessments to embrace the new reality of readily available AI (Sullivan et al., 2023). ChatGPT has the potential to streamline access to information and assist teachers in curriculum

**This study has complied with the Research Publication Ethics stated in "Wager E & Kleinert S (2011) Responsible research publication: international standards for authors. A position statement was developed at the 2nd World Conference on Research Integrity, Singapore, July 22-24, 2010. Chapter 50 in Mayer T & Steneck N (eds) Promoting Research Integrity in a Global Environment. Imperial College Press / World Scientific Publishing, Singapore". For this reason, the author states that he conducted the research within the framework of ethical principles. It is not a human study, so ethical approval is not required. All responsibility belongs to the author.*

To cite this article: Firdaus, T., Mufidah, R., Hamida, R.N., Febrianti, R.I. & Guivara, A.E.R. (2025). Research trends and the impact of ChatGPT on educational environments. *International Journal of Social Sciences and Education Research*, 11 (1), 45-62. DOI: https://doi.org/10.24289/ijsser.1609741

planning, yet integrating this technology raises ethical dilemmas, including privacy, confidentiality, and bias (Srinivasan et al., 2024). Additional concerns regarding ChatGPT relate to the reduced interaction between students and teachers (Limna et al., 2023). Therefore, the deployment of ChatGPT by educators and students requires astuteness in its application, ensuring its ethical, reliable, and efficient use.

ChatGPT has become a popular choice for exploring study subjects in various academic publications. Research related to ChatGPT delves into the development of algorithms that mimic human responses (AI Lily et al., 2023), educators' perceptions (ElSayary, 2024), learners' views (Shoufan, 2023), deficiencies and constraints (Tyson, 2023), as well as the effects and prospects of ChatGPT for students (Gill et al., 2024). Anticipating digital literacy in education seeks to address disruptions caused by ChatGPT and its negative impacts on the educational landscape. However, the historical significance of using ChatGPT is increasing, as the chatbot can research, integrate, interpret, and compile content in a human-like manner, presenting substantial transformations and obstacles for the education domain (Tirado-Olivares et al., 2023).

The presence of ChatGPT has stimulated academic discussions and research on its impact on educational environments. Bibliometric analysis and systematic literature review approaches are crucial for comprehensively understanding the effect. Bibliometric analysis is a method to explore and analyze scientific data to reveal the evolution of a field and highlight emerging areas within it (Donthu et al., 2021). In contrast, a systematic literature review (SLR) aims to identify all studies addressing a research question and its methodology, developed to minimize selection, publication, and data extraction biases (Nightingale, 2009).

Bibliometric analysis enables researchers to identify research trends, collaboration networks, and the evolving knowledge map related to using ChatGPT in education. Conversely, SLR can provide a detailed overview of existing research findings, identify benefits and challenges encountered, and offer insights into the ethical and pedagogical implications of using this technology. The combination of these two methods not only provides a comprehensive picture of ChatGPT's impact on education but also helps identify existing research gaps and potential future developments. Thus, this study aims to explore previous research trends and patterns and analyze the effects of ChatGPT on education.

2. ChatGPT

ChatGPT stands for Generative Pre-Trained Transformer. It was first introduced to the public by OpenAI in November 2022 and originated in the United States. OpenAI has developed a language model called ChatGPT to generate human-like text (OpenAI, 2022). ChatGPT is a processing model composed of deep learning and reinforcement algorithms trained on over 150 billion human-created items (Downling & Lucey, 2023). ChatGPT has garnered widespread popularity across various demographics. Statistical analysis indicates that ChatGPT users have increased significantly since its launch, with nearly 2 billion monthly visits by April 2024 (Duarte, 2024). The latest version of the ChatGPT application is GPT-4o, launched on May 13, 2024, offering enhanced intelligence and significantly faster text, voice, and vision processing capabilities. GPT-4o surpasses existing models in comprehending and discussing images that users share (Open AI, 2024). The AI capabilities within ChatGPT enable it to learn continually, improve, and even self-develop without human programming assistance (George et al., 2023). As long as users continue to pose questions to ChatGPT, its ability to answer various questions, both simple and complex, will improve, eventually matching the precision of human-generated responses (Kabir et al., 2024). Most ChatGPT users are aged 25 to 34, with the top target audience categories including programming and developer software, computer electronics and technology, video game consoles and accessories, education, and graphics (Similarweb, 2024).

3. Method

This study has complied with the Research Publication Ethics stated in "Wager E & Kleinert S (2011) Responsible research publication: international standards for authors. A position statement was developed at the 2nd World Conference on Research Integrity, Singapore, July 22-24, 2010. Chapter 50 in Mayer T & Steneck N (eds) Promoting Research Integrity in a Global Environment. Imperial College Press / World Scientific Publishing, Singapore". For this reason, the author states that he conducted the research within the framework of ethical principles. It is not a human study, so ethical approval is not required.

3.1. Data source and search strategies

The research data for this study comprises documents from the Scopus database that discuss ChatGPT in educational settings. The selection of documents does not impose any publication start date constraints, as ChatGPT

only emerged in 2022. The choice of the Scopus database is based on the platform's extensive geographical and thematic coverage (Dindorf et al., 2023). The sample search utilizes article titles, abstracts, and keywords with the terms “ChatGPT" AND "Education” OR “Learning” or employs the query (TITLE-ABS-KEY (ChatGPT) AND TITLE-ABS-KEY (education) OR TITLE-ABS-KEY (learning)).

3.2. Research design and data analysis

This study employs a mixed-method approach, integrating quantitative and qualitative research methodologies. The quantitative aspect of the research utilizes bibliometric analysis, while the qualitative component employs a systematic literature review.

3.2.1. Bibliometric analysis

Bibliometric analysis is a statistical method characterized by its precision in exploring and analyzing large volumes of scientific data to reveal variations and highlight developments within a specific field (Donthu et al., 2021). This study utilizes bibliometric analysis to identify emerging research trends related to using ChatGPT in education and learning. Based on research publications indexed in Scopus as of July 7, 2024, there are 4,402 research documents published in Scopus. However, for this bibliometric analysis, the study is limited to document types classified as articles and sources limited to journals, resulting in 2,024 documents. This study employs R Studio software with the Biblioshiny package for bibliometric analysis. Thus, documents are exported in BibTeX format. Before the analysis, 20 articles were excluded due to non-compliance with bibliometrics metadata requirements, leaving 2,004 articles for analysis.

Table 1. Completeness of bibliographic metadata

MD	Description	MD	MS %	Status
AU	Author	0	00.00	Excellent
DT	Document Type	0	00.00	Excellent
SO	Journal	0	00.00	Excellent
LA	Language	0	00.00	Excellent
PY	Publication Year	0	00.00	Excellent
TI	Title	0	00.00	Excellent
TC	Total Citation	0	00.00	Excellent
C1	Affiliation	19	0,065972222	Good
DI	DOI	22	01.10	Good
AB	Abstract	75	0,176388889	Good
DE	Keywords	219	0,480555556	Acceptable
RP	Corresponding Author	240	0,525694444	Acceptable
ID	Keywords Plus	1067	53.22.00	Critical
CR	Cited References	2005	100.00.00	Completely missing
WC	Science Categories	2005	100.00.00	Completely missing

MD: Metadata; MS: Missing Count

3.2.2. Systematic literature review

The systematic literature review identifies eligible studies by following the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA). PRISMA is an evidence-based set of policies that enhances the transparency and quality of systematic reviews (Figure 1). These guidelines assist researchers in providing clear, detailed explanations, ensuring that readers can fully understand the methods used and the results obtained (Page et al., 2021).

4. Results

4.1. Bibliometric mapping of extant studies

4.1.1. Descriptive analysis: evolution of publications and the most globally cited articles

The publication rate of articles discussing ChatGPT began in 2022 and has since seen a significant increase in user engagement and rapid development through 2023 and 2024. Numerous authors have contributed articles examining the influence and impact of ChatGPT on the educational domain. The frequency of research outcomes between 2022 and 2024 is depicted in Figure 2.

Table 2, which presents the annual scientific production, demonstrates the increase in articles published yearly. The rise in the number of articles discussing ChatGPT has been recorded over the past three years, namely 2022,

2023, and 2024. In 2022, the initial year of ChatGPT's emergence, two research publications were listed in the Scopus database. In 2023, there was a notable increase, with 813 articles published. This trend continued with a significant surge in 2024, reaching 1,189 published articles. In 2024, the widespread use of ChatGPT was evident, particularly among students, university attendees, educators, and academic staff. Research on ChatGPT in 2024 gained remarkable popularity. The number of publications is closely related to the average number of citations per year. The average citation per year is detailed in Table 2.

Figure 1. PRISMA Model

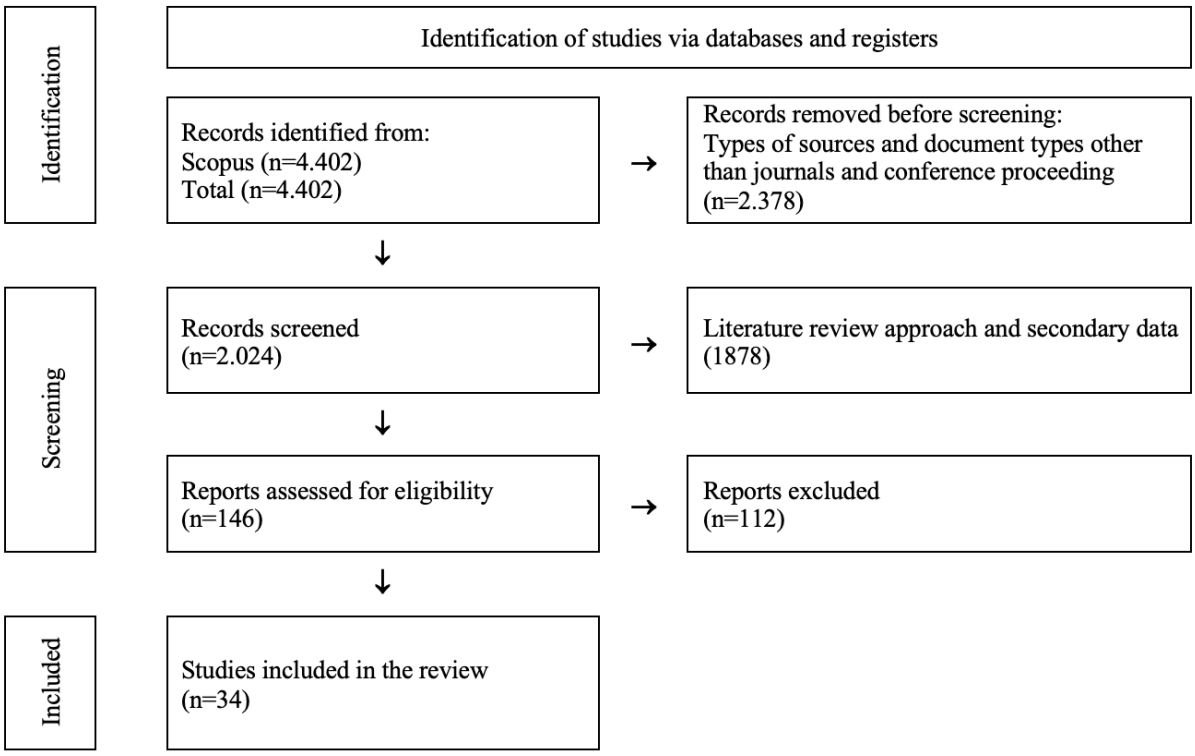


Figure 2. Annual scientific production

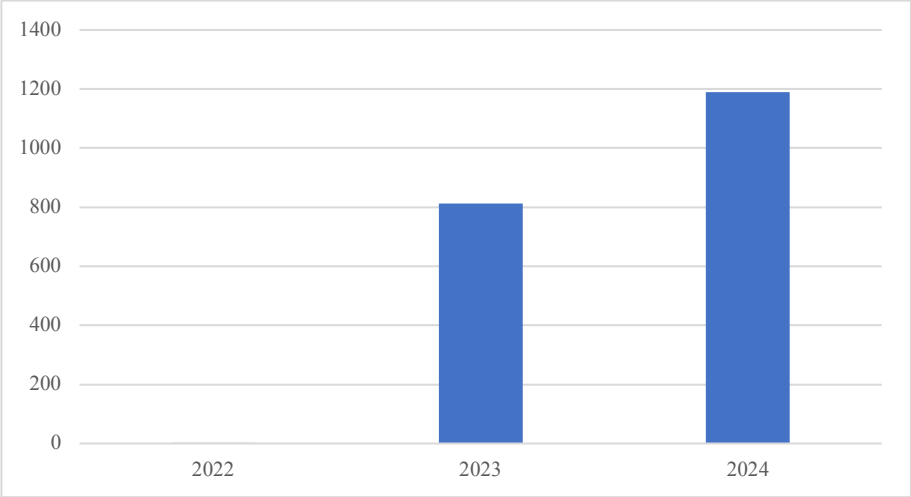


Table 2. Average citation per year

Year	MeanTCperArt	N	MeanTCperYear
2022	117,50	2	39,17
2023	20,12	813	10,06
2024	1,95	1189	1,95

Table 2, "Average Citation per Year," illustrates the average total citations over the three years considered. The average number of citations per article in 2022 is notably high, at 117.50. The small sample size influences this elevated figure, as only two articles were published that year. It suggests that each article was highly impactful. The average citations per year for 2022 is 39.17, indicating that despite the small sample size, the articles were very influential. In 2023, the average number of citations per article decreased significantly to 20.12. This decline corresponds with a substantial increase in the number of published articles on ChatGPT in the Scopus database, which rose to 813. The broader and more diverse coverage in 2023 likely contributed to the lower average citations per article. The average citations per year also decreased to 10.06, reflecting the increased number of articles diluting the citation impact of highly cited works. By 2024, the average citations per article further declined to 1.95, despite the number of articles increasing to 1,189, even though data collection occurred mid-year. This trend indicates a continued rise in publications, leading to broader coverage and reducing the average number of citations per article. The average number of citations per year in 2024 stands at 1.95, showing a sharp decline in citation impact compared to previous years.

4.1.2. Citation source analysis and local impact

Scientific articles indexed in Scopus that investigate and understand the impact of ChatGPT in educational settings total 2,006 documents published in various international journals. The overall number of papers collected has been published across different journals. One measure of a scientific article's success is the number of citations it receives. The higher the citation count, the greater the benefit and significance of the article. Consequently, the potential readership dramatically influences the impact and credibility of a journal. The most relevant sources for this research are presented in Figure 3, and the local impact sources are detailed in Table 3.

Based on Figure 3, the journal with the highest number of scientific articles on the analysis of ChatGPT is JMIR Medical Education, with 46 articles. Education and Information Technologies follows it with 34 articles, and Computers and Education: Artificial Intelligence with 29 articles. The Journal of Applied Learning and Teaching and the Journal of Chemical Education have an equal number of publications, with 23 articles each. IEEE Access occupies the fifth through eighth positions with 22 articles, Education Sciences with 19 articles, Scientific Reports and TechTrends with 18 articles, and Journal of Medical Internet Research with 15 articles. The most relevant sources are closely related to the local impact sources presented in Table 3.

Figure 3. Most relevant sources

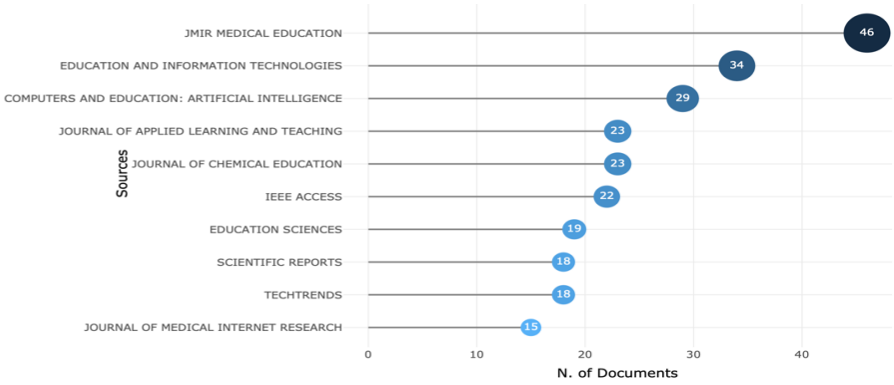


Table 3. Source local impact

Source	h_index	g_index	m_index	TC	NP
JMIR Medical Education	12	33	6	1112	46
Journal of Applied Learning and Teaching	12	23	6	1145	23
Journal of Chemical Education	10	18	5	334	23
Computers and Education: Artificial Intelligence	8	13	4	203	29
Education and Information Technologies	8	18	4	329	34
Education Sciences	6	11	3	139	19
International Journal of Management Education	6	9	3	340	9
Scientific Reports	6	9	3	88	18
European Archives of Oto-Rhino-Laryngology	5	11	2,5	124	13
International Journal of Educational Technology in Higher Education	5	10	2,5	287	10

According to Table 3, JMIR Medical Education is identified as the journal with the highest number of articles and possesses robust metrics. It indicates that many authors have cited articles from this journal as references. JMIR Medical Education discusses language processing models or programming codes by ChatGPT that are revolutionizing medical education. Education and Information Technologies exhibit equally robust metrics as Computers and Education: Artificial Intelligence, though their research foci differ. Education and Information Technologies concentrates on ChatGPT's potential as a learning catalyst, while Computers and Education: Artificial Intelligence explores advanced tool accessibility based on artificial intelligence (AI). The Journal of Applied Learning and Teaching and the Journal of Chemical Education have equal publications and robust metrics. The Journal of Applied Learning and Teaching significantly impacts discussions on the limitations and challenges of AI, as well as the potential of ChatGPT for enhancing student learning. In contrast, the Journal of Chemical Education primarily addresses the performance and potential of ChatGPT to effect substantial changes in chemistry education.

Education Sciences is relevant to integrating ChatGPT into teaching strategies, whereas Scientific Reports focuses on the implications of ChatGPT for scientific communication. The International Journal of Management Education primarily addresses strategies for integrating ChatGPT into education from a constructivist learning perspective. Although these three journals possess similarly strong metrics, their research focuses differ while remaining within the same overarching context. The European Archives of Oto-Rhino-Laryngology is dedicated to identifying responses generated by ChatGPT across various domains. The International Journal of Educational Technology in Higher Education concentrates on evaluating the proficiency and reliability of ChatGPT. Despite having relatively lower metrics, these journals maintain their credibility and value in the academic discourse.

4.1.3. Collaboration analysis: countries and authors

Figure 4, "Most Relevant Authors," illustrates data on authors with the most relevant articles published. Zhang Y is the leading author, having published 15 journals. It is followed by three authors—Wang C, Wang J, and Wang Y—each with 14 articles. Chen J's count decreases to 12 articles, and further down, Li Z, Seth I, Wang X, and Wang Z each have 11 articles. The lowest count is observed for Chen Y, who has published nine articles. The data in Figure 4 shows a non-uniform, zigzag pattern of author contributions, with the colors indicating decreasing publication numbers. As the publication numbers decrease, the colors shift from dark blue to light blue, reflecting the diminishing quantity of published work.

Table 4 provides data on the local impact of authors, including several metrics: h-index, g-index, m-index, total citations (TC), and number of publications (NP). The h-index reflects the number of significant publications an author has. The g-index offers a more detailed view of frequently cited articles, emphasizing quality over quantity. The m-index accounts for the duration of a researcher's productivity, providing insight into how quickly an author establishes their scientific impact.

Figure 4. Most relevant authors

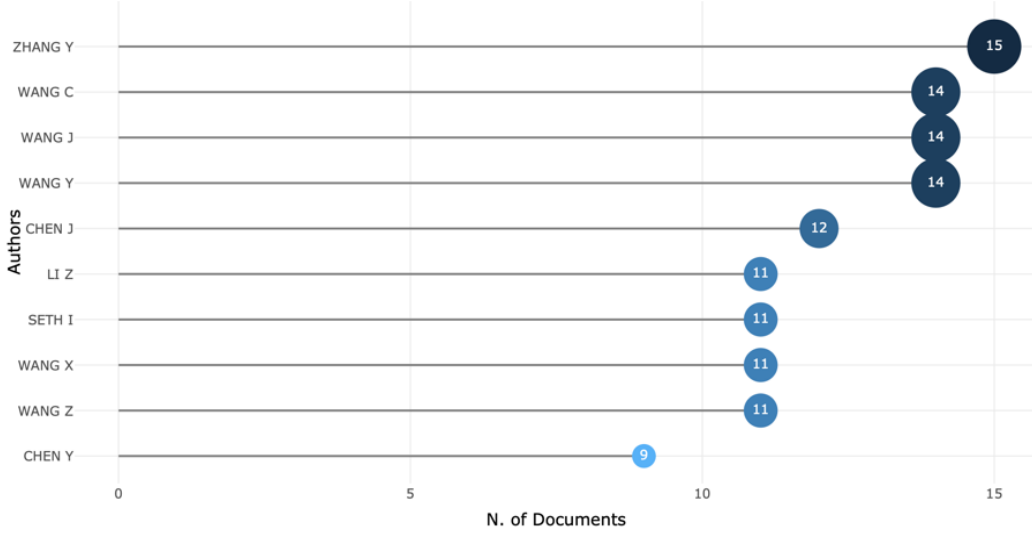


Table 4. Authors local impact

Author	h_index	g_index	m_index	TC	NP
Tan S	7	8	3,5	1253	8
Seth I	6	11	3	153	11
Rozen WM	5	9	2,5	139	9
Xie Y	5	6	2,5	134	6
Currie G	4	4	2	63	4
Gupta R	4	4	2	72	4
Hunter-Smith DJ	4	4	2	129	4
Karakose T	4	4	2	45	4
Li B	4	6	2	41	7
Pack A	4	5	2	52	5

Table 5. Corresponding authors countries

Country	Articles	Articles %	SCP	MCP	MCP %
USA	453	22,6	396	57	12,6
China	193	9,6	151	42	21,8
United Kingdom	85	4,2	59	26	30,6
Australia	66	3,3	47	19	28,8
Germany	58	2,9	42	16	27,6
India	51	2,5	36	15	29,4
Turkey	51	2,5	45	6	11,8
Korea	45	2,2	34	11	24,4
Canada	40	2	22	18	45
Italy	39	1,9	27	12	30,8

Table 4 shows that TAN S has the highest h-index, while SETH I has the highest g-index, indicating differences in citation distribution across their articles. The higher m-index values for TAN S and SETH I reflect their consistent productivity over a more extended period. The author's local impact data is closely related to the countries of the corresponding author, as presented in Table 5. The United States (USA) is the most significant contributor, with 453 articles accounting for 22.6%. Most of these articles are single-country publications (SCP) (396), with a relatively low percentage of multi-country publications (MCP) at 12.6%. Despite the high scientific output, domestic researchers conduct much of the research independently. China is the second-largest contributor, with 193 articles (9.6%) and a higher proportion of MCP at 21.8%, indicating greater involvement in international collaboration than the USA.

The United Kingdom, Canada, Italy, and India show a high tendency for international collaboration, with MCP percentages of 30.6%, 45%, 30.8%, and 29.4%, respectively. It suggests that these countries are active in international cooperation, which enhances the diversification and quality of research. High MCP percentages indicate a positive inclination towards international collaboration, with countries possessing strong global research networks or specialized facilities often attracting collaborative interest from other nations. Government or institutional policies promoting collaboration can further increase MCP percentages. Figure 5 presents a world map where countries are shaded in varying gradients of blue. Darker blue shades represent higher levels of scientific production, while lighter shades indicate lower levels than dark blue. Some countries are not colored, indicating a lack of data or deficient scientific output.

Figure 5 illustrates the distribution of scientific production across countries, with varying shades of blue representing the intensity of scholarly output. Countries such as the United States are shaded in a very dark blue, indicating a very high level of scientific production. It includes factors such as the number of publications, research, or other scientific contributions. Other countries like Canada, China, Germany, and several European and Asian nations are depicted in lighter shades of blue, reflecting moderate levels of scientific output. Many countries in Africa, parts of Asia, and Latin America are shown in even lighter shades or are not colored at all, indicating low levels of scientific production or a lack of available data. Table 6 presents the Top 10 countries with the highest number of citations for their scientific work, highlighting the global impact and recognition of research contributions from these nations.

Table 6 enumerates the ten nations most frequently cited in academic publications based on total citations (TC) and average citations per article (ACA). The United States leads with the highest total citations (4039), despite its

average citations per article (8.90) not being the highest. It indicates a substantial volume of publications. The United Kingdom ranks second in total citations but boasts the highest average citations per article (25.30), indicative of the high quality of research produced by UK scholars. Australia ranks third in total citations, with a similarly high average of citations per article (18.50), underscoring its significant contribution to academic literature. China registers a considerable number of total citations, though with a lower average of citations per article (5.90), suggesting a high volume of publications with varied citation impact. India presents a moderate contribution in total and average citations per article (12.90), reflecting relevant and impactful research.

Hong Kong exhibits relatively high total citations and a notable average of citations per article (15.30). The United Arab Emirates demonstrates a high average of citations per article (17.20), indicative of influential research, despite its total citations being lower than other countries in this list. Italy shows a similar total citation count to the United Arab Emirates, with relatively high average citations per article (13.50), denoting significant academic contributions. Canada makes a moderate impact in both total citations and average citations per article (12.30), highlighting the relevance of its research. South Korea rounds out the list with the lowest total citations but a moderate average of citations per article (10.00), indicating a substantial number of publications with a reasonably good citation impact.

Figure 5. Countries scientific productions

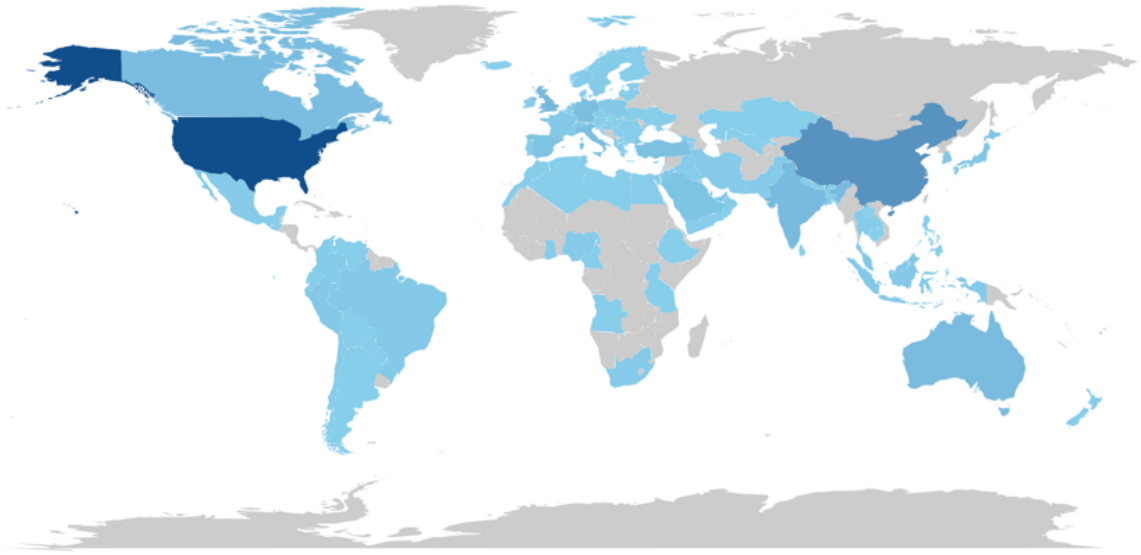


Table 6. Most cited countries

Country	TC	Average article citations
Usa	4039	8,90
United Kingdom	2151	25,30
Australia	1220	18,50
China	1141	5,90
India	658	12,90
Hong Kong	581	15,30
United Arab Emirates	533	17,20
Italy	525	13,50
Canada	492	12,30
Korea	451	10,00

4.1.4. Development of ChatGPT research issues

The total distribution of publications is closely related to the number of citations each year. Table 7 presents the top 10 most-cited documents, highlighting the most influential research outputs in the field.

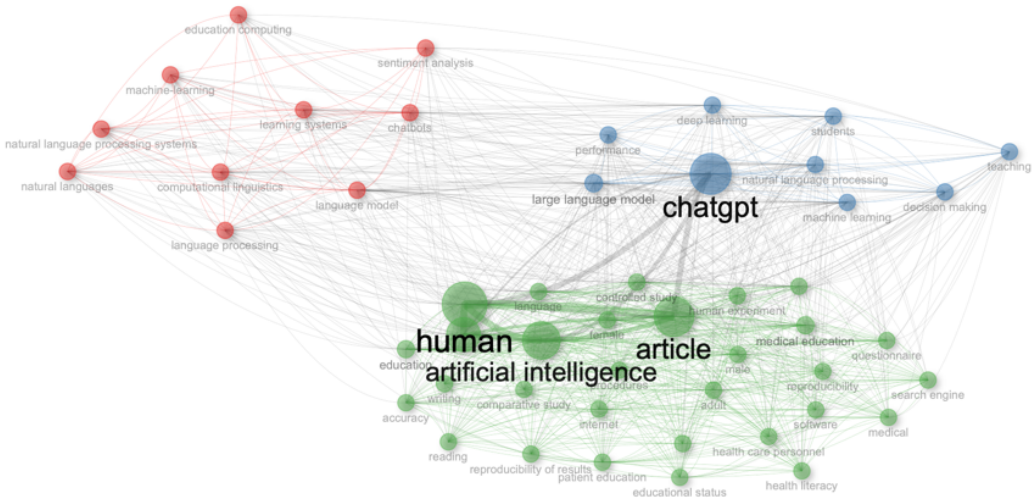
Table 7. Most cited document

Paper	DOI	TC	TCY	NTC
Dwivedi YK, 2023	10.1016/j.ijinfomgt.2023.102642	953	476,50	47,36
Gilson A, 2023	10.2196/45312	611	305,50	30,36
Tlili A, 2023	10.1186/s40561-023-00237-x	418	209,00	20,77
Rudolph J, 2023	10.37074/jalt.2023.6.1.9	400	200,00	19,88
Cotton DRE, 2024	10.1080/14703297.2023.2190148	334	334,00	171,40
Cascella M, 2023	10.1007/s10916-023-01925-4	313	156,50	15,55
Pavlik JY, 2023	10.1177/10776958221149577	289	144,50	14,36
Salvagno M, 2023	10.1186/s13054-023-04380-2	281	140,50	13,96
Cooper G, 2023	10.1007/s10956-023-10039-y	241	120,50	11,98
Lim WM, 2023	10.1016/j.ijme.2023.100790	237	118,50	11,78

Table 8. Trend topics

Term	Frequency	Year (Q1)	Year (Median)	Year (Q3)
Educational status	77	2023	2023	2024
Writing	57	2023	2023	2024
Human experiment	56	2023	2023	2023
Artificial intelligence	809	2023	2024	2024
Human	585	2023	2024	2024
ChatGPT	566	2023	2024	2024

Figure 6. Co-occurrences network



Citations used in scientific writing acknowledge the influence of previous scholarly work. The total number of citations an article receives is considered a precise indicator of its impact within a research domain. The recognition of the top 10 most-cited articles in Table 7 underscores the significance of these research outputs. These articles provide valid and credible information on ChatGPT's understanding, development, and impact. Articles with high citation counts are highly relevant to their research domain and offer critical solutions to existing problems, making them significant contributions that serve as a foundation for discoveries. The most-cited documents are closely related to the trending research topics in Table 8.

Table 8 reveals six prominent research topics and their associated trends. The issues garnering significant attention and influencing citation counts from 2023 to 2024 include Artificial Intelligence (AI), with a term frequency of 809. It is followed by the topics of humans, with a term frequency of 585, and ChatGPT, with a term frequency of 566. These three topics are particularly prominent and are a substantial focus of current research discussions. In addition, the trend topics educational status, writing, and human experiment have term frequencies of 77, 57, and 56, respectively. Despite having lower frequencies, these topics continue to attract the interest of researchers. All six trend topics are closely interconnected. Furthermore, the relationship between these terms extends beyond the six mentioned, encompassing related terms such as language model, medical education, learning, deep learning, controlled study, and others. Figure 6 illustrates the correlations between these various topics.

Table 9. Extraction of the studies in the review

Article	Sample size and characteristics	Type of research	Main finding
(Xu et al., 2024)	Nine experts: 3 ICT experts and six higher education professors.	Qualitative research	ChatGPT has the potential to address challenges faced by learners in Personal Learning Environments (PLEs) by enhancing cognitive, non-cognitive, and metacognitive skills
(Lyu et al., 2023)	Sixty-two low-dose chest computed tomography lung cancer screening scans and 76 brain magnetic resonance imaging metastases screening scans.	Experimental research	ChatGPT can successfully translate radiology reports into plain language with an average score of 4.27
(Papastratis et al., 2024)	3000 virtual user profiles and 1000 real profiles for experiments	Experimental research	AI-based diet recommendation methods can generate accurate and personalized weekly meal plans for different cuisines, population groups, and medical conditions with high accuracy and variability.
(Jo et al., 2024)	Ten evaluators	Evaluative research study	the three Generative Artificial Intelligence (GAI) tools, GPT-4, Google Bard, and CLOVA X, demonstrated similar or better communicative competence compared to the information book regarding questions
(Wang et al., 2023)	Fifty-two medical students participated in the 2020 examination, 49 medical students took part in the 2021 examination, and 65 students participated in the 2022 examination.	Comparative research	ChatGPT's performance in the Chinese National Medical Licensing Examination (NMLE) was lower than that of medical students, with its correct answer rate being influenced by the year of the exam questions released, indicating the potential for improvement through deep learning.
(Sun, 2024)	100 students' English	Experimental in nature research	the ChatGPT-based language model demonstrates the remarkable potential for diverse applications in English language education
(Avello-Martínez et al., 2024)	41 master's students: an experimental group with 20 students and a control group with 21 students.	Experimental research	ChatGPT did not significantly improve digital storytelling skills in the short term.
(Sabir et al., 2024)	219,294 ChatGPT tweets were categorized as 56,012 positive, 107,796 negative, and 554,870 neutral.	Experimental research	achieving an accuracy of 96.41% with the SVM classifier when using TF-IDF as a feature extraction technique.
(Choi, 2023)	312 MCQs derived from the K-CBMSE test items	Evaluative research	ChatGPT exhibited % overall accuracy of 76.0% in answering multiple-choice questions, with limitations in generating correct rationales and references.
(Rahman & Watanobe, 2023)	60 teachers	Primarily exploratory and experimental	ChatGPT offers significant advantages for researchers by supporting the writing process and assisting in code optimization.
(Jošt et al., 2024)	Thirty-two participants	Experimental study	there is a negative correlation between increased reliance on Large Language Models (LLMs) for critical thinking-intensive tasks like code generation
(Alanzi, 2023)	54 participants	Qualitative research	ChatGPT has several advantages for teleconsultants in the healthcare sector, but it is associated with ethical issues.
(Tyni et al., 2024)	88 participants	Mixed-methods study	ChatGPT 3.5 outperforms 4.0 with student input, and integrating expert roles in prompts is unreliable and necessary only with game designer input for ChatGPT 3.5.
(Hosseini et al., 2023)	420 respondents	Developing a research	there was greater interest in using ChatGPT
(Kosar et al., 2024)	182 participants	Experimental research	ChatGPT usage did not significantly influence the students' performance in practical assignments, grading results, or midterm exams
(Fütterer et al., 2023)	5,541,457 users tweets	Analyzing Twitter data	the global reception on Twitter about ChatGPT was rapid and widespread
(Piccolo et al., 2023)	184 programming	Evaluation study	ChatGPT successfully solved 139 out of 184 programming exercises, demonstrating its potential to aid in programming tasks in life sciences education and research.
(Taloni et al., 2023)	1023 questions	Comparative cross-sectional study	GPT-4.0 outperformed both humans and GPT-3.5 in answering multiple-choice questions from the American Academy of Ophthalmology self-assessment program

Article	Sample size and characteristics	Type of research	Main finding
(Tsai et al., 2024)	450 multiple-choice questions	Cross-sectional study	ChatGPT-4 demonstrated strengths in urology with an overall accuracy of 57.8%, surpassing ChatGPT-3.5 significantly
(Cohen et al., 2023)	150 authentic questions	compared the performance of ChatGPT to the real-life actual performance	ChatGPT's performance in Hebrew OBGYN residency examinations was significantly lower compared to the actual performance of residents and English language tests
(Ng & Chow., 2024)	4,251,662 tweets.	quantitative study	the identification of six prominent peaks in ChatGPT conversations across 4 million tweets
(Gobira et al., 2023)	7,006 participating	Assessment of the performance of ChatGPT-4.0	ChatGPT-4.0 exhibited an overall accuracy of 87.7% in non-nullified questions and 71.4% in nullified questions
(Dubiel et al., 2024)	Query scope (3 classes), Query purpose (6 classes), Response format (4 classes), Information feature (13 classes)	Experimental research	Lightweight LLMs can be fine-tuned for on-device deployment
(Li, 2023)	42 students were in the experimental group, and 39 were in the control group.	Quasi-experiment	ChatGPT-based flipped learning guiding approach significantly improved students' performance
(Kämmer et al., 2024)	N=640	Prospective, randomized controlled experiment	Examine the influence on the diagnostic process and outcomes of interacting with a large language model compared with a human coach
(Al Ghazali et al., 2024)	39 students, which was relatively small in size.	Experimental research	ChatGPT shows promise in knowledge recall and reasoning skills. It faces challenges related to student engagement and completion rates.
(Farazouli et al., 2024)	24 university teachers	Qualitative approach	patterns of downgrading in student-written texts indicate a more critical approach to grading these texts than chatbot-generated responses.
(Noy & Zhang., 2023)	600 participants	Experimental research	Replicating treatment effects on time using an objective measure of 'time active.'
(Collins et al., 2024)	a mixed cohort of participants ranging from undergraduate students to professors of mathematics.	Empirical research & experimental research	Correctness and helpfulness of model responses are highly correlated, but they can diverge in exciting ways
(Currie & Barry,2023)	Long-answer–style questions (8 subjects) and calculation-style questions (2 subjects) were included for examinations.	Experimental study	ChatGPT powered by GPT 3.5 performed poorly in calculation examinations and written tasks, scoring significantly lower than students across nuclear medicine subjects, indicating limitations in depth of insight, research breadth, and information currency.
(Groza et al., 2024)	GPT-3.5-turbo and gpt-4.0	The experimental setup of the study	GPT-4.0 surpasses the state-of-the-art performance in phenotype concept recognition tasks when constrained to a subset of the target ontology with prior knowledge of expected terms.
(Kufel et al., 2023)	120 questions	Prospective study focused on one specialty exam in radiology and diagnostic imaging	ChatGPT's performance in the pass rate of Poland's radiology and imaging diagnostics specialty exam is yet to be determined, requiring further research on improved model versions.
(Tong et al., 2023)	160 questions for evaluation	Experimental research type	ChatGPT demonstrated a correct response rate of 81.25% for Chinese and 86.25% for English questions, with good predictive performance indicated by Brier Scores of 0.19 for Chinese and 0.14 for English, showing promise for global healthcare despite language bias challenges.
(Tang et al., 2023)	six domain experts	Evaluating large language models (LLMs) for medical evidence summarization	Large language models (LLMs) like GPT-3.5 and ChatGPT struggle with identifying salient information, are error-prone in generating summaries over more extended textual contexts, and produce factually inconsistent and misleading summaries, which could lead to potential harm due to misinformation.

Based on Figure 6, the classification is elucidated through cluster analysis distinguished by varying colors. The results indicate 49 items meeting the threshold across three clusters. The first cluster, represented in red, contains ten items. The second cluster, depicted in blue, encompasses nine items. The third cluster, illustrated in green, includes 30 items. This correlation also visualizes the network of research topic trends. The cluster analysis reveals that the first cluster (red) primarily addresses topics related to language modeling or language processing within educational systems assisted by ChatGPT or artificial intelligence. The research subtopics include language models, chatbots, computational linguistics, learning systems, natural language processing systems, language processing, sentiment analysis, educational computing, natural languages, and machine learning. Research within this cluster focuses on the potential of language modeling, particularly in computing, medical education, and machine learning using artificial intelligence (AI).

The second cluster (blue) centers on ChatGPT, the principal topic of this research. The research subtopics in this cluster encompass ChatGPT, large language models, machine learning, students, natural language processing, deep learning, decision-making, performance, and teaching. These subtopics relate to ChatGPT during the study analysis process. The third cluster (green) focuses on artificial intelligence, commonly called AI. The subtopics in this cluster include AI, human, article, language, medical education, controlled study, patient education, learning, educational status, human experiment, comparative analysis, reproducibility of results, and several other topics. This cluster concentrates on AI-related issues, with ChatGPT being one of the AI products.

The investigation into and understanding of the impact of ChatGPT within the learning environment, based on the issues raised from the cluster analysis, suggests that natural language processing models or language processing in programming code creation can enhance biomedical information skills among medical students. Furthermore, ChatGPT, an advanced AI model within the learning environment, can significantly foster student engagement and comprehension. The impact of ChatGPT on the learning experience, particularly in writing, grammar, and vocabulary expansion, demonstrates positive outcomes for students. ChatGPT presents opportunities for educational institutions and acts as a catalyst for enhancing the quality and accessibility of education.

4.2. Systematic literature review

The utilization of ChatGPT in education has generated a variety of positive and negative impacts. Analysis from multiple articles reveals diverse outcomes depending on the context and application. Some studies indicate improvements in skills and learning experiences, limitations and challenges, potential in professional training and education, influence on teaching and assessment, and ethical considerations. ChatGPT can enhance cognitive, non-cognitive, and metacognitive skills by providing personalized feedback with tailored learning resources (Xu et al., 2024). Additionally, in the context of flipped learning, a ChatGPT-based approach has been shown to improve student performance, self-efficacy, learning attitudes, intrinsic motivation, and creative thinking (Li, 2023). Despite its substantial potential, ChatGPT also exhibits limitations in certain aspects. For instance, GPT-3.5-powered ChatGPT performs poorly in calculations and writing tasks, indicating limitations in the depth of insight and current information (Currie & Barry, 2023). Furthermore, using large language models (LLMs) for tasks requiring intensive critical thinking, such as code creation and debugging, negatively correlate with final grades in programming education (Jošt et al., 2024).

In professional education, ChatGPT demonstrates varying capabilities. It can accurately and personally provide weekly meal plans and simplify health information while retaining key messages (Papastratis et al., 2024). However, ChatGPT's performance on China's national medical licensing exams is lower than that of medical students, indicating a need for further enhancement in academic abilities (Wang et al., 2023). Additionally, the use of ChatGPT in teaching can influence how instructors assess student-written texts. Farazouli et al. (2024) found that instructors tend to be more critical of texts written by students than responses generated by chatbots, indicating a difference in assessment standards when AI technology is involved in the teaching process. ChatGPT can support telemedicine consultations in the healthcare sector, but ethical issues must be addressed (Alanzi, 2023). Concerns about language bias and the inability to accurately identify critical information also pose challenges (Tang et al., 2023). While ChatGPT offers numerous benefits in learning and education, including assisting in teaching and personalizing learning, significant challenges must be addressed. ChatGPT must be monitored and adapted to ensure its impact remains positive and beneficial to the educational process.

5. Discussions and conclusions

The findings from the bibliometric analysis reveal the evolution of research on ChatGPT from 2022 to 2024. Research on ChatGPT began to emerge in 2022 with two published articles and saw a significant increase in 2023

and 2024, reaching 813 and 1189 articles, respectively. This growth reflects the rising interest in the impact and applications of ChatGPT, particularly in education. However, the average citations per article decreased from 117.50 in 2022 to 1.95 in 2024, coinciding with the increased number of published articles. Although the number of articles increased, the average citations per article declined due to the tendency of a higher volume of publications to reduce the average citation impact.

Citation source analysis indicates that the journal JMIR Medical Education stands out with 46 documents and possesses vital metrics, including an h-index of 12 and 1112 citations. Articles from this journal are frequently cited in research discussing the use of ChatGPT in medical education. Other journals, such as Education and Information Technologies and Computers and Education: Artificial Intelligence, contribute significantly to the related literature. Citation sources are also linked to collaboration. Collaboration analysis shows that the United States has the highest contribution of articles (22.6%), followed by China (9.6%). Countries like the United Kingdom, Canada, and Italy exhibit high levels of international collaboration, potentially enhancing the diversity and quality of research.

The United States possesses the highest total citation count, which indicates its high productivity in academic publications. However, its average citations per article are lower than those of countries like the United Kingdom and Australia. It suggests that while the U.S. has many publications, the quality and impact per article not be as high as those of the countries mentioned above. Despite having a lower total citation count than the U.S., the United Kingdom demonstrates exceptionally high research quality with the highest average citations per article. It signifies that publications from the U.K. are highly influential and frequently cited in academic literature. Australia and the United Arab Emirates also exhibit high research quality with significant average citations per article. In contrast, China, with a high total citation count, has a lower average citation per article, indicating variability in the impact of its publications. Other countries, such as India, Hong Kong, Italy, and Canada, demonstrate substantial contributions to academic research with relatively high average citations per article, indicating the relevance and influence of their research outputs.

Artificial intelligence (AI) emerges as the most frequently discussed topic, with high frequency in related terms such as "ChatGPT," "human," and "educational status." Cluster analysis indicates that the main topics include language processing models, ChatGPT itself, and AI in general. ChatGPT has become a highly intriguing subject in scientific research, particularly regarding its impact on education. ChatGPT in education offers significant opportunities to improve the quality of learning and expand access to education, with substantial potential to develop students' language skills and comprehension.

International collaboration can enhance access to resources and technology, expand researchers' networks, and enrich perspectives in research. It can certainly increase the visibility and impact of research. However, cultural differences, language barriers, regulatory disparities, and varying research priorities can hinder collaboration. Additionally, coordinating researchers from different countries can present logistical challenges. Independent research can focus more on specific national issues and be quicker and more efficient since it does not require cross-country coordination. Collaborative research can yield more holistic and globally applicable knowledge through broader sharing of resources and expertise. Countries with a high percentage of multi-country publications (MCP) demonstrate greater involvement in global collaboration, which can enrich research outcomes. On the other hand, countries with a high proportion of single-country publications (SCP) focus more on national issues. Both approaches have advantages and challenges, and a deep understanding of these dynamics can aid in developing better research strategies.

ChatGPT has proven effective in enriching students' learning experiences by providing personalized feedback and tailored learning materials. However, ChatGPT powered by the GPT-3.5 model has limitations in handling tasks that require complex calculations and producing in-depth writing. ChatGPT is still limited in providing deep and accurate insights in some fields. As demonstrated by the results of China's national medical licensing exam, its academic capabilities are still below the expected standard. Despite its potential as an educational tool, ChatGPT still needs further improvement to be reliable in teaching materials that require deep and specialized knowledge.

The influence of ChatGPT on teaching and assessment also warrants attention. It can introduce bias in evaluation, affecting the objectivity and validity of academic assessments. Establishing transparent and fair guidelines for using this technology is essential to ensure consistent and transparent assessments. The ethical aspects of using ChatGPT are also a primary concern. The use of AI in medical consultation and education requires strict policies to protect user privacy and prevent data misuse. Additionally, language biases in this model could lead to misinterpretations or overlooking critical information, which can negatively impact decision-making processes.

Research on ChatGPT significantly increased from 2022 to 2024, with published articles surging from merely two in 2022 to 813 in 2023 and 1189 in 2024. This increase reflects substantial interest in ChatGPT's applications and impact, particularly in education. Despite the growing number of articles, the average number of citations per article declined from 117.50 in 2022 to 1.95 in 2024, indicating that the proliferation of publications can dilute the average citation impact.

JMIR Medical Education has emerged as a primary citation source with solid contributions, followed by other journals such as Education and Information Technologies and Computers and Education: Artificial Intelligence. The United States stands out as the most significant contributor with a high level of international collaboration, followed by countries such as China, the United Kingdom, Canada, and Italy. This global collaboration is crucial for enriching the perspectives and quality of research, although it also faces challenges such as cultural and regulatory differences. Overall, research on ChatGPT demonstrates significant potential for enhancing the quality of education and global accessibility. ChatGPT offers numerous advantages in learning and education, including assisting in teaching and personalizing learning. However, significant challenges must be addressed, including technical limitations and ethical concerns. Therefore, the use of ChatGPT in education must be carefully monitored and adapted to ensure its sustained positive impact and contribution to improving education quality.

5.1. Theoretical implications

The increased use of ChatGPT in education significantly contributes to developing educational technology theories, particularly in language-based learning, personalized learning, and cognitive development (Dalgıç et al., 2024) . These findings suggest that ChatGPT can be pivotal in expanding existing theoretical frameworks, such as constructivist learning theory, the Technological Pedagogical Content Knowledge (TPACK) framework, and artificial intelligence (AI) -based learning models. Consequently, future research could explore how ChatGPT enriches or challenges established learning models, focusing on personalized feedback, enhanced student engagement, and developing critical thinking skills. These findings also reflect a paradigm shift toward AI-driven education that is more student-centered. ChatGPT can contribute substantially to understanding how AI-based tools can support self-directed learning, metacognitive development, and knowledge construction by students. Future theoretical research should investigate the role of AI in helping or even redefining existing learning theories. However, it is essential to note that the limitations and potential biases inherent in AI models like ChatGPT must be a central concern, leading to discussions on the ethics and reliability of AI in educational contexts.

5.2. Practical implications

The use of ChatGPT in educational settings shows significant potential in enhancing students' cognitive, non-cognitive, and metacognitive skills. ChatGPT can be integrated into personalized learning pathways, offering students tailored feedback and relevant learning resources. For example, ChatGPT can enhance students' intrinsic motivation, improve their understanding, and stimulate creative and critical thinking within the flipped classroom model. As such, educators must be trained to effectively leverage ChatGPT in curriculum design and create interactive learning environments. In the context of professional education, ChatGPT demonstrates a range of capabilities. For instance, ChatGPT can simulate patient interactions or simplify complex medical information in medical education. However, its use still needs refinement to meet the more stringent standards of national examinations. Therefore, educational institutions must provide professional development for instructors on integrating AI into their teaching practices and educating students on critically engaging with AI tools.

Clear ethical guidelines must accompany the implementation of ChatGPT in education to ensure fair and secure usage. These guidelines should include standards for assessing AI-generated written work and methods for ensuring transparency in AI-driven feedback systems. Privacy concerns and potential biases in the responses provided by ChatGPT must also be addressed to ensure the technology is used optimally and does not harm students. Moreover, ChatGPT can be an extremely valuable tool in supporting the learning of students with diverse needs, particularly in inclusive education. For instance, with its ability to provide immediate feedback and support self-directed learning, ChatGPT can assist students with learning disabilities or those who require specialized support, especially in underserved areas with limited access to traditional educational resources.

5.3. Limitations and future research directions

This study offers valuable insights into the impact of ChatGPT usage in education, but it has several limitations that should be acknowledged. One such limitation is the restricted scope of the data analyzed, where the bibliometric analysis only includes articles indexed in the Scopus database, thus excluding other potentially relevant

sources. The result in the omission of important articles that have not been indexed or those published on alternative platforms, which could influence the overall depiction of research trends. Additionally, this study primarily focuses on quantitative analysis, such as publication counts and citation numbers. It does not provide a comprehensive qualitative examination of ChatGPT's influence on learning outcomes across various educational settings. Moreover, the research does not delve deeply into the potential risks and challenges associated with ChatGPT's use, such as the long-term effects of AI on students' critical thinking skills or more complex ethical issues related to applying this technology in education. Many studies focus on immediate impacts within the initial years of ChatGPT's implementation, leaving long-term social and technical concerns largely unexplored.

Future research should aim to conduct more in-depth and extensive studies involving a wider array of platforms and databases beyond Scopus to expand the data coverage and include articles that have not yet been indexed. The research would provide a more comprehensive view of the development of ChatGPT-related research in education and facilitate comparisons across different research sources. Furthermore, subsequent studies should aim to understand the long-term effects of ChatGPT on education. Longitudinal studies are essential to assess how the use of this AI tool impacts the development of critical thinking, problem-solving, and social skills among students. This approach would allow researchers to observe whether there are significant changes in learning quality and student development over time. Future research could also focus on a deeper exploration of ChatGPT usage's ethical and social challenges. For example, how can we ensure that AI is deployed in a manner that is equitable, unbiased, and accessible to all social groups? Research could investigate how educational policies can be crafted to address potential inequalities stemming from uneven technological dependence between more developed and less developed schools or regions.

References

Alanzi, T. M. (2023). Impact of ChatGPT on teleconsultants in healthcare: perceptions of healthcare experts in Saudi Arabia. *Journal of Multidisciplinary Healthcare*, *16*, 2309–2321. https://doi.org/10.2147/JMDH.S419847

Al Ghazali, S., Zaki, N., Ali, L., & Harous, S. (2024). Exploring the potential of ChatGPT as a substitute teacher: A case study. *International Journal of Information and Education Technology*, *14*(2), 271-278.

Al Lily, A. E., Ismail, A. F., Abunaser, F. M., Al-Lami, F., & Abdullatif, A. K. A. (2023). ChatGPT and the rise of semi-humans. *Humanities and Social Sciences Communications*, *10*(1), 1-12.

Avello-Martínez, R., Gajderowicz, T., & Gómez-Rodríguez, V. G. (2024). Is ChatGPT helpful for graduate students in acquiring knowledge about digital storytelling and reducing their cognitive load? An experiment. *Revista de Educación a Distancia*, *24*(78). https://doi.org/10.6018/red.604621

Castillo, A. G. R., Rivera, H. V. H., Teves, R. M. V., Lopez, H. R. P., Reyes, G. Y., Rodriguez, M. A. M., ... & Arias-Gonzáles, J. L. (2023). Effect of ChatGPT on the digitized learning process of university students. *Journal of Namibian Studies: History Politics Culture*, *33*, 1-15.

Choi, W. (2023). Assessment of the capacity of ChatGPT as a self-learning tool in medical pharmacology: a study using MCQs. *BMC Medical Education*, *23*(1), 1-15. https://doi.org/10.1186/s12909-023-04832-x

Cohen, A., Alter, R., Lessans, N., Meyer, R., Brezinov, Y., & Levin, G. (2023). Performance of ChatGPT in Israeli Hebrew OBGYN national residency examinations. *Archives of Gynecology and Obstetrics*, *308*(6), 1797–1802. https://doi.org/10.1007/s00404-023-07185-4

Collins, K. M., Jiang, A. Q., Frieder, S., Wong, L., Zilka, M., Bhatt, U., ... & Jamnik, M. (2024). Evaluating language models for mathematics through interactions. *Proceedings of the National Academy of Sciences*, *121*(24), 1-21.

Currie, G., & Barry, K. (2023). ChatGPT in nuclear medicine education. *Journal of Nuclear Medicine Technology*, *51*(3), 247-254.

Dalalah, D., & Dalalah, O. M. (2023). The false positives and false negatives of generative AI detection tools in education and academic research: The case of ChatGPT. *The International Journal of Management Education*, *21*(2), 1-12.

Dalgıç, A., Yaşar, E., & Demir, M. (2024). ChatGPT and learning outcomes in tourism education: The role of digital literacy and individualized learning. *Journal of Hospitality, Leisure, Sport & Tourism Education*, *34*, 1-13.

Dindorf, C., Bartaguiz, E., Gassmann, F., & Fröhlich, M. (2022). Conceptual structure and current trends in artificial intelligence, machine learning, and deep learning research in sports: a bibliometric review. *International Journal of Environmental Research and Public Health*, *20*(1), 173-185.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of business research*, *133*, 285-296.

Dowling, M., & Lucey, B. (2023). ChatGPT for (finance) research: The Bananarama conjecture. *Finance Research Letters*, *53*, 1-12.

Duarte, A. V., Zhao, X., Oliveira, A. L., & Li, L. (2024). De-cop: Detecting copyrighted content in language models training data. *arXiv preprint arXiv:2402.09910*.

Dubiel, M., Barghouti, Y., Kudryavtseva, K., & Leiva, L. A. (2024). On-device query intent prediction with lightweight LLMs to support ubiquitous conversations. *Scientific Reports*, *14*(1) 1-14. https://doi.org/10.1038/s41598-024-63380-6

Einarsson, H., Lund, S. H., & Jónsdóttir, A. H. (2024). Application of ChatGPT for automated problem reframing across academic domains. *Computers and Education: Artificial Intelligence*, *6*, 1-14.

ElSayary, A. (2024). An investigation of teachers' perceptions of using ChatGPT as a supporting tool for teaching and learning in the digital era. *Journal of Computer Assisted Learning*, *40*(3), 931-945.

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2024). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment and Evaluation in Higher Education*, *49*(3), 363–375. https://doi.org/10.1080/02602938.2023.2241676

Firdaus, T. (2023). Representative platform cyber metaverse terkoneksi BYOD sebagai upaya preventive urgensi digital pada sistem pendidikan Indonesia. *Jurnal Integrasi dan Harmoni Inovatif Ilmu-Ilmu Sosial*, *3*(2), 123-131. https://doi.org/10.17977/um063v3i2p123-131

Firdaus, T., Sholeha, S. A., Jannah, M., & Setiawan, A. R. (2024). Comparison of ChatGPT and Gemini AI in Answering Higher-Order Thinking Skill Biology Questions: Accuracy and Evaluation. *International Journal of Science Education and Teaching, 3*(3), 126-138. https://doi.org/10.14456/ijset.2024.11

Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M., & Gerjets, P. (2023). ChatGPT in education: global reactions to AI innovations. *Scientific reports*, *13*(1), 1-10.

George, A. S., George, A. H., & Martin, A. G. (2023). The environmental impact of AI: a case study of water consumption by chat GPT. *Partners Universal International Innovation Journal*, *1*(2), 97-104.

Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., ... & Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging era of AI Chatbots. *Internet of Things and Cyber-Physical Systems*, *4*, 19-23.

Gobira, M., Nakayama, L. F., Moreira, R., Andrade, E., Regatieri, C. V. S., & Belfort, R. (2023). Performance of ChatGPT-4 in answering questions from the Brazilian National Examination for Medical Degree Revalidation. *Revista Da Associacao Medica Brasileira*, *69*(10), 1-16. https://doi.org/10.1590/1806-9282.20230848

Groza, T., Caufield, H., Gration, D., Baynam, G., Haendel, M. A., Robinson, P. N., ... & Reese, J. T. (2024). An evaluation of GPT models for phenotype concept recognition. *BMC Medical Informatics and Decision Making*, *24*(1), 30-42.

Hosseini, M., Gao, C. A., Liebovitz, D. M., Carvalho, A. M., Ahmad, F. S., Luo, Y., MacDonald, N., Holmes, K. L., & Kho, A. (2023). An exploratory survey about using ChatGPT in education, healthcare, and research. *PLoS ONE*, *18*(10), 1-13. https://doi.org/10.1371/journal.pone.0292216

Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *3*(2), 1-15.

Jo, M. H., Kim, M. J., Oh, H. K., Choi, M. J., Shin, H. R., Lee, T. G., ... & Kang, S. B. (2024). Communicative competence of generative artificial intelligence in responding to patient queries about colorectal cancer surgery. *International Journal of Colorectal Disease*, *39*(1), 94-109.

Jošt, G., Taneski, V., & Karakatič, S. (2024). The impact of large language models on programming education and student learning outcomes. *Applied Sciences*, *14*(10), 4-15.

Kabir, S., Udo-Imeh, D. N., Kou, B., & Zhang, T. (2024, May). Is stack overflow obsolete? an empirical study of the characteristics of ChatGPT answers to stack overflow questions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1-17).

Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, *15*(16), 1-11.

Kämmer, J. E., Hautz, W. E., Krummrey, G., Sauter, T. C., Penders, D., Birrenbach, T., & Bienefeld, N. (2024). Effects of interacting with a large language model compared with a human coach on the clinical diagnostic process and outcomes among fourth-year medical students: study protocol for a prospective, randomised experiment using patient vignettes. *BMJ Open*, *14*(7), 1-9. https://doi.org/10.1136/bmjopen-2024-087469

Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, 1-24.

Kosar, T., Ostojić, D., Liu, Y. D., & Mernik, M. (2024). Computer Science Education in ChatGPT Era: Experiences from an Experiment in a programming course for novice programmers. *mathematics*, *12*(5), 1-17.

Kufel, J., Paszkiewicz, I., Bielówka, M., Bartnikowska, W., Janik, M., Stencel, M., Czogalik, Ł., Gruszczyńska, K., & Mielcarska, S. (2023). Will ChatGPT pass the Polish specialty exam in radiology and diagnostic imaging? Insights into strengths and limitations. *Polish Journal of Radiology*, *88*(1), e430–e434. https://doi.org/10.5114/pjr.2023.131215

Li, H. (2023). Effects of a ChatGPT-based flipped learning guiding approach on learners' courseware project performances and perceptions. *Australasian Journal of Educational Technology*, *39*(5), 40-58.

Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., & Chocksathaporn, P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching*, *6*(1), 64-74.

Lyu, Q., Tan, J., Zapadka, M. E., Ponnatapura, J., Niu, C., Myers, K. J., Wang, G., & Whitlow, C. T. (2023). Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. *Visual Computing for Industry, Biomedicine, and Art*, *6*(1). https://doi.org/10.1186/s42492-023-00136-5

Ng, R., & Chow, T. Y. J. (2024). Powerful tool or too powerful? Early public discourse about ChatGPT across 4 million tweets. *Plos one*, *19*(3), e0296882.

Nightingale, A. (2009). A guide to systematic literature reviews. *Surgery (Oxford)*, *27*(9), 381-384.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187-192.

OpenAi. (2022). Introducing ChatGPT (November 30, 2022). https://openai.com/index/ChatGPT/

OpenAi. (2024). Introducing GPT-4o and more tools to ChatGPT free users. https://openai.com/index/gpt-4o-and-more-tools-to-ChatGPT-free/ (accessed May 13, 2024).

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ, 372,* 1-14.

Papastratis, I., Stergioulas, A., Konstantinidis, D., Daras, P., & Dimitropoulos, K. (2024). Can ChatGPT provide appropriate meal plans for NCD patients?. *Nutrition*, *121*, 1-11.

Piccolo, S. R., Denny, P., Luxton-Reilly, A., Payne, S. H., & Ridge, P. G. (2023). Evaluating a large language model's ability to solve programming exercises from an introductory bioinformatics course. *PLoS Computational Biology*, *19*(9), 1-18. https://doi.org/10.1371/journal.pcbi.1011511

Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences (Switzerland)*, *13*(9), 1-15. https://doi.org/10.3390/app13095783

Sabir, A., Ali, H. A., & Aljabery, M. A. (2024). ChatGPT tweets sentiment analysis using machine learning and data classification. *Informatica (Slovenia)*, *48*(7), 103–112. https://doi.org/10.31449/inf.v48i7.5535

Sadiku, M. N., Musa, S. M., & Chukwu, U. C. (2022). *Artificial intelligence in education*. LA, Universe.

Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. *IEEE Access*, *11*, 38805-38818.

Similiarweb. (2024). Chatopenai. https://www.similarweb.com/website/chat.openai.com/#traffic (accessed May 13, 2024).

Srinivasan, N., Samaan, J. S., Rajeev, N. D., Kanu, M. U., Yeo, Y. H., & Samakar, K. (2024). Large language models and bariatric surgery patient education: a comparative readability analysis of GPT-3.5, GPT-4, Bard, and online institutional resources. *Surgical Endoscopy*, *38*(5), 2522-2532.

Sun, H. (2024). Multi-scenario application of ChatGPT-based language modeling for empowering English language teaching and learning. *Applied Mathematics and Nonlinear Sciences*, *9*(1). https://doi.org/10.2478/amns-2024-0790

Taloni, A., Borselli, M., Scarsi, V., Rossi, C., Coco, G., Scorcia, V., & Giannaccare, G. (2023). Comparative performance of humans versus GPT-4.0 and GPT-3.5 in the self-assessment program of American Academy of Ophthalmology. *Scientific Reports*, *13*(1), 1-16. https://doi.org/10.1038/s41598-023-45837-2

Tang, L., Sun, Z., Idnay, B., Nestor, J. G., Soroush, A., Elias, P. A., Xu, Z., Ding, Y., Durrett, G., Rousseau, J. F., Weng, C., & Peng, Y. (2023). Evaluating large language models on medical evidence summarization. *Npj Digital Medicine*, *6*(1), 1-17. https://doi.org/10.1038/s41746-023-00896-7

Tirado-Olivares, S., Navío-Inglés, M., O'Connor-Jiménez, P., & Cózar-Gutiérrez, R. (2023). From human to machine: investigating the effectiveness of the conversational AI ChatGPT in historical thinking. *Education Sciences*, *13*(8), 803-819.

Tong, W., Guan, Y., Chen, J., Huang, X., Zhong, Y., Zhang, C., & Zhang, H. (2023). Artificial intelligence in global health equity: an evaluation and discussion on the application of ChatGPT, in the Chinese National Medical Licensing Examination. *Frontiers in Medicine*, *10*, 1-16. https://doi.org/10.3389/fmed.2023.1237432

Tsai, C. Y., Hsieh, S. J., Huang, H. H., Deng, J. H., Huang, Y. Y., & Cheng, P. Y. (2024). Performance of ChatGPT on the Taiwan urology board examination: insights into current strengths and shortcomings. *World Journal of Urology*, *42*(1), 1-17. https://doi.org/10.1007/s00345-024-04957-8

Tyni, J., Turunen, A., Bednarik, R., Kahila, J., & Tedre, M. (2024). International Journal of Serious Games can ChatGPT match experts? Comparing input for serious game development. *International Journal of Serious Games I*, *11*(2), 1-14. https://doi.org/10.17083/ijsg

Tyson, J. (2023). Shortcomings of ChatGPT. *Journal of Chemical Education*, *100*(8), 3098-3101.

Wang, X., Gong, Z., Wang, G., Jia, J., Xu, Y., Zhao, J., Fan, Q., Wu, S., Hu, W., & Li, X. (2023). ChatGPT Performs on the Chinese National Medical Licensing Examination. *Journal of Medical Systems*, *47*(1), 1-17. https://doi.org/10.1007/s10916-023-01961-0

Xu, X. S., Wang, X. B., Zhang, Y. F., & Zheng, R. (2024). Applying ChatGPT to tackle the side effects of personal learning environments from learner and learning perspective: An interview of experts in higher education. *PLoS ONE*, *19*(1), 1-16. https://doi.org/10.1371/journal.pone.0295646

Vargas-Murillo, A. R., de la Asuncion, I. N. M., & de Jesús Guevara-Soto, F. (2023). Challenges and opportunities of AI-assisted learning: A systematic literature review on the impact of ChatGPT usage in higher education. *International Journal of Learning, Teaching and Educational Research*, *22*(7), 122-135

Author contribution statements

The authors equally conducted the research design and implementation, analysis, and article writing without using AI applications.

Disclosure statement

The authors reported no potential competing interest.

Ethical committee approval

This study has complied with the Research Publication Ethics stated in "Wager E & Kleinert S (2011) Responsible research publication: international standards for authors. A position statement was developed at the 2nd World Conference on Research Integrity, Singapore, July 22-24, 2010. Chapter 50 in Mayer T & Steneck N (eds) Promoting Research Integrity in a Global Environment. Imperial College Press / World Scientific Publishing, Singapore". For this reason, the author states that he conducted the research within the framework of ethical principles. It is not a human study, so ethical approval is not required. All responsibility belongs to the authors.