

Bibliometric Analysis of Potential Themes and Trend Development of ChatGPT in the Field of Education

Liu Shang

*School of Education and Languages, Hong Kong Metropolitan University, Ho Man Tin, Kowloon,
Hong Kong, 999077, China
1464395803@qq.com*

Keywords: Bibliometrics; Data Analysis; ChatGPT; Education; VOS viewer; SATI

Abstract: The purpose of this article is to review existing research on ChatGPT in education through bibliometric analysis. The research questions are: First, what are the basic characteristics of the research on issues related to ChatGPT in educational applications? Second, what are the main research topics of ChatGPT in educational applications? Third, how has the international topic of ChatGPT in educational applications evolved? The research method is to use VOSviewer and SATI's bibliometric mapping research tools to analyze and visualize the author information and keywords of published articles to observe the basic characteristics, high-frequency research keywords and the evolution trend of the research topic in the field. The expected research contribution is to provide help and reference for researchers in their research in the field of ChatGPT education. The research results reveal several key findings. First, the number of documents published in journals within this research field is on an overall upward trend. Asian scholars and research institutions significantly contribute to the volume of published articles. The region with the highest number of publishing institutions is Asia, specifically Japan. Among the authors of these published articles, Chinese scholars hold the largest proportion, occupying the second and third positions. Second, on an international scale, the primary research topic is Artificial Intelligence. Third, the evolution of ChatGPT's international theme in educational applications has progressed from early scholars' research on ChatGPT school courses to investigations into the performance of artificial intelligence itself. Subsequently, it shifted towards research on ChatGPT medical education and ultimately evolved into studies related to ChatGPT language models and patient education.

1. Introduction

What is ChatGPT? ChatGPT (<https://openai.com/blog/chatgpt>) is the latest release of the Generative Pretrained Transformer (GPT) family of language models released by OpenAI (<https://openai.com>) on November 30, 2022. A language model is a statistical model that can predict the probability of a sequence of words. With this capability, a language model can generate natural language in a human style. Like all statistical models, a language model needs to be trained by many word sequences to calculate the probability of each sequence. The number of word sequences or the training corpus size used to train a model determines how much experience a model can gain about

the language and, more importantly, the knowledge incorporated in the language^[1].(ChatGPT in education: global reactions to AI innovations | Scientific Reports)In education, artificial intelligence technologies, including machine learning and natural language processing (NLP), offer immense potential to improve learning and support decision-making^[2].(Harry, 2023)To save time and money, artificial intelligence is gradually changing the way students learn and teachers teach. One such AI model is OpenAI's ChatGPT, which has been gaining popularity in the education sector for its ability to generate human-like responses and summarize complex information^[3]. (Full article: Discussing ChatGPT in education: A literature review and bibliometric analysis) This bibliometric study aims to comprehensively understand ChatGPT in education through author information and keyword data extracted from the Web of Science (WOS) database from 2022 to 2024. To explore the fundamental characteristics, research topics, and trends in topic evolution, it is essential to examine the current status and development of the research.

2. Literature Review

Since we are still in the early stages of discussing the use of OpenAI in education, particularly about ChatGPT, we would like to contribute by showing the trending topics related to this discussion. Rojas-Sánchez et al. (2022) Judging from the research literature on the application of ChatGPT in the education field, educators get a before-/after-class assistant whilst students attain a personalized learning experience^[4]. (K Aleksić-Maslačal.,2024) ChatGPT In the future of education, both teachers and pupils will be able to use it^[5]. (A de Bem Machado.,2024) ChatGPT is in constant development and students must have a good understanding of how to ethically use and manage Chat-GPT and critically assess its produced content. Educators must develop policies within the institution for using Chat-GPT in research work done by students^[6]. (Atanas Anov., 2023) Although Scholars have done a lot of research on the topic of ChatGPT in the field of education, but there are still a lot of research gaps in the visual analysis literature of ChatGPT and education. Rojas-Sánchez et al. (2022) and Pradana et al. (2023) started analyzing the development of using technology in education by conducting a bibliometric analysis, thus motivated me to conduct the same methodology for this article^[7].

3. Methodology

3.1 Data Sources

The article data used in this study will be based on the Social Science Citation Index (SSCI) and Science Citation Index (SCI) databases that have been included in the core collection of Web of science. The search title contains both "ChatGPT" and "Education", and the language is English. paper. The literature screening period was from November 30, 2022 to March 31, 2024, and 385 relevant documents were obtained. The actual time span of the obtained documents was from January 1, 2023 to March 31, 2024. Exported in RIS (other reference software) format, this study extracted the publication year, author country, keywords and other information of each document to analyze the basic characteristics, themes and theme evolution.

3.2 Data processing

This study uses the visualization tools VOS viewer and SATI to conduct visual analysis of highly cited author information and co-occurring keywords, and conducts data conversion and density based on VOS viewer, 1.6.20 (64-bit) and SATI, 4.0 (online version)^[8]. To explore visualization techniques, including network visualization, overlay visualization, and journal time frequency, it is essential to

analyze their characteristics. Finally, conclusions and reflections are drawn, and the research ideas are shown in Figure 1:

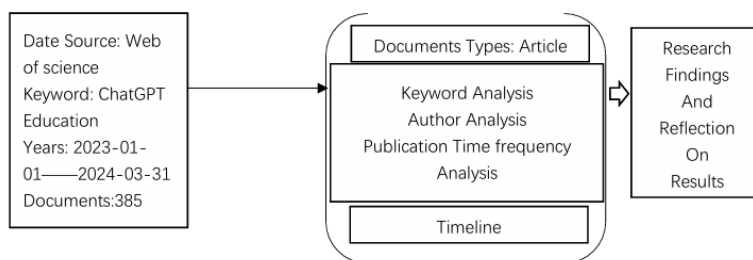


Figure 1: Research Approach Diagram.

4. General overview

4.1 Document frequency time

Judging from the frequency and time of documents, the number of documents published in journals studying "ChatGPT" and "Education" is on an overall upward trend: among them, JOURNAL OF CHEMICAL EDUCATION shows a downward trend in the figure, from 14 articles published in 2023 to 1 article in 2024, but currently 2024 is only counted to the third month, that is, at the beginning of 2024, scholars will continue to publish articles in this journal. The same goes for SUSTAINABILITY, which ranks second in the list of fluctuating changes. It dropped from 10 articles published in 2023 to 1 article in 2024. This is also inseparable from the short period at the beginning of the year. It can be expected that these two journals will grow strongly in the future. In addition to MEDICAL TEACHER, which published only 4 articles in 2023 to 6 articles in 2024, and DIGITAL HEALTH, which published 1 article in 2023 to 3 articles in 2024, only in the first three months of the year, other journals have The number of publications on this topic has shown a small increase. However, at the beginning of the year alone, so many journals have begun to publish articles on this topic. It can be seen that the number of publications on this topic in major journals will grow steadily. See Figure 2 and Figure 3:

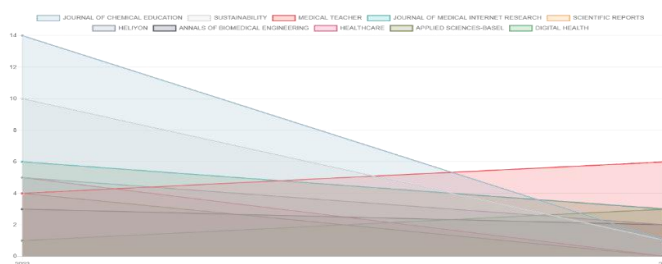


Figure 2: Document source frequency time series chart.

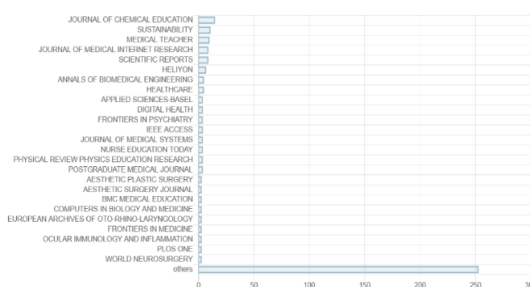


Figure 3: Literature source data distribution chart.

The fact that 2023 has become the starting point of research is mainly due to the fact that after ChatGPT was released on November 30, 2022, many scholars began to study the application of ChatGPT in various fields. Therefore, in 2023, "ChatGPT" and "Education" were included as one of the ChatGPT research categories. The project began to become a research hotspot for many scholars, and the number of journal articles began to increase. After 2024, thanks to the release and subsequent commercial use of ChatGPT4.0 in 2023, the number of publications in this field will still increase.

4.2 Distribution of author institutions

As can be seen from the figure below, the institutions with the largest number of publications are Nagoya Univ, Grad Sch Engr, Dept Mat Proc Engr (6 articles), Univ Queensland, Australian Inst Bioengn & Nanotechnol AIBN, Brisbane (5 articles) and King Saud Univ. Coll Sci, Chem Dept (4 articles). The countries where the top three institutions belong are: Japan, Australia and Saudi Arabia. Among them, Japan and Saudi Arabia ranked first and third respectively, and they are also Asian countries. Therefore, in the two years since the rise of ChatGPT, Asian scholars and research institutions have compared the publications of scholars and research institutions in other regions in the field of ChatGPT education. The contribution of the number of articles is relatively high. See Figure 4 and Figure 5:



Figure 4: Publishing House Statistical Chart1.

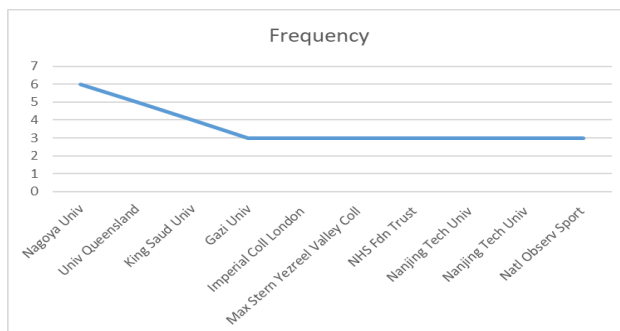


Figure 5: Publishing House Statistical Chart2.

4.3 Distribution of the number of articles published by authors

According to the number of publications, the top 10 authors are Yamauchi Yusuke, Chen Tingting, Chen Zhongwei, Elyoseph Zohar, Jiang Juncheng, Seth Ishith, Yu Yuan, Agrawal Rajdeep, Agrawal Rupesh, Cifuentes-Gonzalez Carlos. The countries where the top five authors belong are: Australia, China, China, Israel and the United States. It can be seen from this that after adding up the number of articles published by Chinese scholars in the second and third places, Chinese scholars are at the forefront of research in the field of ChatGPT education, followed by Australia, Israel and the United States. See Figure 6 and Figure 7:

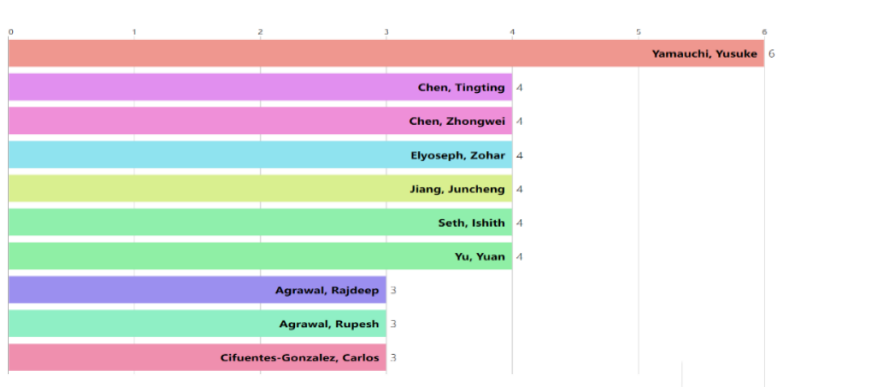


Figure 6: Statistical chart of the number of articles published by authors 1.

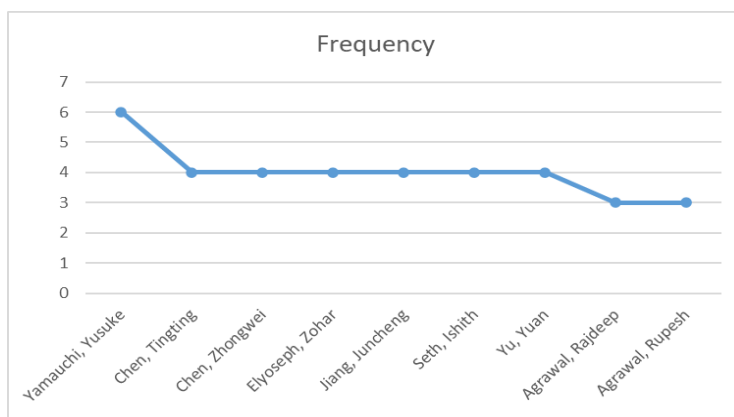


Figure 7: Statistical chart of the number of articles published by authors 2.

5. Research hotspots

5.1 The field of Artificial Intelligence research is gradually expanding internationally

Based on the keyword co-occurrence analysis of "ChatGPT" and "Education", this study obtained the hot spots of "ChatGPT" and "Education" and set the minimum number of occurrences of the topic keyword to 5 times. Among the 1517 keywords, a total of We screened out 55 subject keywords that matched the frequency of occurrence, and calculated the co-occurrence intensity of the subject keywords. According to the frequency of occurrence of the subject keywords, the 10 most frequent connection strength subject keywords were "ChatGPT", "Artificial Intelligence" ", "Natural Language Processing", "Medical Education", "Education", "Large Language Model", "Machine Learning", "Patient Education", "chatbot", "gpt-4". Among these 10 words, after excluding the subject words ChatGPT and Education that you searched for, you can then substitute the 11th and 12th words in order, which are "Readability" and "ethics" respectively. "Artificial Intelligence" appears frequently in the co-occurrence analysis, which occurs 161 times and the total link strength is 405, revealing that Artificial Intelligence is the highest hot spot in the research of "ChatGPT" and "Education". It can be seen that scholars' research topics are mainly concentrated on Artificial Intelligence field. See Figure 8 and Figure 9:

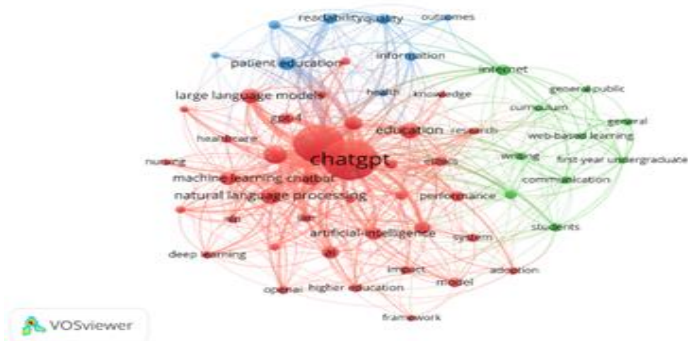


Figure 8: Keywords network visualization.

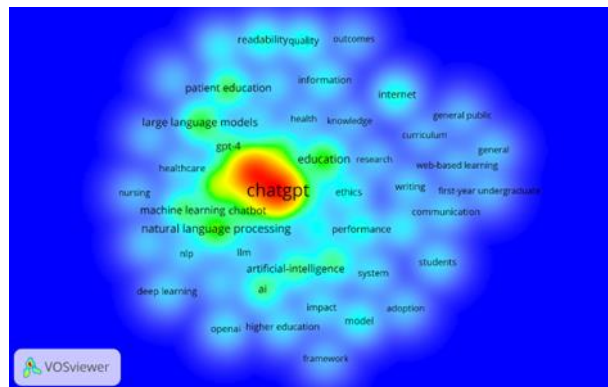


Figure 9: Keywords density visualization.

5.2 Analysis of the evolution of research theme development

According to the year distribution chart provided by Vos Viewer, we can see the hot topics in different periods. The darker the color, the longer the topic has been around, and the lighter the color, the newer the topic. After excluding the two subject words "ChatGPT" and "Education", the remaining keywords can be roughly divided into four categories: dark blue, green, light green and yellow. Dark blue represents the keywords concentrated in the right half of the keyword network, including: "General public", "Curriculum", "General", "Research", "Web based learning", "First year undergraduate", and key "Deep Learning" for the left half of the word network. The middle and lower parts of the keyword network are green keywords, including: "Higher Education", "AI", "System", "Students", "NLP", "performance", "ethics", "communication", and the upper right corner "Knowledge" and "Internet". Light green keywords are mainly concentrated in the middle core area of the keyword network, including: "Machine Learning Chatbot", "Nursing", "healthcare", and "Health". The yellow keywords representing the most recent relevant time period include: "Large Language Model", "Patient Education", "readability", "quality", "information", and "model" in the area below the keyword network. It can be found that the evolution trend of the topics in "ChatGPT" and "Education" is as follows: from the early scholars' research on ChatGPT school courses, to the research on the performance of artificial intelligence itself, to the research on ChatGPT medical education, and finally evolved into the research on ChatGPT Research on language modeling and patient education. See Figure 10:

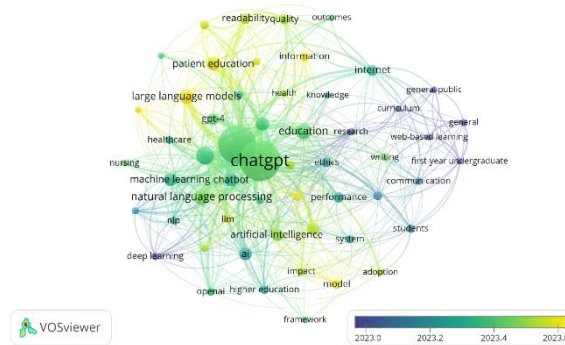


Figure 10: Keywords overlay visualization.

6. Conclusion

Since the release of ChatGPT on November 30, 2022, more and more scholars have begun to pay attention to and conduct research, resulting in rapid breakthroughs in various branches of research in this field, and the research on ChatGPT in the field of education has quickly become a research Hotspot. Author information and keywords are two important factors determined by visual analysis in this study. After eliminating all irrelevant or repeated data (such as repeated subject words) during statistics, 385 relevant documents were obtained by delimiting the screening time. Then export it through VOS viewer and SATI to obtain important frequency information data such as authors and keywords. This review provides researchers and educators with a detailed review of ChatGPT research trends and topics in the field of education as of March 31, 2024, which can help them continue to develop in this field, solve new research problems, and find Research partners.

References

- [1] Rojas-Sánchez, M., Palos-Sánchez, P., & Folgado-Fernández, J. A. (2022). Systematic literature review and bibliometric analysis on virtual reality and education. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-022-11167-5>.
- [2] Anov, A. (2023, December). From Wikipedia to Chat-Gpt Addressing the Issue of Ai in Education through Unesco 'S Guidelines. In *Conference Proceedings* (pp. 296-303).
- [3] Pradana, M., Elisa, H. P., & Syarifuddin, S. (2023). Discussing ChatGPT in education: A literature review and bibliometric analysis. *Cogent Education*, 10(2), 2243134.
- [4] Aleksić-Maslač, K., Borović, F., & Biočina, Z. (2024). Perception and Usage of Chat Gpt in the Education System. In *INTED2024 Proceedings* (pp. 1842-1848). IATED.
- [5] Harry, A. Role of AI in Education. *Interdisciplinary J. Hummanity Inj.* 2, 260–268 (2023).
- [6] Liu Qiyuan, Ye Ying (2012). Bibliographic information mining technology methods and implementation of software SATI - taking Chinese and foreign library and information science as an example. *Journal of Information Resources Management*, (01), 50-58.
- [7] A de Bem Machado, Pesqueira, A., & Sousa, M. J. (2024). A Review ChatGPT and the Future of Education. In *Digital Transformation in Higher Education Institutions* (pp. 155-176). Cham: Springer Nature Switzerland.
- [8] Full article: Discussing ChatGPT in education: A literature review and bibliometric analysis. <https://www.tandfonline.com/doi/full/10.1080/2331186X.2023.2243134>.