Visual Analysis of Research Hotspots and Frontiers in Domestic and International Medical Large Language Models Based on CiteSpace: Postprint

Authors: Niu Ben, Zhu Xiaoqian, Yang Chen, Liang Wannian, Liu Jue, Liang Wannian, Liu Jue

Date: 2024-09-24T00:00:00+00:00

Abstract

Background: Owing to their formidable language processing capabilities and extensive application potential, large language models, exemplified by ChatGPT, have spearheaded a new trend in natural language processing within the medical domain. Objective: This study employs bibliometric analysis to uncover research hotspots, thematic distributions, and future developmental trajectories of medical large language models since 2017. Methods: Relevant literature on medical large language models published between January 2017 and June 2024 was systematically collected and screened from Web of Science, CNKI, Wanfang Data Knowledge Service Platform, and VIP databases. CiteSpace software was utilized to extract thematic keywords and other information from the literature, enabling analysis and comparison of the evolution, hotspots, and trends in domestic versus international research. Results: A total of 1,071 relevant articles were included. The findings indicate that international research concentrates on the application of artificial intelligence, large language models, deep learning, knowledge graphs, and related technologies in medicine, whereas domestic research remains relatively scarce, with emphasis on constructing Chinese medical question-answering systems and addressing unstructured medical data processing. Conclusion: It is imperative to deepen medical data mining, expand multi-scenario applications, and draw upon international experience in fine-tuning and application evaluation of large language models to foster the advancement of medical large language model technology and its clinical implementation in China.

Full Text

Evolution and Trends of Domestic and International Research Hotspots in the Field of Large Language Models in Medicine Based on CiteSpace

Ben Niu[1], **Xiaoqian Zhu**[2], **Chen Yang**[3], **Wannian Liang**[4], ***Jue Liu***[5]

[1]Hospital Management Institute, Shenzhen University, Shenzhen 518060, China
[2]Greater Bay Area International Institute for Innovation, Shenzhen University, Shenzhen 518060, China
[3]College of Management, Shenzhen University, Shenzhen 518060, China
[4]Vanke School of Public Health, Tsinghua University, Beijing 100084, China
[5]School of Public Health, Peking University, Beijing 100091, China

Corresponding authors: Wannian Liang, Professor/Doctoral supervisor; E-mail: liangwn@tsinghua.edu.cn
Jue Liu, Research Professor/Doctoral supervisor; E-mail: jueliu@bjmu.edu.cn

Abstract

Background: With advanced language processing capabilities and broad application potential, large language models (LLMs) such as ChatGPT are driving a new wave of natural language processing in the medical field. **Objective:** This study aims to identify research hotspots, topic distribution, and future development trends of medical LLMs through bibliometric analysis. **Methods:** A systematic search was conducted across Web of Science, CNKI, Wanfang Data, and VIP databases for literature on medical LLMs published between January 2017 and June 2024. CiteSpace software was used to extract thematic keywords and other information from the literature to analyze and compare the evolution, hotspots, and trends of domestic and international research. **Results:** A total of 1,071 relevant papers were included. The results revealed that international research primarily focuses on applying artificial intelligence, LLMs, deep learning, and knowledge graphs in medicine, while domestic research is relatively limited, concentrating on developing Chinese medical question-answering systems and addressing unstructured medical data problems. **Conclusion:** It is recommended to deepen medical data mining, expand multi-scenario applications, and leverage international experiences in fine-tuning and application evaluation of LLMs to promote the development of medical LLM technology and its application in the medical field in China.

Keywords: Delivery of health care; Healthcare; Large language models; Bibliometric analysis; CiteSpace; Artificial intelligence

---

Healthcare is a knowledge- and data-intensive industry. The rich medical data resources provide extensive possibilities for the application of large language models in medicine. Medical large language models (LLMs) refer to generative language models trained on large-scale medical data using deep learning and natural language processing technologies, capable of understanding and processing massive multimodal medical data including text and images [1]. Through natural language processing and deep learning, LLMs can not only understand and generate natural language but also support medical decision-making, optimize patient care, and drive innovative applications in healthcare.

Natural language processing has a long history of application in the medical field. Traditional medical language models, including statistical language models such as n-gram and neural language models such as word2vec, were widely used in the analysis of medical texts and other data. With Google' s introduction of the bidirectional Transformer-based pre-trained language model in 2018, research on medical language models began to shift toward the exploration and application of pre-trained language models [2]. In 2022, OpenAI launched ChatGPT, which, due to its outstanding natural language generation capabilities, sparked widespread global attention toward large language models. Research on LLMs has experienced rapid growth since the introduction of ChatGPT, and domestic and international research on medical language models has consequently shifted from pre-trained language models to LLMs.

Understanding the hotspots and trends of LLMs and related research is crucial for the future development of LLMs in China and for promoting the application of large language models in the medical field. Therefore, this study aims to systematically review the current state of domestic and international LLM research from 2017 to the present and analyze its research hotspots to provide directional insights for future technological development and clinical applications in this field.

1.1 Data Sources and Search Methods

We searched Web of Science, CNKI, Wanfang Data Knowledge Service Platform, and VIP databases from inception to June 28, 2024. Detailed English and Chinese search terms are provided in Table 1 . NoteExpress 4.0 software was used for literature management, and CiteSpace software was employed for analysis.

1.2 Research Methods

This study adopted bibliometric research methods to analyze the evolution of research hotspots and frontier trends of domestic and international LLMs. Inclusion criteria were: (1) relevant to LLM themes; (2) research articles or review papers; and (3) literature in Chinese or English. Exclusion criteria were: (1) completely unrelated to LLM themes or content; (2) duplicate publications; and (3) literature with unknown author or institutional information. Two researchers independently screened literature according to the inclusion and exclusion criteria; disagreements were resolved by a third researcher or through collective discussion. The initial search yielded 1,340 articles, and after screening titles, abstracts, and full texts, 1,071 articles were finally included.

1.3 Statistical Analysis

This study used CiteSpace 6.3.R1 software to set parameters and process literature data. The time slice parameter was set to 2017–2024 with a time span of 1 year, and the parameter k was set to 25 for publication volume analysis, keyword co-occurrence analysis, keyword clustering analysis, and keyword frontier analysis. Visualization analysis was performed and keyword burst tables were constructed to explore research frontiers and development trends, indicating the development direction of domestic LLM research. Microsoft Office Excel 2019 was used for statistical analysis of publication years, volumes, and keyword frequencies.

2.1 Annual Publication Volume Analysis

Figure 2 [Figure 2: see original paper] shows the number of publications in the LLM field domestically and internationally. As of June 2024, the total number of publications was 1,071, including 960 English articles and 111 Chinese articles. Since the introduction of ChatGPT, global research interest in large language models in the medical field has surged, showing explosive growth. Before November 2022, although some studies involved large-scale pre-trained language models, these models had significantly smaller parameter scales compared to current mainstream LLMs. Therefore, the period from 2023 to the present represents the rapid development stage of LLM research. As shown in Figure 2, the number of domestic and international LLM publications has shown a rapid growth trend since 2022. English literature in this field began in June 2023, while Chinese literature began in November 2023, both reaching their peak in 2024, with total publications of 27 and 616, respectively. Scholarly research on LLMs has gradually gained momentum.

2.2.1 Keyword Co-occurrence Analysis

Keywords are core vocabulary used to express the main content of literature, and co-occurrence analysis of keywords is an important basis for identifying disciplinary research hotspots and development trends. CiteSpace was used for keyword co-occurrence network analysis (see Figures 3a [Figure 3: see original paper] and 3b). The top ten high-frequency keywords in domestic and international LLM research are shown in Table 2 . In literature retrieved from the WOS database, main keywords such as artificial intelligence, large language models, large language model, natural language processing, machine learning, medical education, generative ai, "chatgpt" , language model, and deep learning constitute research hotspots in the international LLM field. In contrast, in literature retrieved from Chinese databases including CNKI, Wanfang Data, and VIP, main keywords such as deep learning, artificial intelligence, knowledge graph, entity recognition, electronic medical records, evidence-based medicine, medical question answering, literature classification, machine learning, and medical domain constitute research hotspots in the domestic LLM field.

2.2.2 Keyword Clustering Analysis

Based on keyword co-occurrence networks, keyword clustering further deepens thematic relationships by grouping closely related keywords into different categories to visually display various research directions in the field, allowing exploration of research hotspots through clustering themes. The LLR algorithm was used for keyword clustering domestically and internationally, yielding keyword clustering maps shown in Figures 4a [Figure 4: see original paper] and 4b. The clustering modularity value (Q value) was 0.7705 > 0.3000, indicating effective clustering, and the clustering average silhouette value (S value) was 0.9256 > 0.5000, indicating high clustering consistency (Figure 4a). The Q value was 0.8470 > 0.3000, indicating effective clustering, and the S value was 0.9625 > 0.5000, indicating high clustering consistency, demonstrating that the keyword clustering map is significant and structurally reasonable (Figure 4b). The maps show that there are 14 main clusters in international research in this field and 5 main clusters in domestic research. The main keyword clusters domestically and internationally are shown in Table 3 . Domestic research focuses on entity recognition and medical question-answering systems, electronic medical records and pattern recognition, and the integration and application of knowledge graphs. International research focuses on the application and progress of electronic health records, evidence-based medicine and medical decision support systems, considerations of medical ethics, and diversified research on large language models.

2.3 Research Frontier Analysis

To understand research frontiers and predict hotspots and future trends in the field, keyword burst analysis, as a visualization method for rapidly increasing co-occurrence frequency, has been applied by many scholars. This study can intuitively reflect the changing patterns of research hotspots in the field through the burst strength of literature keywords. Table 4 shows the top 20 keywords with strong burst strength in English literature. Higher burst strength better reflects research frontiers. In the burst bars, red represents the time period when the burst word was active. The analysis reveals that "chat generative pre-trained transformer," "transfer learning," "efficacy," "prevalence," "ai chatbot," "cancer," and "conversational agents"are prominent keywords in recent years, representing important trends in international LLMs. Table 5 shows the top 20 keywords with strong burst strength in Chinese literature. Higher burst strength better reflects research frontiers. In the burst bars, red represents the time period when the burst word was active. The analysis reveals that "artificial intelligence,""medical domain," "drug discovery," and "Chinese medicine" are prominent keywords in recent years, representing important trends and developments in domestic LLMs.

3.1 Evolution Analysis of Domestic and International LLM Research

Based on the annual publication volume in the retrieval databases, the study shows that LLM research has generally maintained a continuous growth trend. The development of medical language models can be roughly divided into three stages: traditional medical language models (before 2018), medical pre-trained language models (2019–2022), and LLMs (2023 to present). Before November 2022, although some studies involved large-scale pre-trained language models, these models did not meet the parameter scale standards of later LLMs. The year 2023 marked the beginning of systematic international research evaluating models like ChatGPT for their applications in the medical field [3]. In contrast, domestic research started later, with studies since November 2023 mainly exploring the feasibility of LLMs in clinical medicine and their potential for application in clinical consultation systems. The reason for the explosive growth of research concentrated in the past two years is the advancement of large language model technology internationally, whose powerful understanding capabilities can process large amounts of medical data, making medical data resources no longer idle [4], coupled with factors such as population aging and increasing medical demands, which together have promoted the rapid growth of domestic LLMs in recent years.

3.2 Analysis of Domestic and International LLM Research Hotspots

Based on keyword clustering analysis, combined with Table 3 and Figures 4a and 4b, this study reorganizes and classifies common research directions such as artificial intelligence and deep learning domestically and internationally, and sorts out differences in research directions between domestic and international studies. According to the relevance of directions, the research status was reclassified and deeply analyzed.

3.2.1 Domestic Research Hotspot Analysis

(1) Entity Recognition Promotes Performance Improvement in Medical Question Answering. Medical question-answering systems serve as important tools for assisting diagnosis, enabling rapid responses to queries from patients or healthcare personnel and improving the efficiency of medical services [5]. However, the complexity and diversity of Chinese medical terminology make it more difficult for systems to understand the true intent of patient queries, limiting the development of medical question-answering systems. To overcome these obstacles, Qiao Kai et al. [6] proposed a question-answering matching model based on an attention mechanism to improve the performance of question-answering matching models. Wang Runzhou et al. [7] proposed a medical knowledge graph question-answering framework based on hybrid dynamic masking and multi-strategy fusion to enhance entity recognition technology, solving the problem of unclear recognition in medical question-answering systems and promoting performance improvement in medical question-answering.

(2) Pattern Recognition Promotes Data Structure Optimization of Electronic Medical Records. As the core of China' s medical informatization, electronic medical records are an essential data source for training LLMs. However, the complexity of their unstructured data poses challenges for training and applying large language models, and traditional data warehousing strategies can no longer meet the needs of big data analysis [8-9], greatly affecting the effectiveness of medical data mining and leaving pre-trained LLMs lacking sufficient, authentic medical information. Machine learning can solve unstructured data problems. Liang Lirong et al. [10] used machine learning algorithms to establish a CCRFs model, which not only optimized the training process of LLMs but also enhanced the model' s application potential in real medical scenarios.

(3) Knowledge Graphs Promote the Expansion of Precision Medicine Applicability. Knowledge graphs contain rich knowledge semantics of entities and relationships in corresponding domains, providing important references for responses in domain dialogues [11], which plays a crucial role in improving the quality and accuracy of dialogue system responses. In this regard, Lü Xueqiang et al. [11] proposed incorporating knowledge semantic graph information into dialogue systems based on medical knowledge graphs, using the TransE model to integrate existing medical graph data into multi-turn dialogues, and demonstrated through experiments that this method can meet the rigor requirements of dialogue system responses. To address the current lack of high-quality question-answering data in the Chinese medical domain, domestic scholars have combined medical knowledge graphs with BERT models [6], which not only enhances the model' s understanding of medical terminology but also improves its ability to handle complex medical consultations [12]. Integrating knowledge graphs can expand the applicability of LLM dialogue systems and optimize their ability to handle complex medical consultations and provide precise medical advice, which can improve the model' s medical professionalism and deepen its understanding of medical terminology and concepts. Therefore, this fusion strategy is gradually becoming key to promoting the development of LLMs toward higher-level specialization. In the retrieved literature, it was found that traditional language model research mostly focuses on entity recognition, electronic medical record information extraction, and pre-trained language models mostly focus on knowledge graph fusion research. These research hotspots are all basic work for strengthening the application of knowledge graphs in LLMs and optimizing medical question-answering systems.

3.2.2 International Research Hotspot Analysis

(1) Application and Progress of Electronic Health Records (EHR). As a comprehensive patient health information database, EHR records not only clinical data but also multidimensional information including patients' lifestyles and social environments. Compared with domestic electronic medical records that focus on recording medical services and treatment processes within single medical institutions, resulting in relatively low information interoperability, international EHRs are more advanced in information completeness and data sharing due to their adoption of international standards such as HL7 and FHIR. International scholars have addressed the de-identification of code-mixed text in EHR through training and fine-tuning methods for PLMs and LLMs [13]. Moreover, international research has effectively solved the identification problem of code-mixed text in EHR and promoted the automation and intelligent development of EHR by training and fine-tuning pre-trained models such as KLUERoBERTa.

(2) Evidence-Based Medicine Promotes Clinical Decision Support Systems. In the field of evidence-based medicine, international research emphasizes the application of large language models in healthcare and clinical decision support systems. Eppler et al. [14] investigated the potential uses of the large language model ChatGPT in urology by collecting opinions from urologists worldwide. Pool et al. [15] evaluated the application and usage potential of large language models and generative AI in telemedicine. At the same time, international research also focuses on the ability of large language models to evaluate drug efficacy and support medical evidence. For example, Moskatel et al. [16] evaluated the accuracy and reliability of ChatGPT-3.5 in evidence-based medicine by querying drug efficacy, while Gwon et al. [17] considered evidence-based medicine as a theoretical foundation for clinical support decision-making. Although studies have found that large language models still face challenges in accuracy and reliability, international research has deeper applications in evidence-based medicine and clinical support decision-making compared to domestic research, becoming one of the research hotspots.

(3) Consideration of Medical Ethics. International research comprehensively considers medical ethics issues when exploring the application of large language models in the medical field, including informed consent, bioethics, and privacy ethics. The United States' HIPAA law sets standards for the application of models through strict regulations on medical data, while in some other countries, the enforcement and standards of data privacy vary, posing ethical challenges for international cooperation. Allen et al. [18] discussed the ethical implications of delegating consent to LLMs in medicine and pointed out their potential advantages in certain clinical situations. Chen et al. [19] assessed the accuracy of ChatGPT-3.5 in answering bioethics questions, while Balas et al. [20] evaluated GPT-4' s performance in responding to complex medical ethics cases. These studies indicate that although large language models still have limitations in understanding ethical dilemmas, improving medical ethics can enhance their reliability in clinical applications.

(4) Diversified Research on Large Language Models. International research has not only focused on ChatGPT and Google Bard but also prioritized Llama 2 as a research focus. For example, Bak et al. [21] evaluated the performance of different large language models in identifying motivational states.

Many scholars have used the Llama 2 model as a base model for fine-tuning to develop new methods. For instance, Wang et al. [22] used domain-specific corpora to fine-tune Llama 2 and developed a new method for rare disease concept normalization. Additionally, international research emphasizes the application of large language models in generating text summaries, such as Goswami et al.' s [23] research on Llama 2 and QLoRA, demonstrating the potential to maintain data quality while reducing memory usage. Therefore, Llama 2 has become one of the hotspots in international LLM research as a base model. Overall, in the retrieved literature, popular schools of international research mainly include pre-trained language models and LLMs. Most pre-trained language model research focuses on optimizing models for electronic health record information extraction, while most LLM research focuses on applications in evidence-based medicine and clinical decision support systems, considerations of medical ethics, and model optimization.

3.3 Frontier Analysis of Domestic and International LLM Research

Based on keyword burst analysis, the concentrated emergence of burst words in 2023-2024 reflects the increasing activity in this field in the past two years. This trend is attributed to the rapid development of technology, especially in processing complex medical texts and multimodal data. These models meet the market demand for efficient medical solutions, and their application potential has been recognized. However, due to differences in research paths between domestic and international studies, future research trends in this field are expected to diverge.

3.3.1 International Research Frontiers

(1) Research on Technology and Method Applications is the Technical Foundation for LLM Development. As shown in Table 4, "Chat Generative Pre-trained Transformer," "Transfer Learning," and "Conversational Agents" are key technological methods for LLMs that are gradually becoming research hotspots. Scholars are committed to developing models capable of understanding and generating natural language texts and exploring their diverse applications in the medical field. For example, pre-training and generation technologies have been used to build models that can provide continuous clinical decision support [1]. Scholars are also exploring methods to combine LLMs with medical guidelines to solve specific medical problems [24]. This indicates that research on LLM technology methods and their application in the medical field will be an important direction for future research.

(2) Research on Application Effects is the Results-Oriented Focus of LLMs in the Medical Field. As the application of large language models in the medical field becomes increasingly widespread, their efficacy and prevalence have become the focus of academic attention. Scholars are exploring the performance and application potential of these models in different medical scenarios. For example, by designing identification protocols for treatment plans for lung cancer patients, scholars have evaluated the performance of large language models in clinical decision support [25]. Additionally, Li et al. [26] filled gaps in mental health assessment by synthesizing evidence and evaluating the effectiveness of large language models. These research results indicate that future research will pay more attention to the application effects and general applicability of large language models in the medical field and their potential role in improving clinical decision support and patient care quality. Therefore, for international research, the pre-trained language model school will tend to combine with LLMs to optimize and build models suitable for clinical decision support systems, while the LLM school will tend to continue exploring the performance and potential of large language models in medical field applications.

3.3.2 Domestic Research Frontiers

(1) "Artificial Intelligence" Has Broad Application Prospects in the Medical Field, Particularly Showing Initial Effectiveness in Clinical Decision Support Systems and Medical Image Analysis. Chinese scholars mostly focus on the possibility of applying artificial intelligence in the medical field and discuss the development trends and future prospects of large language models in review form. For example, the application prospects, challenges, and countermeasure analysis of ChatGPT in the nursing field provide theoretical foundations and practical directions for the application of artificial intelligence in nursing education, research, and clinical nursing [27-28]. Some scholars have used ChatGPT and GPT-like technologies to conduct case demonstrations from five aspects: early prevention, disease diagnosis, treatment methods, prognosis and nursing, and image presentation, to obtain the application prospects of large models in the medical field [29-30]. These studies not only enrich the theoretical foundation of artificial intelligence in the medical field but also point out directions for future empirical research and application practice.

(2) LLMs Will Continue to Drive Challenges and Innovation in Chinese Medical Information Processing. The hotspot "Chinese Medicine" focuses on identifying Chinese electronic medical records and solving the problem of unstructured Chinese medical records. Domestic scholars mainly combine traditional deep learning methods with BERT models to address the unstructured problem of Chinese medical record data [31-32]. Given the professionalism of Chinese medical data, constructing knowledge graphs containing sufficient professional terminology is particularly important [33-34], providing necessary database support for Chinese medical question-answering systems. With the growth of Chinese medical data volume and the advancement of language model technology, developing LLMs capable of efficiently processing and analyzing Chinese medical information has become increasingly important. Furthermore, constructing professional terminology knowledge graphs will significantly improve the model' s understanding of medical texts and the accuracy of answering medical-related questions, which is crucial for improving clinical decision support and patient care quality. Therefore, for domestic research, traditional language model schools and pre-trained language model schools will tend to combine machine learning methods with pre-trained language models for optimization to promote Chinese medical record information processing, while LLM schools will tend to explore the feasibility of large language model applications in medical fields such as clinical decision support systems and medical imaging, driving the application of large language models in different medical scenarios.

4 Summary and Outlook

Research on large language models in the medical field is built upon traditional language models and pre-trained language models. With the maturation of generative large language model technologies such as ChatGPT, the number of relevant research papers has increased significantly in recent years. After in-depth analysis of the development history of medical language models, this study found that domestic research focuses mainly on knowledge graph construction, entity recognition, electronic medical record applications, and medical question-answering system development, particularly emphasizing the application prospects of knowledge graphs and artificial intelligence in medicine. International research, in contrast, focuses more on the development of large language models, improvement of natural language processing technologies, and innovative applications of generative artificial intelligence, especially in fine-tuning large language models to adapt to diverse medical scenarios and improve their application accuracy.

This study is limited by the finite databases supported by CiteSpace, which may result in incomplete literature retrieval. Nevertheless, this study also considered literature data from 2017–2022. Although early relevant research did not directly target LLMs, these studies are highly relevant to current LLMs, and therefore, this study included a small amount of research on traditional medical language models and pre-trained language models to provide a more comprehensive examination of LLM evolution.

Overall, compared with international research, China still has room for improvement in the application and fine-tuning of large language models. In the future, LLM research in China should continue to deepen the mining of local medical data and the construction of knowledge graphs, which are key to improving model performance. At the same time, China should learn from international advanced experiences in fine-tuning large language models and multi-scenario application evaluation. Furthermore, China should promote the application of models in clinical decision support systems to achieve comprehensive optimization and innovation of technical methods. In-depth exploration of these research directions will enhance China' s international competitiveness in the LLM field and improve our contribution and influence in the global medical artificial intelligence field.

References

[1] THIRUNAVUKARASU A J, TING D S J, ELANGOVAN K, et al. Large language models in medicine[J]. Nat Med, 2023, 29(8): 1930-1940. DOI:10.1038/s41591-023-02448-8.

[2] SI Y Q, WANG J Q, XU H, et al. Enhancing clinical concept extraction with contextual embeddings[J]. J Am Med Inform Assoc, 2019, 26(11): 1297-1304. DOI:10.1093/jamia/ocz096.

[3] MOOR M, BANERJEE O, ABAD Z S H, et al. Foundation models for generalist medical artificial intelligence[J]. Nature, 2023, 616(7956): 259-265. DOI:10.1038/s41586-023-05881-4.

[4] SHAH N H, ENTWISTLE D, PFEFFER M A. Creation and adoption of large language models in medicine[J]. JAMA, 2023, 330(9): 866-869. DOI:10.1001/jama.2023.14217.

[5] GUAN L B, LI S. Chinese medical question answering matching model integrating multi-granularity semantic information and knowledge graph[J]. Computer Engineering and Applications, 2024, 60(14): 152-161. DOI:10.3778/j.issn.1002-8331.2305-0453.

[6] QIAO K, CHEN K J, CHEN J Q. Chinese medical question answering matching method based on knowledge graph and keyword attention mechanism[J]. Pattern Recognition and Artificial Intelligence, 2021, 34(8): 733-741. DOI:10.16451/j.cnki.issn1003-6059.202108006.

[7] WANG R Z, ZHANG X S. Medical knowledge graph question answering based on hybrid dynamic masking and multi-strategy fusion[J/OL]. Computer Science and Exploration, 1-20. DOI:10.3969/j.issn.1671-3982.2021.11.001.

[8] WU Z Y, BAI K L, YANG L R, et al. A survey of electronic medical record text mining[J]. Journal of Computer Research and Development, 2021, 58(3): 513-527. DOI:10.7544/issn1000-1239.2021.20200402.

[9] MU Y, CHEN H B. Machine learning promotes the development of laboratory medicine[J]. Chinese Journal of Laboratory Medicine, 2018, 41(8): 627-630. DOI:10.3760/cma.j.issn.1009-9158.2018.08.013.

[10] LIANG L R, LI C W, SHEN Y, et al. Electronic medical record text information extraction based on cascaded conditional random field model[J]. Computer Applications and Software, 2019, 36(10): 47-54, 112. DOI:10.3969/j.issn.1000-386x.2019.10.009.

[11] LÜ X Q, ZHANG J, MU T Y, et al. Medical domain dialogue system embedded with knowledge semantics[J]. Computer Engineering and Design, 2023, 44(12): 3794-3799. DOI:10.16208/j.issn1000-7024.2023.12.037.

[12] ABDALLA M, ABDALLA M, RUDZICZ F, et al. Using word embeddings to improve the privacy of clinical notes[J]. J Am Med Inform Assoc, 2020, 27(6): 901-907. DOI:10.1093/jamia/ocaa038.

[13] LEE Y Q, CHEN C T, CHEN C C, et al. Unlocking the secrets behind advanced artificial intelligence language models in deidentifying chinese-english mixed clinical text: development and validation study[J]. J Med Internet Res, 2024, 26: e48443. DOI:10.2196/48443.

[14] EPPLER M, GANJAVI C, RAMACCIOTTI L S, et al. Awareness and use of ChatGPT and large language models: a prospective cross-sectional global survey in urology[J]. Eur Urol, 2024, 85(2): 146-153. DOI:10.1016/j.eururo.2023.10.014.

[15] POOL J, INDULSKA M, SADIQ S. Large language models and generative AI in telehealth: a responsible use lens[J]. J Am Med Inform Assoc, 2024, 31(9): 2125-2136. DOI:10.1093/jamia/ocae035.

[16] MOSKATEL L S, ZHANG N S. The utility of ChatGPT in the assessment of literature on the prevention of migraine: an observational, qualitative study[J]. Front Neurol, 2023, 14: 1225223. DOI:10.3389/fneur.2023.1225223.

[17] GWON Y N, KIM J H, CHUNG H S, et al. The use of generative AI for scientific literature searches for systematic reviews: ChatGPT and microsoft Bing AI performance evaluation[J]. JMIR Med Inform, 2024, 12: e51187. DOI:10.2196/51187.

[18] ALLEN J W, EARP B D, KOPLIN J, et al. Consent-GPT: is it ethical to delegate procedural consent to conversational AI?[J]. J Med Ethics, 2024, 50(2): 77-83. DOI:10.1136/jme-2023-109366.

[19] CHEN J, CADIENTE A, KASSELMAN L J, et al. Assessing the performance of ChatGPT in bioethics: a large language model' s moral compass in medicine[J]. J Med Ethics, 2024, 50(2): 97-101. DOI:10.1136/jme-2023-109366.

[20] BALAS M, WADDEN J J, HÉBERT P C, et al. Exploring the potential utility of AI large language models for medical ethics: an expert panel evaluation of GPT-4[J]. J Med Ethics, 2024, 50(2): 90-96. DOI:10.1136/jme-2023-109549.

[21] BAK M, CHIN J. The potential and limitations of large language models in identification of the states of motivations for facilitating health behavior change[J]. J Am Med Inform Assoc, 2024, 31(9): 2047-2053. DOI:10.1093/jamia/ocae057.

[22] WANG A, LIU C, YANG J Y, et al. Fine-tuning large language models for rare disease concept normalization[J]. J Am Med Inform Assoc, 2024, 31(9): 2076-2083. DOI:10.1093/jamia/ocae133.

[23] GOSWAMI J, PRAJAPATI K K, SAHA A, et al. Parameter-efficient fine-tuning large language model approach for hospital discharge paper summarization[J]. Appl Soft Comput, 2024, 157: 111531. DOI:10.1016/j.asoc.2024.111531.

[24] LI Z Z, ZHANG J F, ZHOU W, et al. GPT-agents based on medical guidelines can improve the responsiveness and explainability of outcomes

for traumatic brain injury rehabilitation[J]. Sci Rep, 2024, 14(1): 7626. DOI:10.1038/s41598-024-58514-9.

[25] BENARY M, WANG X D, SCHMIDT M, et al. Leveraging large language models for decision support in personalized oncology[J]. JAMA Netw Open, 2023, 6(11): e2343689. DOI:10.1001/jamanetworkopen.2023.43689.

[26] LI H, ZHANG R W, LEE Y C, et al. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being[J]. NPJ Digit Med, 2023, 6(1): 236. DOI:10.1038/s41746-023-00979-5.

[27] LUO H Y, XU M, ZENG C R, et al. Prospects and challenges of ChatGPT application in the nursing field[J]. Chinese Journal of Nursing Education, 2023, 20(12): 1520-1523. DOI:10.3761/j.issn.1672-9234.2023.12.019.

[28] SHI Q H, ZHANG Z F, HU B, et al. Research progress on the application of deep learning and artificial intelligence in the diagnosis and treatment of cervical and lumbar degenerative diseases[J]. Medical Journal of Chinese People's Liberation Army, 2021, 46(10): 1034-1039. DOI:10.11855/j.issn.0577-7402.2021.10.13.

[29] XU L L, HONG Y, YE Y. Application prospect discussion of ChatGPT and GPT-like technologies in medical informatics[J]. Information Studies: Theory & Application, 2023, 46(6): 38-42. DOI:10.16353/j.cnki.1000-7490.2023.06.006.

[30] DING W J. "Assistant" or "Killer": risk analysis of seeking medical advice with ChatGPT[J]. Medicine and Philosophy, 2023, 44(23): 22-25. DOI:10.12014/j.issn.1002-0772.2023.23.05.

[31] ZHANG Y R, TANG Y, ZHU M, et al. Construction of an early recognition model for acute respiratory infectious diseases based on natural language processing and deep learning[J]. Chinese Journal of Nosocomiology, 2024(15): 2394-2400. DOI:11.3436/r.20240722.1302.062.

[32] LIANG W T, ZHU Y H, ZHAN F, et al. Medical named entity recognition based on deep learning multi-model fusion[J]. Computer Applications and Software, 2022, 39(10): 162-168, 229. DOI:10.3969/j.issn.1000-386x.2022.10.025.

[33] XI Y J, LI M, DENG Y S, et al. Research on knowledge graph construction for question-answering content in Chinese online medical communities[J]. Library and Information Service, 2024, 68(4): 124-136. DOI:10.13266/j.issn.0252-3116.2023.24.010.

[34] XU R Q. Medical question-answering system integrating knowledge graph and semantic matching[J]. Modern Electronics Technique, 2024, 47(8): 49-54. DOI:10.16652/j.issn.1004-373x.2024.08.008.

Author Contributions: Ben Niu was responsible for determining the research topic, literature screening and verification, manuscript writing and revision, and funding support. Xiaoqian Zhu and Chen Yang were responsible for literature retrieval, screening, deduplication, and data analysis. Jue Liu was responsible for figure and table production, manuscript revision and quality control, and

funding support. Wannian Liang was responsible for research topic determination, quality control, guidance, and revision, and takes overall responsibility for the article.

Conflict of Interest: The authors declare no conflict of interest.

(Received: August 15, 2024; Revised: September 10, 2024)

Note: Figure translations are in progress. See original paper for figures.