

Large Language Model an Emerging Technology in Medical Education: An Analysis of Scholarly Collaboration, Citation, and Vocabulary Co-occurrence

Shahnaz Khademizadeh^{1,2*}, Zeinab Jozi³, Hajar Beigi⁴

¹Visiting Faculty in Medical Library and Information Sciences Department, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran

²Associate Professor, Department of Knowledge and Information Science; Faculty of Educational Sciences and Psychology, Shahid Chamran University of Ahvaz; Ahvaz, Iran

³PhD Candidate in Knowledge and Information Science, Faculty of Educational Sciences and Psychology, Shahid Chamran University of Ahvaz; Ahvaz, Iran

⁴PhD Candidate in Knowledge and Information Science, Faculty of Educational Sciences and Psychology, Alzahra University, Tehran, Iran

Received: 2025 January 03

Revised: 2025 September 15

Accepted: 2025 October 12

Published online: 2025 October 12

***Corresponding author:**

Department of Knowledge and Information Science; Faculty of Educational Sciences and Psychology, Shahid Chamran University of Ahvaz; Ahvaz, Iran.

Email: s.khademi@scu.ac.ir

Citation:

Khademizadeh S, Jozi Z, Beigi H. Large Language Model an Emerging Technology in Medical Education: An Analysis of Scholarly Collaboration, Citation, and Vocabulary Co-occurrence. *Strides Dev Med Educ.* 2025 October; 23(1):e1504. doi:10.22062/sdme.2025.200769.1504

Abstract

Background: The emergence of large language models has led to significant progress in various fields. In recent years, researchers have had a special focus on its applications in medical education.

Objectives: This study explores the social and intellectual framework of research on large language models as a growing technology in medical education across the Web of Science database.

Methods: The present study is a quantitative study performed using scientometric techniques. The statistical population consisted of texts containing 643 documents from the year of this study. Following the Prisma guidelines, studies were screened based on their titles, abstracts, and keywords, leading to the selection of 233 documents for the research population. No restrictions (such as year, time, country) were applied to the search. Data analysis was undertaken using VOSviewer software.

Results: The study found that the United States was the most active country, generating 90 papers and receiving 537 citations. Author Lee, Hyunsoo, received 163 citations, with *JMIR Medical Education* being among the most frequent authors and journals with 30 papers and 352 citations. The keywords "artificial intelligence" with 144 repetitions, "ChatGPT" with 131 repetitions, "medical education" with 97 repetitions, and "large language models" with 89 repetitions make them highly frequent words. Four clusters were formed from the keywords obtained in the study".

Conclusion: The research in large language models and medical education is growing rapidly, with organizations revealing a strong interest in international cooperation. This field has great potential for development. Nevertheless, countries with more publications have not necessarily received more citations.

Keywords: Large Language Models; Artificial Intelligence; Medical Education; Scientometrics; Citation Analysis; Co-Occurrence

Background

Recent advances in artificial intelligence have enabled the development of more robust language models capable of generating fluent and coherent text. Among these models, "Large language models" stand out given their impressive size and ability to learn vast amounts of textual data (1). Large language models are semi-supervised and generative artificial language

models trained on massive amounts of textual data. They can contextualize the ordinal nature of words effectively in a sentence to predict the most reasonable response (2). Essentially, large language models employ deep neural networks—complex structures with multiple layers of statistical correlation, or "hidden layers," that facilitate complex and fine-grained relationships as well as advanced information abstraction (3).

The emergence of large language models can be traced back to n-grams models utilized in the 1980s (1, 4). Topic models such as hidden Dirichlet allocation came into existence during the 2000s (5). However, these language models were restricted by traditional statistical methods in natural language processing, which relied on word counts. As such, they suffered from limitations such as short-term memory, data fragmentation, difficulty in training, lack of semantic understanding, lack of generalizability, and low efficiency. In response to these challenges, OpenAI introduced the Transformer architecture in 2017. The core component of transformers is the attention mechanism. This mechanism enables the model to grasp the contextual connections between words (or other elements) in a sentence. Transformers apply multi-headed self-attention for analyzing the relationship between words in a sentence. Self-attention means that words pay attention to their relationship with other words, regardless of their relative or absolute position in a sequence (1, 6).

One of the AI tools developed based upon the TransqurMator model was GPT, which was introduced in 2018. In the subsequent years, various versions of GPT were released. For instance, GPT-2, released in 2019, garnered attention for its remarkable capability to generate coherent and relevant text on a wide range of topics (7). The introduction of GPT-3 by OpenAI in 2020 marked a significant milestone in the field. The performance of the GPT-3 model, which employs 175 billion training data parameters, is noteworthy for its more accurate responses compared to its predecessors (8). Since its introduction in November 2022, Chat GPT 3.5 has been equipped with Transformer architecture capabilities (9). Continuing the development of generations of large language models, the emergence of GPT-4 in 2023 indicated significant progress (10). This version (GPT-4) has the ability to process inputs beyond text, including images and data (11).

Recent developments in large language models, such as ChatGPT and its newer model, ChatGPT-4, have transformed teaching methods within the field of medical education. In traditional medical education, the primary teaching method involves lectures delivered at specific times and locations (12). With the introduction of large language models, especially GPT chat and its subsequent versions, it is anticipated that the usage of this technology in medical education would revolutionize teaching methods in this field. GPT chat

enhances students' understanding of medical knowledge by offering tailored and efficient learning experiences through simulated conversations, intelligent instruction, and automated question responses (13, 14). With the assistance of AI technologies, students are essentially placed in a simulated clinical situation. This simulation intends to mimic interactions between a student and a doctor, enabling the student to engage in critical thinking, make decisions, and engage in active learning (15, 16). For small group sessions, GPT-4 can facilitate discussions by creating thought-provoking questions, encouraging peer-to-peer interactions, and establishing an attractive and collaborative learning environment. To simulate a virtual patient, LLM can create realistic scenarios involving virtual patients, pose questions, interpret responses, and provide feedback (10, 17, 18).

Further, integrating GPT-4 large language and chat models into medical curriculum planning can assist the faculty in developing, updating, or revising medical curricula (16). Recent studies have indicated significant benefits of large language models in medical knowledge and reasoning; for example, in another study by Zei et al. (2023), they examined the potential of large language models such as ChatGPT, Bard, and Bing AI in medical education as well as clinical decision support for young physicians. Sivaner et al. (2022) explored the use of large language models in medical school curricula from the perspectives of researchers and medical students (19). Another study found that ChatGPT achieved a score close to the passing level on the United States Medical Licensing Examination (USMLE) (20).

Scientometrics refers to the quantitative and qualitative study of scientific production, aimed at extracting patterns, trends, and knowledge networks. In other words, it is the science of measuring science, capturing all quantitative methods and patterns for measuring the production of science and technology (21). Although many studies have explored the usage of large language models in medicine and health from a scientometric perspective (including Mayta-Tovalino et al. (2024) (22), Mousavi et al. (2024) (23), Liu et al. (2024) (11), and Wang et al. (24)), there has been no comprehensive analysis of the trends, gaps, and key topics in global research on medical education. It is critical for educators, physicians, and students to have a thorough understanding of the research landscape and emerging opportunities in medical education. Thus, this study aims to conduct a Scientometric analysis of studies

in this field, ascertaining the scientific standing of countries, influential researchers, collaborative networks, popular topics, and overlooked areas of study. The goal is to provide insights into future research directions for researchers and policymakers in the field of medical education.

Objectives

Considering the issues raised in the previous section, the objectives of this research are as follows:

- Examine the scientific status of countries and identify leading countries.
- Analyze scientific collaboration networks between countries.
- Identify the scientific status of active and highly cited authors.
- Recognize hot topics and inspect the time trend of topics in the use of large language models in medical education.

Methods

The current research applied a quantitative-descriptive approach, employing scientometric techniques. The method used in this study was documentary, with data extracted from the Web of Science database on 06/28/2024. For data extraction from the database, two strategies were implemented. The study population consisted of 643 articles. Data were consolidated through removing overlapping and unrelated cases, leading to the selection of 233 documents for analysis and addressing the research questions. To eliminate overlap, the first step was to download and merge the two Excel outputs. Next, the duplicate items were removed and the necessary data were selected. Ultimately, the selected items were extracted from the database in Plain Text format.

In the first search strategy, the following were combined using Boolean operators, yielding 460 documents: ("Artificial Intelligence*" OR "Machine Learning" OR "Neural Networks*" OR "Natural Language Processing*" OR "Computer Vision*" OR "Generative Adversarial Networks" OR "AI Ethics" OR "AI Explainability" OR "Autonomous Agents" OR "Robotics" OR "Cognitive Computing" OR "AI Algorithms" OR "AI Applications" OR "speech processing*" OR "Audio processing*" OR "Image Processing*" OR "Smarting*" OR "Emotion processing*" OR "Environmental intelligence*" OR "expert systems*" OR "chat bots*" OR "Computer Reasoning" OR "Knowledge Acquisition") AND ("Large language models" OR "LLM architectures*" OR "Zero-Shot

Learning" OR "Few-Shot Learning" OR "Transfer Learning" OR "Text Summarization*" OR "Machine Translation" OR "Question Answering" OR "Model Interpretability" OR "ChatGPT") AND ("Clinical Training" OR ("Medical education*" OR "Evidence-Based Medicine" OR "Personalized Learning" OR "Resource Optimization" OR "Skill Development" OR "Virtual Patient Simulations" OR "Medical Chatbots" OR "Academic & Research Integrity").

For the second strategy, keywords ("Large Language Model*" OR "chatgpt*" OR "LLM") and ("Medical Education*" OR "Medical Training*") were merged, resulting in 183 documents.

In this study, the results of two search strategies were analyzed, culminating in a total of 605 documents after merging and removing duplicates. Once the titles, abstracts, and keywords were reviewed, 327 documents were excluded, leaving a final selection of 233 documents (Figure 1).

In order to standardize keywords such as "medical education" and "medical training", "LLM" and "large language models", and "artificial intelligence" and "AI", a terminology was developed and applied in VOSviewer software.

Note that the first strategy aimed for comprehensive coverage, identifying studies that indirectly referenced large language models. The second strategy focused on directly retrieving studies associated with large language models and medical education. It should be noted that in our study, as in other scientometric studies, the threshold for creating scientific maps was determined based on the authors' opinions, which are presented below.

In this study, we utilized VOSviewer software to analyze scientific and citation networks of 54 countries, 101 authors (at least one article and 30 citations), 46 sources (at least one study and 10 citations), as well as word co-occurrence networks (with a frequency of at least 3). This analysis allowed us to identify 67 main keywords. Further, we examined the contribution of countries to the production and citation of documents using frequency and percentage frequency tests.

In this study, we ensured the quality of the selected studies through extracting data from the most prominent global database. Further, all selected articles had been published in reputable journals, with the researchers aiming to select only studies related to the application of large language models in medical education, rather than focusing on specific individuals

or organizations. Thus, there has been no bias in the selection of data.

Results

1- Country Performance

In order to ascertain the performance of different countries, it was determined that 53 countries were actively involved in producing work in this field. As displayed in [Figure 1](#), countries with larger nodes indicate higher levels of activity. The United States, with 90 documents, China with 26 documents, and England with 23 documents, are the top three countries in this field ([Figure 2](#)).

According to the image above, India, Germany, Australia, Canada, Italy, the United Arab Emirates, and Saudi Arabia are the top countries in large language model research in the field of medicine.

The United States, with 90 documents producing 19.04% and receiving 537 citations, claims nearly 39% of the total citations in this field, making it a prominent country in the subject area. Nevertheless, the status of document production and receiving citations for some countries differs from their citation share, as outlined in [Table 1](#). For example, Canada, with 11 documents, claims 11.06% of citations, while China, with 26 documents and a citation share of 8.58, and the United Kingdom, with 23 documents and a citation share of 4.72, do not have a favorable status.

Countries such as South Korea and India, though not among the top countries in document production, are in a good position in terms of receiving citations (9.01 and 3.00, respectively).

The trend of countries undertaking studies in the field of large language models in medical education was also analyzed based on the time period of the studies presented in [Figure 3](#). The Overlay Visualization map applies colors to reflect the average year of publication of the articles in which a term appeared; yellow represents words mainly employed in recent articles, while blue indicates words primarily utilized in older articles (25). If items have a score value, their color is determined by the score value. By default, the map color ranges from blue (lowest value) to green and then yellow (highest value) (26).

According to the image, countries in the blue region have an average publication year between 2022.5 and 2023.00, revealing the lowest average level. Countries in the green range from 2023.2 to 2023.4 are considered medium-level. Dark green countries fall between 2023.5 and 2023.8, signifying high-level countries. Lastly,

countries within the range of 2024.0 are emerging countries in this area.

2- Status of Cited Authors

In order to determine the citation status of authors, we found that 1,112 authors contributed to the studies. To identify highly cited authors, we selected authors whose works had at least 30 citations. As a result, we selected the names of 101 authors and generated maps of their citations. According to [Figure 4](#), the authors with the highest number of citations are located in the center and the hot area.

Author Density Map [Figure 4](#) indicates that author Lee, Hyunsu is the most cited author with 163 citations and no co-authors. The citation status of the other top 5 authors, including Antaki, Fares, Duval, Renaud, El-Khoury, Jonathan, Milad, Daniel, Milad, Daniel, who are affiliated with the University of Montreal and Canada, is presented in [Table 2](#). According to the citation density map, the most cited authors are located in the center of the map with the greatest heat.

3- The status of citations to sources that publish scientific productions

It was found that out of 133 sources in this field, only three works were presented at conferences. For this review, sources that received at least 10 citations were chosen to create a citation map. The map, depicted in [Figure 5](#), includes 46 sources.

According to [Table 3](#), JMIR Medical Education is the most produced and cited source. Anatomical Sciences Education Journal is another journal with only one published document in this field, while ranking second in terms of receiving citations, with 163 citations. Ophthalmology Science is the third publication with 13 citations.

Table 3. The most cited sources of publishing studies

Sources	Documents	Citations
JMIR medical education	30	352
Anatomical sciences education	3	163
Ophthalmology science	1	143
Cureus journal of medical science	26	127

4- Thematic Clusters

To investigate the main research topics in the field of large linguistic models and medical education, the frequency of vocabulary and hot topics in this area was examined. Words with a frequency of at least 3 were selected for this analysis. A total of 67 keywords were employed to form 4 subject clusters, which are illustrated in [Figure 6](#). Word co-occurrence is a highly

useful technique for identifying popular topics, trends, and the overall structure of scientific knowledge within a specific subject. Both word co-occurrence and scientific mapping serve as powerful tools for uncovering the conceptual framework of scientific fields (27, 28). These methods enable identifying thematic clusters, new concepts, and knowledge gaps by quantitatively examining scientific documents. The distance between specific keywords in co-occurrence maps reflects the degree of connection between them; the smaller the distance, the greater the connection between them. Further, more related keywords form a cluster that can be utilized to describe a core topic in research. These co-occurrence networks can be applied to analyze a complete set of topics (29). In this study, we applied content analysis to analyze clusters.

In this map, the keyword "artificial intelligence" appeared 144 times, followed by "ChatGPT" with a frequency of 131, "medical education" 97, "large language models" 89, and "natural language processing" 24, representing the most frequent words. Content analysis was applied to examine thematic clusters, with the results of the analysis for each cluster listed in Table 4.

Trends and Average Year of Occurrence of Subjects

Figure 7 exhibits the Overlay visualization. This map indicates the overlap of keywords over time using a color spectrum, with colors closer to yellow representing more recent document dates. The denser a cluster appears in this map, the more significant it is relative to other clusters (30). In simpler terms, the more often a keyword appears, the more yellow its color will be (26, 31). Keywords found in studies within the 2022 timeframe in cluster "two" such as "virtual reality," "simulation," "clinical training," "student," "surgical training," "curriculum," and "health care" reflect the emergence of research on the application of interactive artificial intelligence technologies in the medical field, including surgery and clinical medicine. It presents the trend of studies in the area being studied in clusters "one and three" for the beginning of 2023. Keywords such as "artificial intelligence," "chatGPT," "medical education," "natural language processing," "chatbots," "bard," "bing," machine learning, and "scientific integrity" indicate the emergence of research on the application of large language models in the field of medical education.

Further, keywords in clusters "one and four" for 2023 and 2024 (nodes in yellow) including ChatGPT 3.5, ChatGPT 4, Ethics, Bias, Explainable AI, Decision Making, Answer, Question, and Patient Education are

topics that have surfaced in the field of medical education in late 2023. These topics well indicate the advancement of large language models and their application in medical research, as well as the biases and ethical considerations surrounding the use of such technologies in medical education.

Discussion

The results demonstrate the growth and development of the field of large language models and the growing attention of the medical field to maximize the usage of this technology in education and the development of new methods in medical education.

The results indicate that the United States has had a significant impact on this field by generating 90 documents and receiving 537 citations, making it the leading country in the field of large language models and medical education. In addition to conducting extensive research in this field, the United States also claimed the largest share of citations and is ranked first. The research by Mousavi et al., (2024) (23), Liu et al. (2024) (11), and Raman et al. (2023) (32), also confirms the leadership of the United States in producing documents and receiving citations. This indicates that adequate funding results in high-quality research, and the United States, as the richest country in the world, continuously invests in advanced scientific topics.

A time trend analysis of countries' activities based on the average year revealed that there has been little research generated in this field since the beginning. However, many countries began entering research in this field in 2023. The start of research in 2023 is visible on the map in purple, with a small volume. By the middle of the year, countries marked in green emerged with a significant extent of research in this field. Research in the United States also grew along this period. These findings align with the findings of Liu et al. (2024) (11), who examined global trends and hot spots in medical research, identifying countries such as the United States as leaders in this field.

The results regarding highly cited authors revealed that authors such as Lee, Hyunsu, despite having written only one document in this field, have received the most citations. Further, a review of journals publishing scientific resources indicated that the JMIR Medical Education journal is the most productive and cited journal in this field, with 30 articles and 352 citations. Our study's results show minor overlap with the research of Mayta-Tovalino et al. (2024) (22), and Raman et al. (2023) (32). However, our study, along with

the two mentioned above, has focused on the scientists actively writing studies about large language models. This demonstrates that a scientific article will have a high impact if it contains suitable content and is written based upon the needs of the scientific community. A higher number of articles, while containing inappropriate content that does not fulfill the needs of the scientific community, may not necessarily have an impact on the research community.

From the cluster analysis, four clusters were created: "Application of artificial intelligence in clinical medical research", "Application of interactive technologies in surgical education", "Application of chatbots in medical education evaluation and conversations", and "Types of large language models and accuracy in medical research". Among these clusters, Cluster 1 is the central and strategic cluster that captures a combination of technological foundations, educational-clinical applications, and ethical concerns. These results indicate that these technologies have had a significant influence on optimizing medical decisions and ameliorating the quality of education. Nevertheless, challenges such as algorithmic bias, scientific integrity, and ethical issues of applying these technologies still persist as potential obstacles in this field.

For example, a study found that the use of large language models in the clinical field can result in accurate diagnosis when analyzing patient medical records (33). Further, Ravi et al. (2023) exhibited the impact of large language models on clinical decision-making in their research (2). Further, Liu et al. (2024) (11) conducted scient metrics research identifying clusters such as "Artificial Intelligence and Machine Learning", "Education and Training", "Healthcare Applications", and "Data Analytics and Technology" as the main clusters. These clusters somewhat overlap with the findings of this study.

The results also reveal a rapid expansion in the use of these technologies since 2022, particularly in medical education and related research. This growth necessitates attention to ethical considerations and optimization of usage methods. Topics in 2022 focused on simulation in medical education based on AI technology in surgery. In 2023, the focus shifted towards language models, natural language processing, and text generation tools for creating exam questions, ensuring scientific integrity. By mid-2023 and into early 2024, attention turned to new GPT chat models such as GPT 3.5 and GPT 4. Discussions centered on the ethics and biases of using these technologies, explainable AI, and preventing

scientific misconduct. These discussions aimed to enhance transparency and strengthen accountability in medical AI systems.

In medical education, explainable AI offers numerous capabilities. For instance, when AI provides an answer, it is critical for students, professors, or doctors to understand why the AI chose that answer. Explainable AI helps clarify the reasoning behind AI-generated answers, aiding in error identification and correction (34). In spite of the positive outcomes from research, the risks of using AI in medical education cannot be overlooked. A study by Ahn (2025) (35) reported that large language models sometimes provide misleading or incorrect information and have limitations. Another study by Weber et al. (2025) (36) indicated the shortcomings of large language models in solving clinical problems, highlighting their inflexibility compared to human doctors.

The limitations of this study included relying solely on a single database and not evaluating the quality of the included studies.

Conclusion

In general, with the growth of research in the field of LLMs and medical education, as well as the high willingness of organizations to collaborate internationally, it is clear that this field has great potential for development. The overall results of this study indicate that the usage of artificial intelligence technologies in medical education currently focuses on large language models, chatbots, and explainable artificial intelligence. These technologies have the potential to enhance medical education and clinical decision-making. However, given that some previous studies have revealed practical weaknesses of LLMs, validation of AI tools is essential for the full and responsible use of these technologies in medical education. Thus, in order to develop and improve the use of large language model technology in medical education, it is suggested that national and international cooperation be developed, especially with identified active countries. Qualitative research and case studies should be undertaken to examine the application of LLMs in real-world and educational environments. It is important to examine the evolution of thematic clusters in this field based on 5-year time intervals to identify temporal changes and developments in the field under study. In order to develop and boost capacity for the use of LLMs in medical education, policymakers are advised to establish clear guidelines for data protection and

prevention of misuse. The use of LLMs should be accompanied by explanations to earn the trust of faculty and students. Further, the integration of this technology into clinical skills training should be gradual, and empowerment courses should be conducted to acquaint faculty and students with its opportunities and constraints.

Acknowledgements: The authors would like to thank and acknowledge all those who assisted them in this study.

Conflict of interests: There is no conflict of interest.

Ethical approval: Not applicable.

Funding/Support: Not applicable.

References

- Nassiri K, Akhlofi MA. Recent advances in large language models for healthcare. *BioMedInformatics*. 2024;4(2):1097-143. doi: [10.3390/biomedinformatics4020062](https://doi.org/10.3390/biomedinformatics4020062).
- Ravi A, Neinstein A, Murray SG. Large language models and medical education: preparing for a rapid transformation in how trainees will learn to be doctors. *ATS Sch*. 2023 Jun 14;4(3):282-292. doi: [10.34197/ats-scholar.2023-0036PS](https://doi.org/10.34197/ats-scholar.2023-0036PS). [PMID: 37795112] [PMCID: [PMC10547030](https://pubmed.ncbi.nlm.nih.gov/PMC10547030/)]
- Safraneck CW, Sidamon-Eristoff AE, Gilson A, Chartash D. The role of large language models in medical education: applications and implications. *JMIR Med Educ*. 2023 Aug 14;9:e50945. doi: [10.2196/50945](https://doi.org/10.2196/50945). [PMID: 37578830] [PMCID: [PMC10463084](https://pubmed.ncbi.nlm.nih.gov/PMC10463084/)]
- Bahl L, Baker J, Cohen P, Jelinek F, Lewis B, Mercer R. Recognition of continuously read natural corpus. *IN-CASSP'78. IEEE International Conference on Acoustics, Speech, and Signal Processing 1978 Apr 10(3)*, 422-424. IEEE. doi: [10.1109/ICASSP.1978.1170402](https://doi.org/10.1109/ICASSP.1978.1170402).
- Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of Machine Learning Research*. 2003;3:993-1022.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Advances in Neural Information Processing Systems*. 2017;30:1-11. doi: [10.2196/46900](https://doi.org/10.2196/46900).
- Heilbron M, Armeni K, Schoffelen JM, Hagoort P, De Lange FP. A hierarchy of linguistic predictions during natural language comprehension. *PNAS*. 2022 Aug 9;119(32):e2201968119. doi: [10.1073/pnas.2201968119](https://doi.org/10.1073/pnas.2201968119).
- Shao CY, Li H, Liu XL, Li C, Yang LQ, Zhang YJ, et al. Appropriateness and comprehensiveness of using ChatGPT for perioperative patient education in thoracic surgery in different language contexts: survey study. *Interact J Med Res*. 2023 Aug 14;12:e46900. doi: [10.2196/46900](https://doi.org/10.2196/46900). [PMID: 37578819] [PMCID: [PMC10463083](https://pubmed.ncbi.nlm.nih.gov/PMC10463083/)]
- Kleinig O, Gao C, Bacchi S. This too shall pass: the performance of ChatGPT-3.5, ChatGPT-4 and New Bing in an Australian medical licensing examination. *Med J Aust*. 2023 Sep 4;219(5):237. doi: [10.5694/mja2.52061](https://doi.org/10.5694/mja2.52061). [PMID: 37528548]
- Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large language models in medical education: opportunities, challenges, and future directions. *JMIR Med Educ*. 2023 Jun 1;9:e48291. doi: [10.2196/48291](https://doi.org/10.2196/48291). [PMID: 37261894] [PMCID: [PMC10273039](https://pubmed.ncbi.nlm.nih.gov/PMC10273039/)]
- Liu L, Qu S, Zhao H, Kong L, Xie Z, Jiang Z, et al. Global trends and hotspots of ChatGPT in medical research: a bibliometric and visualized study. *Front Med (Lausanne)*. 2024 May 16;11:1406842. doi: [10.3389/fmed.2024.1406842](https://doi.org/10.3389/fmed.2024.1406842). [PMID: 38818395] [PMCID: [PMC11137200](https://pubmed.ncbi.nlm.nih.gov/PMC11137200/)]
- Li Z, Li F, Fu Q, Wang X, Liu H, Zhao Y, et al. Large language models and medical education: a paradigm shift in educator roles. *Smart Learning Environments*. 2024;11(1):26. doi: [10.1186/s40561-024-00313-w](https://doi.org/10.1186/s40561-024-00313-w).
- Wu Y, Zheng Y, Feng B, Yang Y, Kang K, Zhao A. Embracing ChatGPT for medical education: exploring its impact on doctors and medical students. *JMIR Med Educ*. 2024 Apr 10;10:e52483. doi: [10.2196/52483](https://doi.org/10.2196/52483). [PMID: 38598263] [PMCID: [PMC11043925](https://pubmed.ncbi.nlm.nih.gov/PMC11043925/)]
- Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: A conversation with ChatGPT and a call for papers. *JMIR Med Educ*. 2023 Mar 6;9:e46885. doi: [10.2196/46885](https://doi.org/10.2196/46885). [PMID: 36863937] [PMCID: [PMC10028514](https://pubmed.ncbi.nlm.nih.gov/PMC10028514/)]
- Khalpey Z, Kumar U, King N, Abraham A, Khalpey AH. Large language models take on cardiothoracic surgery: a comparative analysis of the performance of four models on American Board of Thoracic Surgery exam questions in 2023. *Cureus*. 2024 Jul 22;16(7):e65083. doi: [10.7759/cureus.65083](https://doi.org/10.7759/cureus.65083). [PMID: 39171020] [PMCID: [PMC11337141](https://pubmed.ncbi.nlm.nih.gov/PMC11337141/)]
- Kitamura FC. ChatGPT is shaping the future of medical writing but still requires human judgment. *Radiology*. 2023 Apr;307(2):e230171. doi: [10.1148/radiol.230171](https://doi.org/10.1148/radiol.230171). [PMID: 36728749]
- Morreel S, Mathysen D, Verhoeven V. Aye, AI! ChatGPT passes multiple-choice family medicine exam. *Med Teach*. 2023 Jun;45(6):665-6. doi: [10.1080/0142159X.2023.2187684](https://doi.org/10.1080/0142159X.2023.2187684). [PMID: 36905610]
- Lee H. The rise of ChatGPT: Exploring its potential in medical education. *Anat Sci Educ*. 2024 Jul-Aug;17(5):926-931. doi: [10.1002/ase.2270](https://doi.org/10.1002/ase.2270). [PMID: 36916887]
- Civaner MM, Uncu Y, Bulut F, Chalil EG, Tatli A. Artificial intelligence in medical education: a cross-sectional needs assessment. *BMC Med Educ*. 2022 Nov 9;22(1):772. doi: [10.1186/s12909-022-03852-3](https://doi.org/10.1186/s12909-022-03852-3). [PMID: 36352431] [PMCID: [PMC9646274](https://pubmed.ncbi.nlm.nih.gov/PMC9646274/)]
- Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023 Feb 9;2(2):e0000198. doi: [10.1371/journal.pdig.0000198](https://doi.org/10.1371/journal.pdig.0000198). [PMID: 36812645] [PMCID: [PMC9931230](https://pubmed.ncbi.nlm.nih.gov/PMC9931230/)]
- De Bellis N. Bibliometrics and citation analysis: from the science citation index to cybermetrics. *Journal of the American Society for Information Science and Technology*; 2010;61(1):205-7. doi: [10.1002/asi.21181](https://doi.org/10.1002/asi.21181).
- Mayta-Tovalino F, Espinoza-Carhuancho F, Alvez-Temoche D, Mauricio-Vilchez C, Munive-Degregori A, Barja-Ore J. Scientometric analysis on the use of ChatGPT, artificial intelligence, or intelligent conversational agent in the role of medical training. *Medical Education*. 2024 Mar 1;25(2):100873. doi: [10.1016/j.edumed.2023.100873](https://doi.org/10.1016/j.edumed.2023.100873).
- Mosavi A, Imre F, Hung VT. ChatGPT and large language models in healthcare: a bibliometrics analysis and review. *Proceedings of the IEEE 11th International Conference on Computational Cybernetics*

- and Cyber-Medical Systems; 2024 Apr 4-6; Hanoi, Vietnam. 2024: 350-4.
24. Wang S, Yang L, Li M, Zhang X, Tai X. Medical Education and Artificial Intelligence: Web of Science–Based Bibliometric Analysis (2013-2022). *JMIR Med Educ.* 2024 Oct 10:10:e51411. doi: [10.2196/51411](https://doi.org/10.2196/51411). [PMID: [39388721](https://pubmed.ncbi.nlm.nih.gov/39388721/)] [PMCID: [PMC11486481](https://pubmed.ncbi.nlm.nih.gov/PMC11486481/)]
 25. VOSviewer workbook. VOSviewer: A Guide for Research Visualization. [cited 2023 Jul 8]. Available from: https://www.griffith.edu.au/__data/assets/pdf_file/0030/1905483/VOSviewer-guide-2023.pdf?utm_source=chatgpt.com
 26. Van Eck NJ, Waltman L. VOSviewer: Visualizing scientific landscapes. Version 1.6.20. Leiden: Centre for Science and Technology Studies, Leiden University. [cited 2023 Oct 31]. Available from: https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.20.pdf.
 27. Van Eck N, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics.* 2010 Aug;84(2):523-38. doi: [10.1007/s11192-009-0146-3](https://doi.org/10.1007/s11192-009-0146-3). [PMID: [20585380](https://pubmed.ncbi.nlm.nih.gov/20585380/)] [PMCID: [PMC2883932](https://pubmed.ncbi.nlm.nih.gov/PMC2883932/)]
 28. Chan S. Knowledge Management in Smart Organizations. In: *Knowledge Management. In: Mezgar I. Integration of ICT in smart organizations.* Hershey, Pennsylvania: IGI Global Scientific Publishing; 2008: 3385-408.
 29. Liu Q, Ali NL, Lee HY. Applying VOSviewer in a bibliometric review on English language teacher education research: An analysis of narratives, networks and numbers. *Cogent Education.* 2025;12(1):2449728. doi: [10.1080/2331186X.2025.2449728](https://doi.org/10.1080/2331186X.2025.2449728).
 30. Hosseini E, Ghaebi A, Baradar R. Bibliometrics and Mapping of Co-words in the Field of Linked Data. *Scientometrics Research Journal,* 2021; 7((Issue 1, spring & summer)): 91-116. doi: [10.22070/rsci.2020.4904.1333](https://doi.org/10.22070/rsci.2020.4904.1333).
 31. Fallah Barzeghar M, Khalili A. Bibliometric Analysis and Systematic Review of Global Theoretical Context in the Field of Urban Health. *Journal of Iranian Architecture & Urbanism(JIAU).* 2022;13(2):351-72. [In Persian]
 32. Raman R, Lathabhai H, Diwakar S, Nedungadi P. Early research trends on ChatGPT: insights from altmetrics and science mapping analysis. *International Journal of Emerging Technologies in Learning (IJET).* 2023;18(19):13-31. doi:[10.3991/ijet.v18i19.41793](https://doi.org/10.3991/ijet.v18i19.41793).
 33. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. *NPJ Digit Med.* 2022 Dec 26;5(1):194. doi: [10.1038/s41746-022-00742-2](https://doi.org/10.1038/s41746-022-00742-2). [PMID: [36572766](https://pubmed.ncbi.nlm.nih.gov/36572766/)] [PMCID: [PMC9792464](https://pubmed.ncbi.nlm.nih.gov/PMC9792464/)]
 34. Hildt E. What Is the Role of Explainability in Medical Artificial Intelligence? A Case-Based Approach. *Bioengineering (Basel).* 2025 Apr 2;12(4):375. doi: [10.3390/bioengineering12040375](https://doi.org/10.3390/bioengineering12040375). [PMID: [40281735](https://pubmed.ncbi.nlm.nih.gov/40281735/)] [PMCID: [PMC12025101](https://pubmed.ncbi.nlm.nih.gov/PMC12025101/)]
 35. Ahn S. The impending impacts of large language models on medical education. *Korean J Med Educ.* 2023 Mar;35(1): 103-107. doi: [10.3946/kjme.2023.253](https://doi.org/10.3946/kjme.2023.253). [PMID: [36858381](https://pubmed.ncbi.nlm.nih.gov/36858381/)] [PMCID: [PMC10020064](https://pubmed.ncbi.nlm.nih.gov/PMC10020064/)]
 36. Weber MT, Noll R, Marchl A, Facchinello C, Grünewaldt A, Hügel C, et al. MedBot vs RealDoc: efficacy of large language modeling in physician-patient communication for rare diseases. *Journal of the American Medical Informatics Association.* 2025 May;32(5):775-83. doi: [10.1093/jamia/ocaf034](https://doi.org/10.1093/jamia/ocaf034).

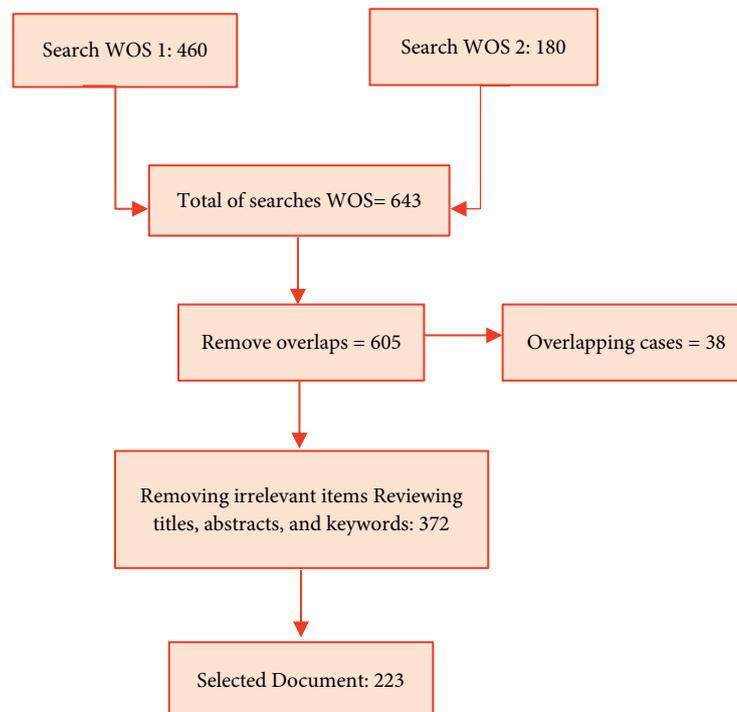


Figure 1. Study Selection Process

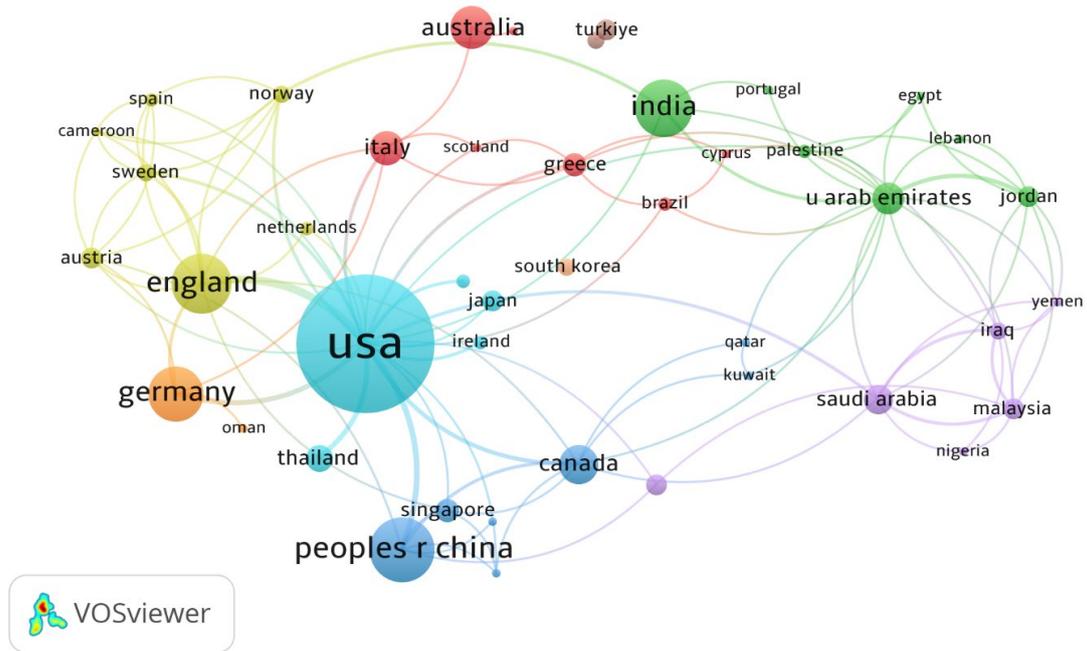


Figure 2. Scientific Map of the Activity of the Countries of the World in the Research of Language Models

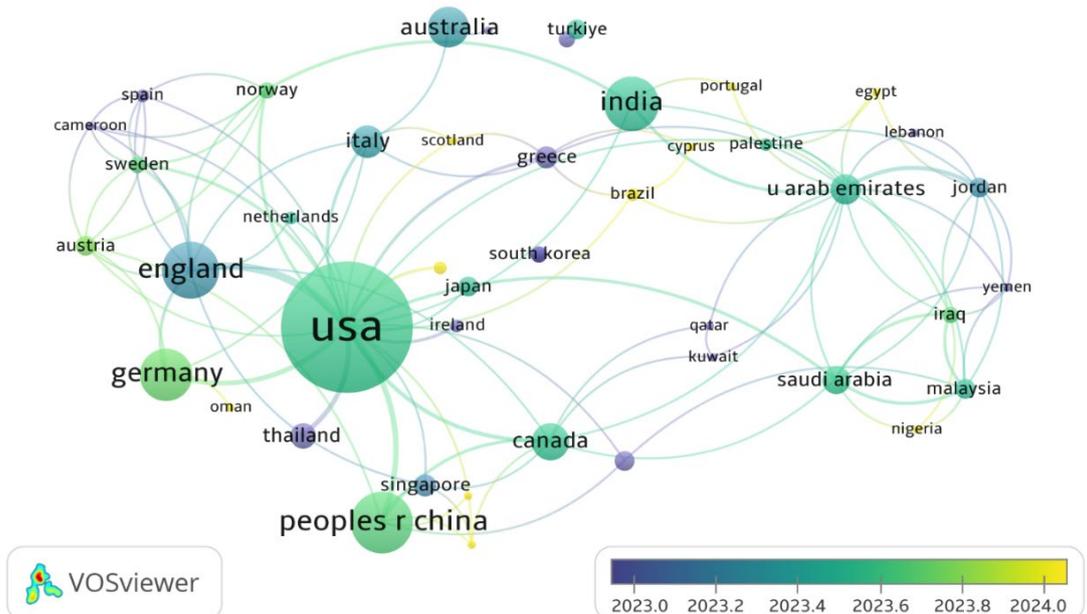


Figure 3. The Performance of Countries Based on Trends and Time Periods



Figure 4. Density Map of Highly Cited Authors in Studies

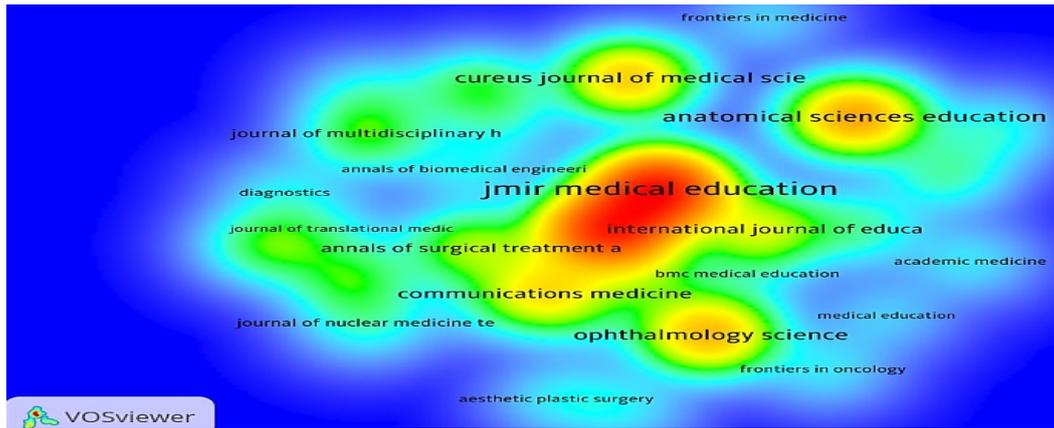


Figure 5. Density Map of Highly Cited Sources that Publish Studies

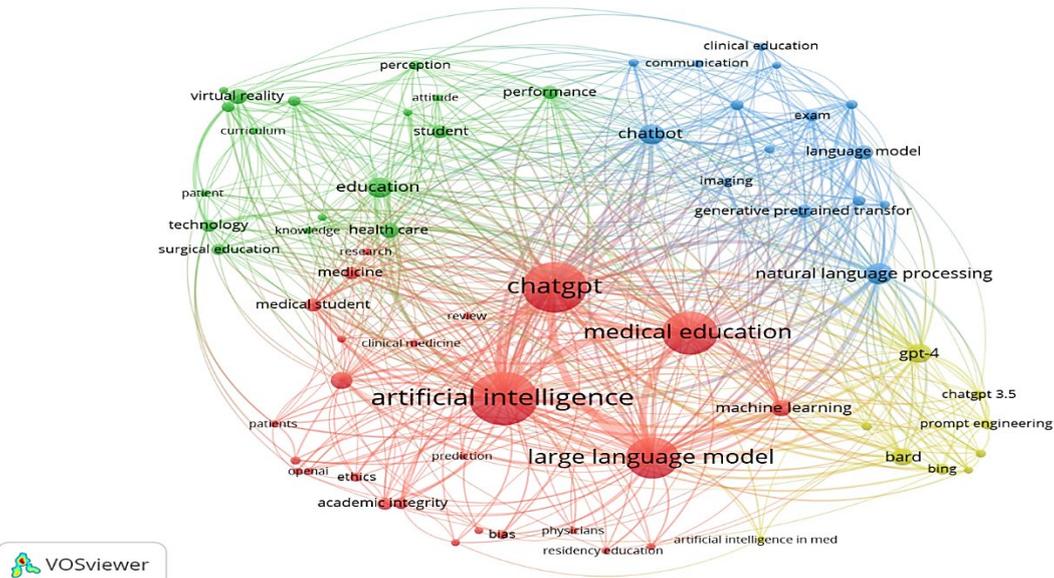


Figure 6. Thematic Clusters of Keywords of Studies in the Field of Large Language Models and Medical Education

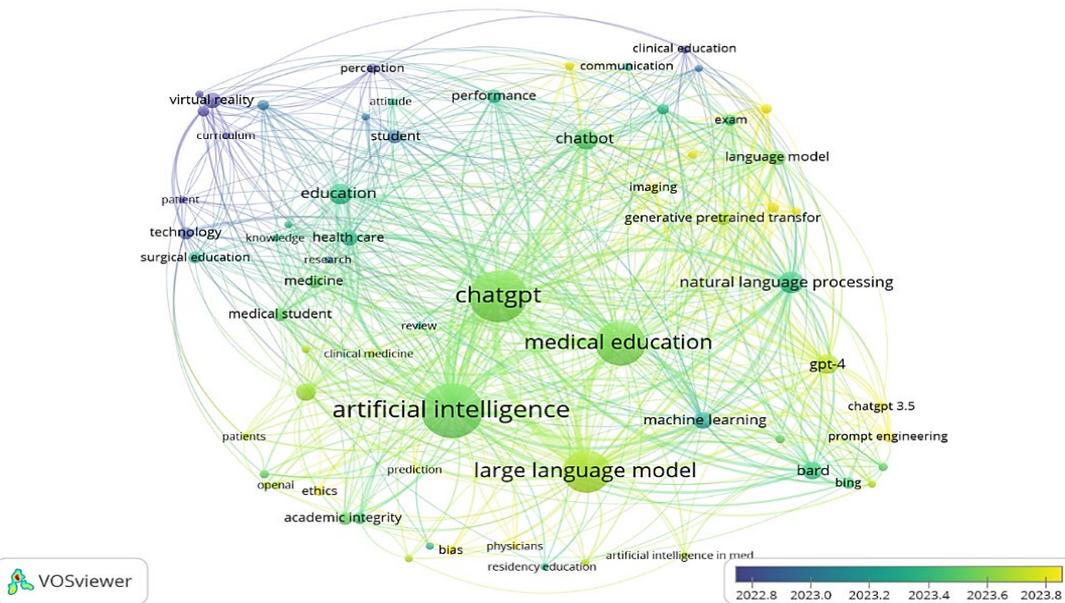


Figure 7. Vocabulary Overlay Map of Studies in the Field of Large Language Models and Medical Education Based on the Average of Normalized Citations

Table 1. Statistics of the Top Countries in Writing Research on Large Language Models

Rank	Country	Documents	Percentage Documents	Rank	Country	Citations	Percentage Citations
1	United States	90	19/04	1	United States	537	38/63
2	China	26	9/97	2	Canada	300	11/16
3	England	23	8/01	3	Germany	241	9/87
4	India	21	7/64	4	South Korea	230	9/01
5	Germany	20	5/15	5	China	155	8/58
6	Australia	13	4/72	6	Australia	142	5/58
7	Canada	11	4/42	7	England	133	4/72
8	Italy	9	4/39	8	India	132	3/86
9	United Arab Emirates	8	4/22	9	United Arab Emirates	127	3/43
10	Saudi Arabia	7	2/53	10	Kuwait	76	3/00

Table 2. Most Cited Authors in Published Studies

Authors	Documents	Citations	Country	Organization
Lee, Hyunsu	1	163	South Korea	Keimyung Univ
Antaki, Fares	2	143	Canada	University of Montreal
Duval, Renaud	1	143	Canada	University of Montreal
El -Khoury, Jonathan	1	143	Canada	University of Montreal
Milad, Daniel	1	143	Canada	University of Montreal
Touma, Samir	1	143	Canada	University of Montreal

Table 4. Nouns are the Most Important Words in Thematic Cluster Analysis Studies of Large Language Models in the Field of Medical Education

Cluster	Important keyword	Analysis
Cluster 1: Application of artificial intelligence in clinical medical research 	Artificial intelligence, large language model, machine learning, chatgpt, clinical medicine, Medical Education, Clinical Decision Making, Residency Education, Integrity, Bias, Ethics, Plagiarism	This cluster focuses on integrating artificial intelligence, large language models, and machine learning with educational and clinical applications in the field of medical education, while also considering ethical and scientific implications.
Cluster 2: Application of interactive technologies in surgical education 	Augmented reality, simulation, surgical education, curriculum, student, performance, patient	This cluster focuses on educating medical students and surgeons using advanced technologies. It examines the role of the curriculum in integrating hands-on training with technology, emphasizing learning, surgical simulation, performance, and attitude.
Cluster 3: Application of chatbots in medical education evaluation and conversation 	Chatbot, conversational agent, clinical education, generative pretrained transformer, multiple-choice question, exam, natural language processing, communication, reasoning,	This cluster explores the application of large language models and chatbots to enhance learning, clinical reasoning, and interactive teaching of medical students through natural language processing and exam simulation.
Cluster 4: Types of large language models and accuracy in medical research 	Artificial intelligence in medicine, accuracy, Bing, Bard, ChatGPT 3.5, GPT-4, prompt engineering, Explainable AI	This cluster explores the role of advanced language models (such as ChatGPT 3.5, GPT-4, Bard, and Bing) and explainable AI in accuracy, prompt engineering, and medical education management to enhance the quality of medical student learning.