# Bibliometric analysis and knowledge graph visualization of artificial intelligence used in medical education: a systematic review

Yu-qing Zhang[1,2,3,§], Xuan Wang[1,3,§], Ya-ping He[4], Hong-guo Rong[1,3], Yan-yan Meng[5,✉]

[1] Institute for Excellence in Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, 100029, China
[2] The First Clinical Medical College, Beijing University of Chinese Medicine, Beijing, 100700, China
[3] School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing, 102488, China
[4] Qingdao Traditional Chinese Medicine Hospital, Qingdao Hiser Hospital Affiliated of Qingdao University, Qingdao, 266033, China
[5] Beijing Research Institute of Chinese Medicine, Beijing University of Chinese Medicine, Beijing, 102488, China

**Received:** February 1, 2026; **Revised:** February 21, 2026; **Accepted:** February 28, 2026

## ABSTRACT

**Introduction** Artificial intelligence (AI) is penetrating medical education and its research status needs to be reviewed and suitably integrated. This study aimed to analyze the current status, hotspots, and trends of AI in medical education. **Methods** Data were retrieved from the Web of Science Core Collection (WOSCC), PubMed, China National Knowledge Infrastructure (CNKI), Wanfang, Chinese Scientific Journal Database (VIP) databases. Using CiteSpace, relevant data were extracted to analyze burst citation detection for keywords, clustering, keyword timelines, and keyword emergence. VOSviewer was used to generate visual collaborative network graphs for the keyword timelines. **Results** This study identified 2437 published in English and 326 Chinese papers. Cluster analysis identified three core themes: technology-driven medical educational innovation, the intelligent transformation of clinical ability training, and the construction of ethical risk and governance systems. In western research, the focus has been on specific technologies, such as deep learning and robotic surgery, with concentrated ethical discussions emerging in 2022. However, Chinese research is driven by the New Medicine policy, which focuses on the macro-integration of AI and the medical education system, with ethical review mechanisms appearing in clusters in 2024. In general, both Chinese and English articles considered the improvement of practical ability using augmented reality and virtual reality technology and the paradigm shift from knowledge transfer to training in higher-order thinking. **Conclusion** While specific technologies and earlier ethical debates have been prioritized in English research, policy-driven systemic integration and later governance frameworks have been emphasized in Chinese research, both coverage on enhancing practical abilities and shifting educational paradigms. In future research, AI should be further integrated with medical education values and student-centered innovations should be promoted.

## KEYWORDS

artificial intelligence, medical education, visualization, research hotspots, comparative knowledge integration framework, knowledge mapping technology

## CITE THIS ARTICLE

Zhang YQ, Wang X, He YP, Rong HG, Meng YY. Bibliometric analysis and knowledge graph visualization of artificial intelligence used in medical education: a systematic review. Evidence-Based Chinese Medicine and Technology Assessment, 2026, 2: 9570027. https://doi.org/10.26599/eCMTA.2026.9570027

## 1. Introduction

As artificial intelligence (AI) involves multiple domains including healthcare, transportation, as well as social interactions, and will become more and more involved in our lives and profoundly change the way people live. AI reconstructs the knowledge-transfer paradigm and skill-training mode of traditional medical education through virtual patient simulations, adaptive learning systems, intelligent diagnosis assistive devices, and other means. At the basic teaching level, intelligent learning guidance systems based on knowledge graphs improved the knowledge retrieval efficiency of medical students through the triplet-structured expression of disease–symptom–treatment plans.[1] In clinical skills training, virtual reality (VR) systems have shown significant improvements in training outcomes.[2] In educational administration, natural language processing (NLP) technology has been used to achieve a normalized application of curriculum evaluation sentiment analysis.[3] AI is increasingly being used in medical education; therefore, learning to apply it more effectively in this field is urgent.

However, despite the rapid increase in literature indicating active inquiry, the current research field is marked by significant fragmentation, application gaps, and a lag in theoretical exploration, as the low aggregation of knowledge production suggests a lack of cohesive dialogue and cumulative building. In this study, a comparative knowledge-integration framework was adopted to address this gap. This framework posits that meaningful integration requires not just synthesis but also a structured comparison of distinct knowledge streams—in this case, a technology-driven, globally focused English literature and a policy-oriented, system-integration-focused Chinese literature. This comparative lens enables moving beyond descriptive mapping to explanatory analysis by asking how the intellectual structures, research priorities, and evolutionary paths of AI in medical education differ and converge across these two major academic spheres, and what this reveals about the overarching dynamics and theoretical underpinnings of the field.

Therefore, the aim of this study was not only to “analyze the status, hotspots, and trends,” but also to theoretically delineate and compare the distinct knowledge architectures and developmental trajectories within the Chinese and English academic literature on AI in medical education. Specifically, we sought to identify and contrast the core thematic clusters and conceptual networks defining each corpus, analyze the divergent temporal emergence of key issues, and synthesize a holistic view of the intellectual landscape of the field by integrating these comparative insights, thereby clarifying the complementary roles of context-specific policy drives and global technological dis-

✉ Corresponding author.
*E-mail address:* 13261267976myy21_2001@163.com (YY. Meng)
§ These authors contributed equally to this work.

© 2026 Beijing University of Chinese Medicine. Production and hosting by Tsinghua University Press. This is an open access article under the CC-BY license (http://creativecommons.org/licenses/by/4.0/).

course in shaping evolution in this field.[4]

Using bibliometrics and knowledge mapping technology, a global cognitive map of AI applied in medical education over the past 20 years, from 2004 to 2025, was drawn. Summarizing the published research with VOSviewer and CiteSpace was necessary. In this study, the high-frequency co-occurrence keyword network, the keyword cluster network in the discipline knowledge system, and the highlights of keywords were analyzed. The synchronous emergence characteristics of knowledge hotspots in discipline research frontiers, such as technology ethics and human-computer collaborative teaching, were revealed, including the evolution of diachronic research fields, which include digital twin technology and generative AI.

## 2. Methods

### 2.1. Sources

The following databases were data sources for this study: the Web of Science Core Collection (WOSCC), PubMed, China National Knowledge Infrastructure (CNKI), Wanfang Database, and Chinese Scientific Journal Database (VIP).

All potentially relevant articles in the database, including author names, titles, year, keywords, abstracts, and journals, were downloaded and converted into NoteExpress 3.5.0 (Beijing Aegean, Beijing, China) to check for duplicates. The built-in deduplication function was employed to identify and merge duplicate records. The process involved automatically matching entries based on key fields, such as title, author(s), publication year, and source journal. Subsequently, a manual review was conducted to verify the software suggestions and resolve ambiguous cases. This two-step procedure ensured the integrity of the final dataset by eliminating redundant entries before bibliometric analysis, guaranteeing that each publication was counted only once in the subsequent analytical metrics.

### 2.2. Search strategy

We downloaded data from the WOSCC, PubMed, CNKI, Wanfang, and VIP on March 20, 2025. Considering the centralization and accuracy of the literature, the following search strategy was used: In the English database, the search function for PubMed was set as: (((((Artificial Intelligence [MeSH Terms]) OR (Intelligence, Artificial [Title/Abstract])) OR (Computer Reasoning [Title/Abstract])) OR (Reasoning, Computer [Title/Abstract])) OR (AI [Title/Abstract])) AND ((Education, Medical [MeSH Terms]) OR (Medical Education [Title/Abstract])); the search formula for WOS is set as: (((TS=(Artificial Intelligence)) OR TS=(Intelligence, Artificial)) OR TS=(Computer Reasoning)) OR TS=(Reasoning, Computer) OR TS=(AI) AND (TS=(Education, Medical)) OR TS=(Medical Education). In the Chinese database, the search function was set as: Artificial Intelligence AND Medical Education.

### 2.3. Inclusion and exclusion criteria

The inclusion criteria were as follows: the research content was related to AI and medical education, and the language used was Chinese or English.

The exclusion criteria were:[5] (Ⅰ) duplicate literature and (Ⅱ) non-original research articles, including conference abstracts, news reports, popular science articles, and patents. Other publication types, such as editorials, letters, and commentaries, were also excluded. (Ⅲ) Articles without one or more items, such as author, institution, keywords, publication time, and other information, were excluded.

### 2.4. Data analysis

We used VOSviewer and CiteSpace for a bibliometric analysis of research on AI application in medical education. VOSviewer, developed by the CWTS at Leiden University, visualizes extensive bibliometric networks because of its strong graphical display capabilities. CiteSpace, created by Professor Chaomei Chen using JAVA, is a popular tool for hotspot analysis that enables multi-perspective co-occurrence analysis and visualization of research interconnections. We made full use of the strengths of both complementary bibliometric software tools (CiteSpace [version 6.3. R1, Laiden University, Leiden, The Netherlands] and VOSviewer [version 1.6.20, Drexel University, Philadelphia, PA]) to conduct network analysis and visualization of the literature dataset.

In CiteSpace, we configured the analysis using the following key parameters to ensure robust network construction and trend detection: English literature, timespan: 2004–2025 (Slice Length = 1); selection criteria: g-index (k = 9); pruning: none; burstness: the parameters were set to a minimum burst duration of 2 years, and the gamma (γ) parameter, which controls the sensitivity for state transitions in the algorithm, was set to 0.8; Chinese literature: timespan: 2004–2025 (Slice Length = 1); selection criteria: g-index (k = 25); pruning: none; burstness: the parameters were set to a minimum burst duration of 1 year, and the γ parameter was set to 0.5.

Regarding VOSviewer, for co-occurrence network construction (e.g., author keywords), the following parameters were used: English literature; counting method: the full counting method was applied, where each co-occurrence link was counted fully for each source publication; minimum number of keyword occurrences was five; Chinese literature counting method: The full counting method was applied, in which each co-occurrence link was counted fully for each source publication; minimum number of keywords was two.

To conduct the search, we merged the English- and Chinese-keywords respectively (Appendix 1).

## 3. Results

After screening, 2437 articles in English and 326 articles in Chinese were retrieved. The literature-screening process is illustrated in Figure 1.

### 3.1. Trends in AI applications for medical education

Over the past two decades, the research output in the field of artificial intelligence applied in medical education has shown a significant upward trend. Literature output in this field has increased by more than sixfold over the past 5 years (Figure 2).

### 3.2. Productive countries/regions

An analysis of 2437 English publications revealed a dominant research landscape led by developed economies. The United States was the most productive country, followed by China and Germany. The United States, Germany, and Canada serve as core hubs that promote extensive international cooperation (Figure 3). China ranked second in the number of published papers; nonetheless, it had relatively few international collaborations. The top 10 countries/regions are presented in Table 1.

### 3.3. Analysis of the Chinese corpus

For the 326 Chinese publications, the analysis revealed a distinct pattern dominated by domestic medical universities and comprehensive universities with strong medical disciplines (Figure 4). Table 2 presents the top institutions involved. The strong presence of top-tier medical universities directly under national ministries and their close collaboration with engineering schools are direct manifestations of the New Medicine policy, which represents an overall reshaping of the medical system based on systems theory and integrating technologies such as precision and intelligent medicine. Its connotations include building interdisciplinary connections and promoting the integration of medicine, engineering, science, and the humanities through collaboration with AI; establishing a complete health lifecycle that covers prevention, treatment, and wellness; and promoting education reconstruction to build a cross-disciplinary, multifaceted talent cultivation system. This policy-driven ecosystem channels resources and mandates systemic integration, explaining the clustering of Chinese literature around policy and curriculum

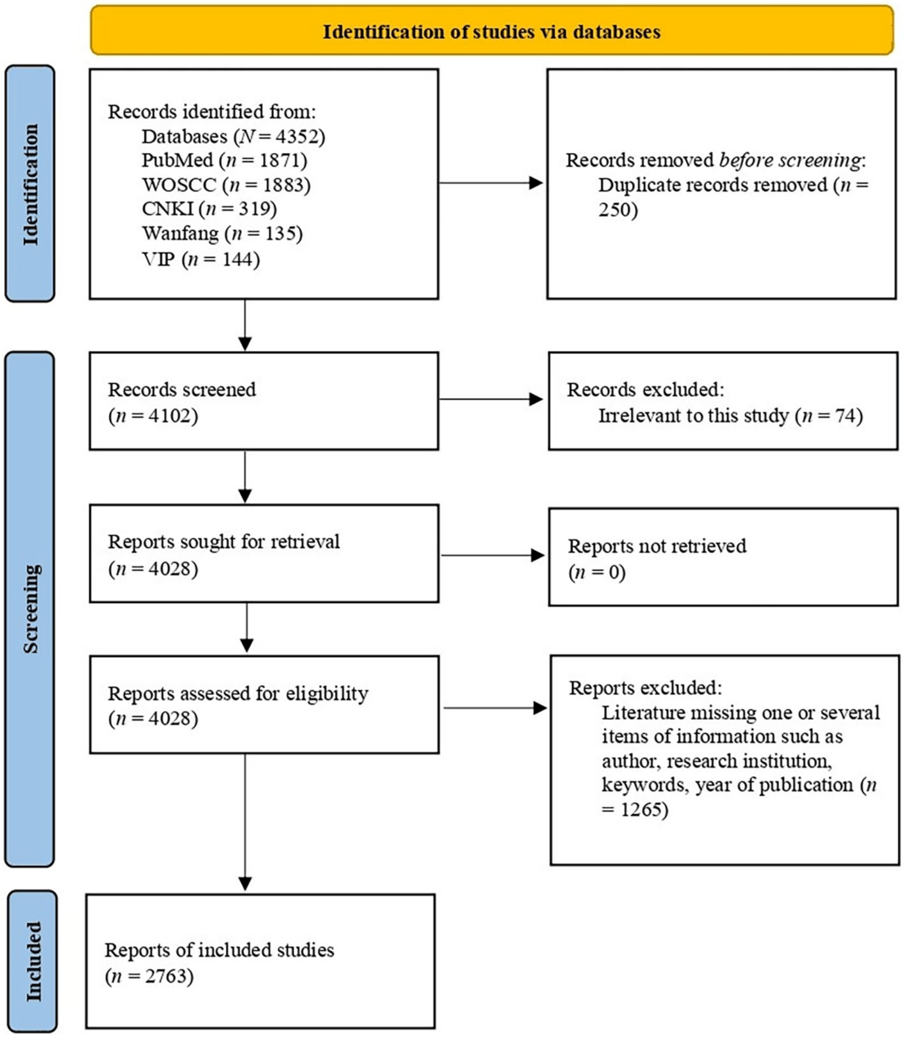


**Figure 1　Flowchart depicting the article selection process**

**Notes:** CNKI: China National Knowledge Infrastructure; VIP: Chinese Scientific Journal Database; WOSCC: Web of Science Core Collection.

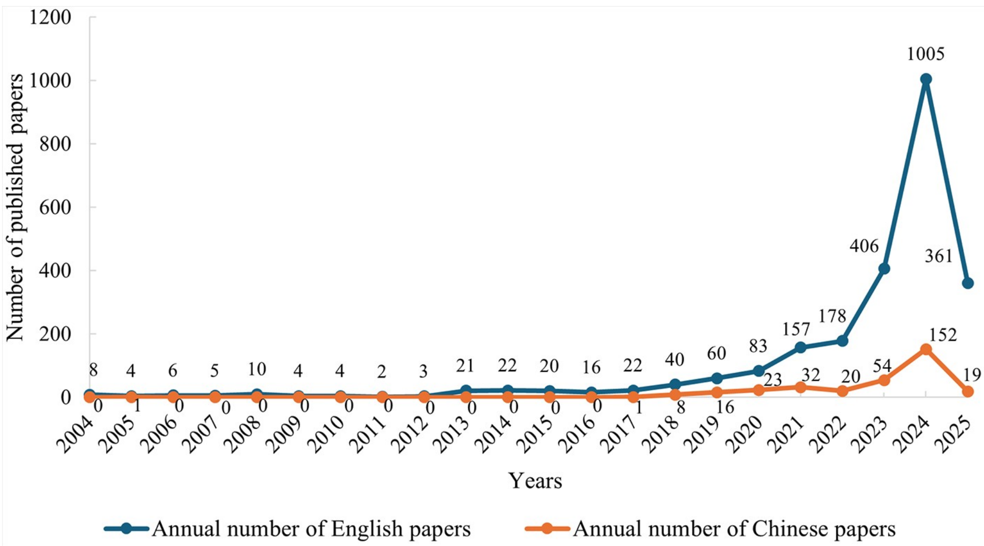


**Figure 2　Analysis of the trend of articles on the application of artificial intelligence in medical education from 2004 to 2025**

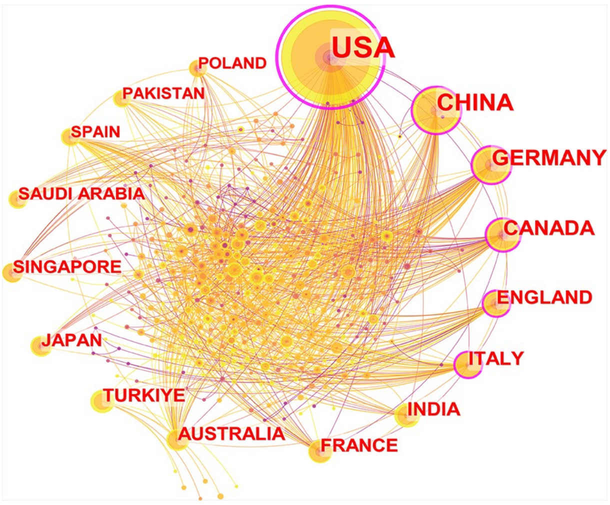


**Figure 3 Co-countries analysis of English publications on artificial intelligence applied in medical education during 2004–2025 (with CiteSpace)**

reform and macro-level system construction.

## 3.4. Research hotspots of Chinese and English keywords analysis

### 3.4.1. English keywords analysis: deep cultivation of technology and globalization issues

Keywords serve as core representations of articles and research topics. Analyzing high-frequency keywords helps reveal popular topics in AI applied in medical education. The "centrality" in the table refers to betweenness centrality. In this study, keywords with high centrality served as bridges between the two research topics. From the retrieved English literature, 20 core keywords were selected (Table 3), among which *AI* (1164 times), *medical education* (674 times), and *machine learning* (251 times) appeared the most frequently. The rapid development of AI and machine learning was highlighted (Figure 5).

In the retrieved literature, the keywords *VR* (57 times) and *surgical education* (56 times) had high co-occurrence rates. The keywords *patient education* (58 times) and *digital health* (48 times) frequently appeared in the literature retrieval, suggesting emerging directions for AI technology application.

**Table 1 Top 10 productive countries/regions in English literature**

| Ranking | Country | Frequency | Centrality |
|---|---|---|---|
| 1 | USA | 678 | 0.83 |
| 2 | China | 172 | 0.09 |
| 3 | Germany | 114 | 0.13 |
| 4 | Canada | 77 | 0.13 |
| 5 | England | 67 | 0.1 |
| 6 | India | 52 | 0.01 |
| 7 | Türkiye | 52 | 0 |
| 8 | Italy | 51 | 0.07 |
| 9 | France | 48 | 0.03 |
| 10 | Japan | 45 | 0.01 |

**Table 2 Top 10 productive institutions in Chinese literature**

| Ranking | Institution | Frequency | Centrality |
|---|---|---|---|
| 1 | Tongji Medical College | 10 | 0 |
| 2 | West China Hospital of Sichuan University | 9 | 0.01 |
| 3 | Harbin Medical University | 6 | 0 |
| 4 | Nanjing Medical University | 6 | 0 |
| 5 | Zhejiang University | 5 | 0 |
| 6 | Capital Medical University | 4 | 0 |
| 7 | Central South University | 4 | 0 |
| 8 | Lanzhou University | 4 | 0 |
| 9 | Gansu University of Chinese Medicine | 4 | 0 |
| 10 | Chinese Academy of Medical Sciences | 4 | 0 |

### 3.4.2. Chinese keywords analysis: policy orientation and localization practice

In the retrieved Chinese studies, 20 keywords were selected (Table 4), among which *AI* (183 times), *medical education* (161 times), and *teaching*

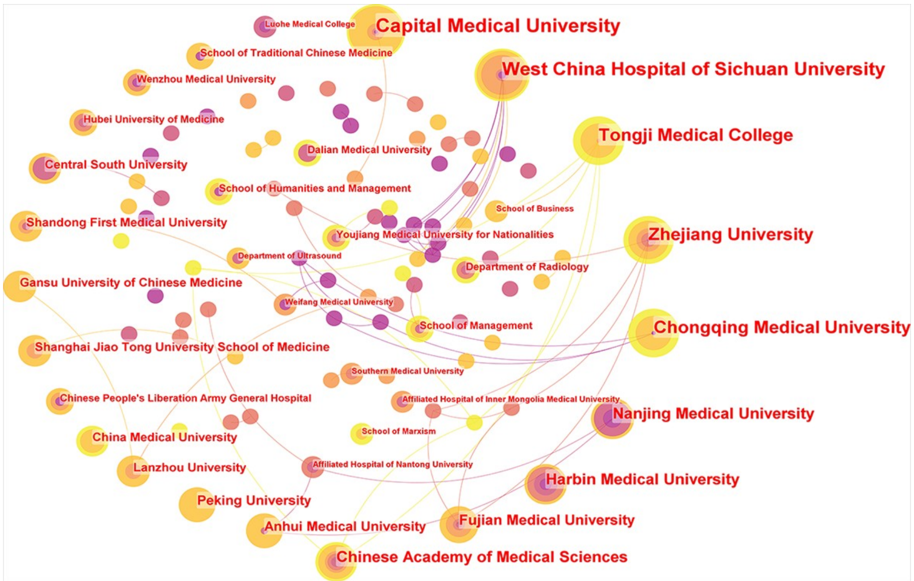


**Figure 4 Co-institutions analysis of Chinese publications on artificial intelligence applied in medical education during 2004–2025 (with CiteSpace)**

**Table 3 Top 20 keywords in terms of frequency of occurrence of core keywords in English articles**

| Ranking | Keywords | Frequency | Centrality |
|---|---|---|---|
| 1 | Artificial intelligence | 1164 | 0.57 |
| 2 | Medical education | 674 | 0.58 |
| 3 | Machine learning | 251 | 0.2 |
| 4 | Large language model | 222 | 0.13 |
| 5 | Medical student | 135 | 0.1 |
| 6 | Generative ai | 100 | 0.1 |
| 7 | Deep learning | 95 | 0.06 |
| 8 | Natural language processing | 95 | 0.03 |
| 9 | Patient education | 58 | 0.04 |
| 10 | Virtual reality | 57 | 0.05 |
| 11 | Surgical education | 56 | 0.05 |
| 12 | Digital health | 48 | 0.03 |
| 13 | Language model | 30 | 0.03 |
| 14 | Clinical reasoning | 29 | 0.04 |
| 15 | Graduate medical education | 29 | 0.07 |
| 16 | Clinical decision support | 19 | 0.03 |
| 17 | Medical education & training | 19 | 0.03 |
| 18 | Clinical decision-making | 18 | 0.02 |
| 19 | Prompt engineering | 18 | 0.01 |
| 20 | Medical imaging | 18 | 0 |

**Table 4 Top 20 keywords in terms of frequency of core keywords in Chinese articles**

| Top | Keywords | Frequency | Centrality |
|---|---|---|---|
| 1 | Artificial intelligence | 183 | 1.00 |
| 2 | Medical education | 161 | 0.74 |
| 3 | Reform of teaching | 21 | 0.01 |
| 4 | New medicine | 20 | 0.08 |
| 5 | Training of talents | 14 | 0.02 |
| 6 | Education | 11 | 0.06 |
| 7 | Teaching and learning | 9 | 0.04 |
| 8 | Virtual reality | 9 | 0.02 |
| 9 | Medical student | 7 | 0.01 |
| 10 | Clinical teaching | 6 | 0.00 |
| 11 | Oral medicine | 6 | 0.00 |
| 12 | Clinical medicine | 6 | 0.05 |
| 13 | Teaching model | 6 | 0.01 |
| 14 | Big data | 5 | 0.01 |
| 15 | Application | 5 | 0.00 |
| 16 | Education reform | 5 | 0.00 |
| 17 | Ethical risks | 4 | 0.05 |
| 18 | Internet | 4 | 0.03 |
| 19 | Internship | 4 | 0.00 |
| 20 | Ophthalmology department | 4 | 0.00 |

*reform* (21 times) appeared most frequently, indicating that "medical education reform" was the main focus in this field. In addition, the keyword "*New Medicine*" (20 times) appears frequently and reflects the Ministry of Education of China's New Medicine construction strategy, which emphasizes the cross-integration of AI and medicine. The strategy reflects the focus on government policies in this field. The keyword further highlighted the research focus of AI in medical education, including the reconstruction of the curriculum systems of medical colleges and universities, the addition of AI+ medicine cross-courses, and the large-scale AI technology application in teaching, such as intelligent classrooms and Massive Open Online Courses (MOOC) platforms (Figure 6).

Additionally, keywords such as *clinical teaching* and *oral medicine* (each mentioned six times) highlighted the demands for AI-assisted clinical skill development. Practical implementations include virtual standardized patient

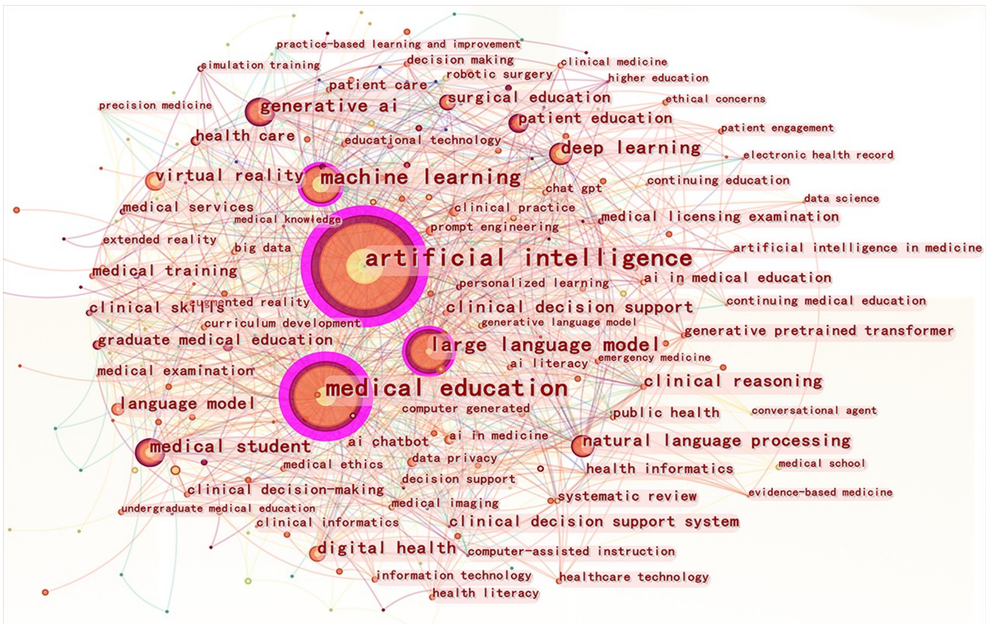


**Figure 5 Keyword analysis of English literature on the application of artificial intelligence in medical education from 2004 to 2025**

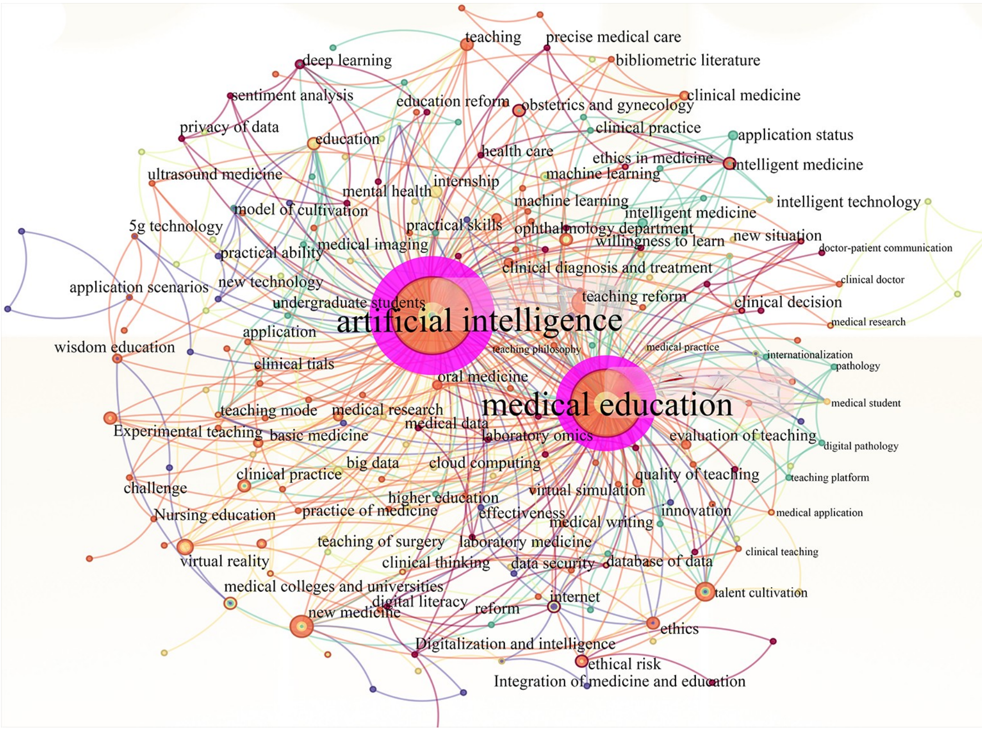


**Figure 6** **Keyword analysis of Chinese literature on the application of artificial intelligence in medical education from 2004 to 2025**

systems and regional medical data-sharing platforms, demonstrating the transformative potential of AI in clinical competency training.

Crucially, the emergence of *ethical risk* (4 times) signals concerns regarding AI deployment challenges. It reflects social concerns about the possible violation of the ethics and morality of using AIs and the role of AI technology in education equity.

### 3.4.3. Comparison and integration trend of hot spots in Chinese and English

A comparative analysis of Chinese and English literature reveals converging technological synergies between AI and VR/ augmented reality (AR). Both academic traditions emphasize immersive VR teaching scenarios, as evidenced by parallel keyword associations. We observed that VR in Chinese publications correlates with surgical education in English literature. Furthermore, a shared research priority emerges in the transition from knowledge transfer to competency development: English keywords such as *clinical reasoning* (29 times) and their Chinese counterparts, *training of talents* (14 times), collectively underscore pedagogical commitment to cultivating higher-order cognitive skills. This intersection highlights a common research imperative: leveraging AI to enhance critical thinking and practical abilities.

The Chinese keyword *intelligent education* signals comprehensive AI integration across educational ecosystems, whereas specialized technical refinement—particularly in large language models (LLMs) and NLP—is prioritized in the English literature. Ethical dimensions also exhibit distinct emphasis. Chinese research predominantly addresses explicit policy risks in discussions on AI ethics, whereas English scholarship focuses on implicit technical ethics. Therefore, the application contexts differ between the two bodies of literature. Chinese studies prioritize institutional teaching reforms, in contrast to the focus of English literature on clinical skill development and patient education methodologies.

## 3.5. Keyword clustering (research topic)

### 3.5.1. Clustering analysis of English keywords

Through keyword cluster analysis (Figure 7), 10 distinct clusters were

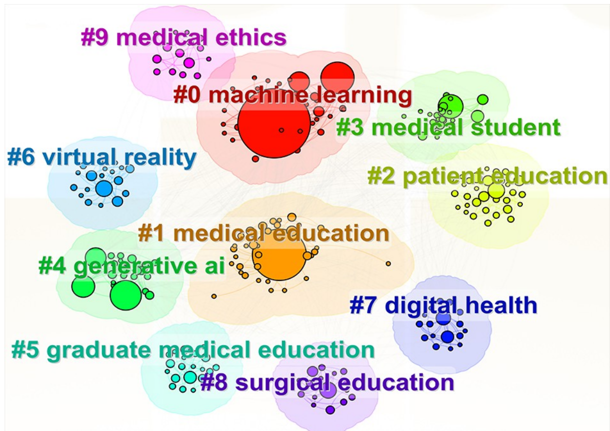


**Figure 7** **Keyword clustering in English literature from 2004 to 2025**

identified in the English language corpus (Table 5). Based on thematic focus, these clusters can be organized into three overarching categories that reveal the intellectual structure of AI in medical education research.

The first category comprises technology-centered clusters that underpin the foundational infrastructure of the field. #0 (machine learning) and #4 (generative ai) represent the core computational technologies that drive innovation. These clusters demonstrate that deep learning architectures and LLMs are not merely technical tools but also fundamental forces reshaping educational possibilities. For #6 (virtual reality) and #7 (digital health), the technological foundation is extended to immersive and digital health domains. The second category comprises education-focused clusters, which represent the primary application domains of these technologies. In #1 (medical education), #3 (medical students), #5 (graduate medical education), and #8 (surgical education), the sections collectively address the continuum of undergraduate medical training through graduate and specialized surgical education. #2 (patient education) represents a distinctive application domain in which AI technologies are deployed beyond traditional medical training to

**Table 5 Keyword clustering and keywords in English literature**

| Cluster ID | Size | Cluster | Keywords |
|---|---|---|---|
| 0 | 46 | machine learning | machine learning; artificial intelligence; deep learning; large language model; medical imaging |
| 1 | 41 | medical education | medical education; machine learning; public health; automatic item generation; faculty development |
| 2 | 32 | patient education | patient education; machine learning; clinical decision support; health care; prompt engineering |
| 3 | 28 | medical student | medical student; clinical reasoning; clinical decision-making; machine learning; breast cancer |
| 4 | 27 | generative ai | generative ai; large language model; natural language processing; machine learning; artificial intelligence |
| 5 | 22 | graduate medical education | graduate medical education; undergraduate medical education; ai (artificial intelligence); curriculum development; intelligent tutoring system |
| 6 | 19 | virtual reality | virtual reality; augmented reality; extended reality; mixed reality; metaverse |
| 7 | 17 | digital health | digital health; language model; generative language model; patient outcomes; orthopedic surgery |
| 8 | 16 | surgical education | surgical education; robotic surgery; practice-based learning and improvement; medical knowledge; resident training |
| 9 | 15 | medical ethics | medical ethics; medical education & training; health informatics; learning strategy; statistics & research methods |

support patient learning and engagement. The third category encompasses the ethical and governance clusters that address the societal implications of AI integration. #9 (Medical ethics) represents research examining the ethical challenges associated with AI adoption. In this study, the coexisting privacy and data-sharing mechanisms for electronic health records that balance educational value with ethical risks were explored. Collectively, these three categories form a comprehensive framework for understanding how AI is reshaping medical education within the English language research tradition.

### 3.5.2. Clustering analysis of Chinese keywords

A keyword cluster analysis identified 15 thematic research groups in the Chinese language corpus (Figure 8, Table 6). These clusters can be systematically organized into three thematic categories reflecting the distinct priorities and institutional contexts of Chinese research on AI in medical education.

The first category comprises the systemic reform clusters that reflect the policy-driven transformation of the medical education infrastructure. #0 (medical education), #2 (education), #3 (big data), #6 (experimental teaching), #7 (clinical medicine), and #14 (standardized training) collectively address the structural reorganization of medical education in response to national policy directives. Research in #3 (big data) and #6 (experimental teaching) demonstrates that VR/mixed reality (MR) technology has become a core tool for experimental teaching. Cluster #14 (standardized training) reflects the policy emphasis on creating uniform training standards across the national medical education system. The second category encompasses technology-driven transformation clusters that represent specific innovations deployed within this systemic reform context. Clusters #1 (artificial intelligence), #4 (teaching and learning), #5 (New Medicine), #10 (training of talents), #12 (intelligent medicine), and #13 (information literacy) demonstrate how deep learning and LLMs reconstruct the ecology of medical education. The strong correlations between clusters #4 (teaching and learning), #2 (education), and #3 (big data) indicate that data-driven approaches to pedagogy are central to the Chinese research agenda. Cluster #13 (information literacy) represents an emerging focus on ensuring that medical students and practitioners possess the skills needed to engage with AI tools effectively and critically. The third category comprises the ethical governance and competency development clusters, which address the human and social dimensions of AI integration. In clusters #8 (medical applications), #9 (medical research), and #11 (doctor-patient communication), the implications of AI in professional practice and ethical governance were explored.

### 3.5.3. Synthesis of cross-linguistic patterns

The comparison between English and Chinese clustering patterns reveals convergences and divergences with significant implications when organized using the systematic framework in this study. Both research traditions

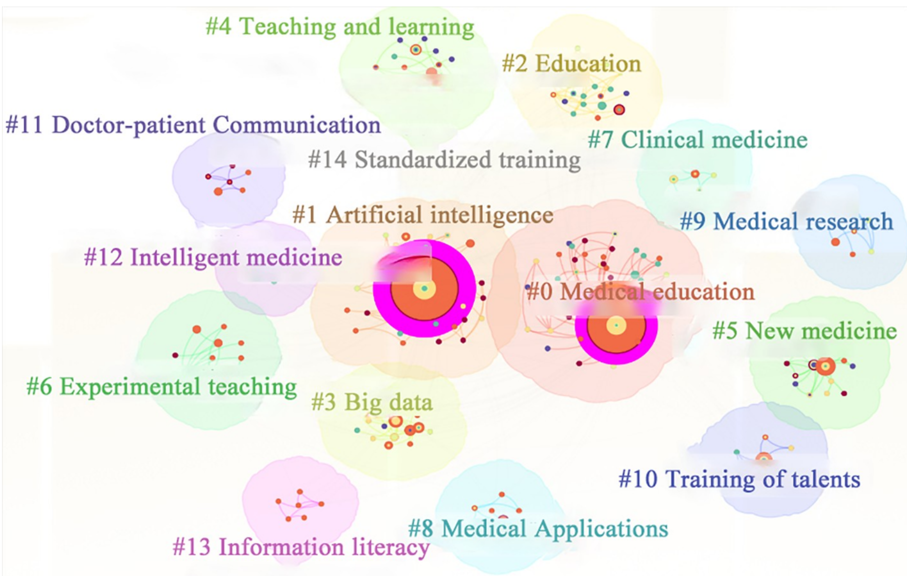


**Figure 8 Keyword clustering in Chinese literature from 2004 to 2025**


清華大学出版社 Tsinghua University Press | SciOpen

**Table 6 Keyword clustering and keywords in Chinese literature**

| Cluster ID | Size | Cluster | Keywords |
|---|---|---|---|
| 0 | 42 | medical education | medical education; medical students; application status quo; teaching reform; department of otolaryngology |
| 1 | 36 | artificial intelligence | artificial intelligence; deep learning; the clinician; teaching methods; medicine |
| 2 | 20 | education | education; teaching model; medical imaging; education reform; model of cultivation |
| 3 | 18 | big data | big data; virtual reality; clinical practice; ethics; medical research |
| 4 | 17 | teaching and learning | teaching and learning; application scenarios; wisdom education; medical education; data mining |
| 5 | 16 | new medicine | new medicine; combination of medicine and education; the internet; digital intelligence; precision rehabilitation |
| 6 | 12 | experimental teaching | experimental teaching; teaching reform; basic medical science; application; oral medicine |
| 7 | 9 | clinical medicine | clinical medicine; the influencing factors; college students; bibliometrics; human-machine collaboration |
| 8 | 8 | medical applications | medical applications; current status; governance; advantages; quality control |
| 9 | 7 | medical research | medical research; ethics of science and technology; regional ethics committees; ethical review; ethical governance |
| 10 | 6 | training of talents | personnel training; professional development; intelligent medical engineering; artificial intelligence education; traditional Chinese medicine |
| 11 | 6 | doctor-patient communication | doctor-patient communication; department of respiratory medicine; language of empathy; teaching evaluation; clinical decision making |
| 12 | 6 | intelligent medicine | intelligent medicine; healthcare; medical prevention integration: medical reform; medical ethics; medical school |
| 13 | 6 | information literacy | information literacy; opportunity; challenges; nursing education; chatgpt |
| 14 | 4 | standardized training | standardized training; practical skills; mode of education; ultrasonic medicine; medical specialists |

demonstrate the tripartite structure of technology infrastructure, educational applications, and ethical governance, suggesting that these three dimensions represent fundamental categories for understanding AI in medical education, regardless of the context. However, relative emphasis and internal organization differ. English-language research exhibits greater granularity in technology, focusing on clusters and the earlier emergence of ethical discourse, reflecting its position at the technological frontier and its responsiveness to innovation-driven research incentives. Chinese language research exhibits a closer integration between technology and systemic reform clusters, with ethical discourse emerging later and more directly tied to policy frameworks, reflecting its position within a coordinated national strategy for educational transformation. These patterns are not merely descriptive, but reveal how the same global technologies are being adapted to and shaped by distinct institutional and policy contexts.

## 3.6. Time chart (research development trajectory)

We analyzed the development of AI applications in medical education using a time graph of literature keyword clustering. As shown in the CiteSpace timeline chart (Figure 9), node color intensity reflects the year of occurrence of the keyword; darker, purpler nodes indicate earlier publications closer to 2020, and lighter, yellow nodes indicate the latest publications closer to 2025. From the timeline chart of the English literature, we observe that the research on #0 (machine learning) has remained a major topic, #4 (generative ai) and #5 (graduate medical education) have only been active since 2021, while #7 (digital health) and #9 (medical ethics) were active until 2024. #0 (machine learning) has maintained strong prominence since 2020 and is directly related to the protein structure prediction revolution triggered by AlphaFold2. The data show that research on ML-based clinical decision support systems has increased significantly since 2020 (PubMed statistics); in addition, a closed loop of technological penetration has been formed in the

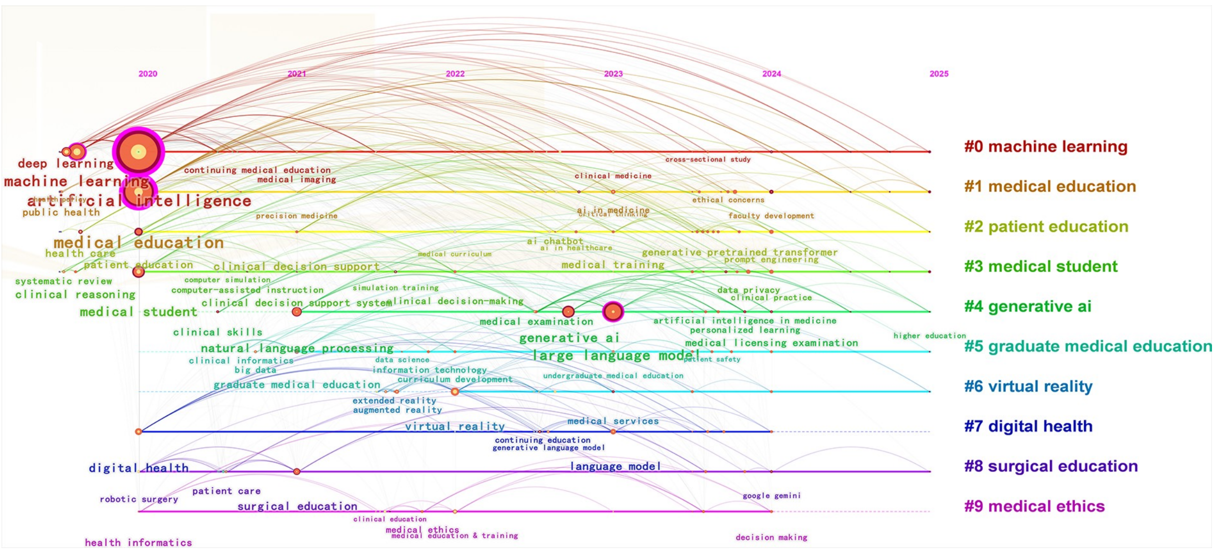


**Figure 9 Timeline visualization of English literature from 2004 to 2025 (with CiteSpace)**

field of imaging diagnosis education. The activity of cluster #4 (generative ai) increased since 2021 and was released simultaneously with GPT-3.5 and ChatGPT. #7 (digital health) and #4 (generative ai) were active in 2023, reflecting the accelerated application of health records in teaching after implementing the Fast Healthcare Interoperability Resources (FHIR) standard.

From the analysis of the Chinese literature timeline chart generated using CiteSpace (Figure 10), research on #0 (medical education), #1 (artificial intelligence), #5 (New Medicine), and #6 (experimental teaching) has remained a main topic. Research on #2 (education), #3 (big data), and #4 (teaching and learning) was active until 2024. #12 (intelligent medicine), #8 (medical applications), and #11 (doctor–patient communication) were active in 2021, 2023, and 2024, respectively. The research density of #8 (medical applications) and #9 (medical research) increased in 2024. #2 (education) and #6 (experimental teaching) remained active, reflecting the accelerated digital transformation of medical education in China after the coronavirus disease 2019 (COVID-19) pandemic.

Analysis of the timeline chart for English literature generated by VOSviewer (Figure 11) shows a high correlation between #1 (education), #2 (medical education), #3 (chatgpt), and #4 (artificial intelligence). This find-

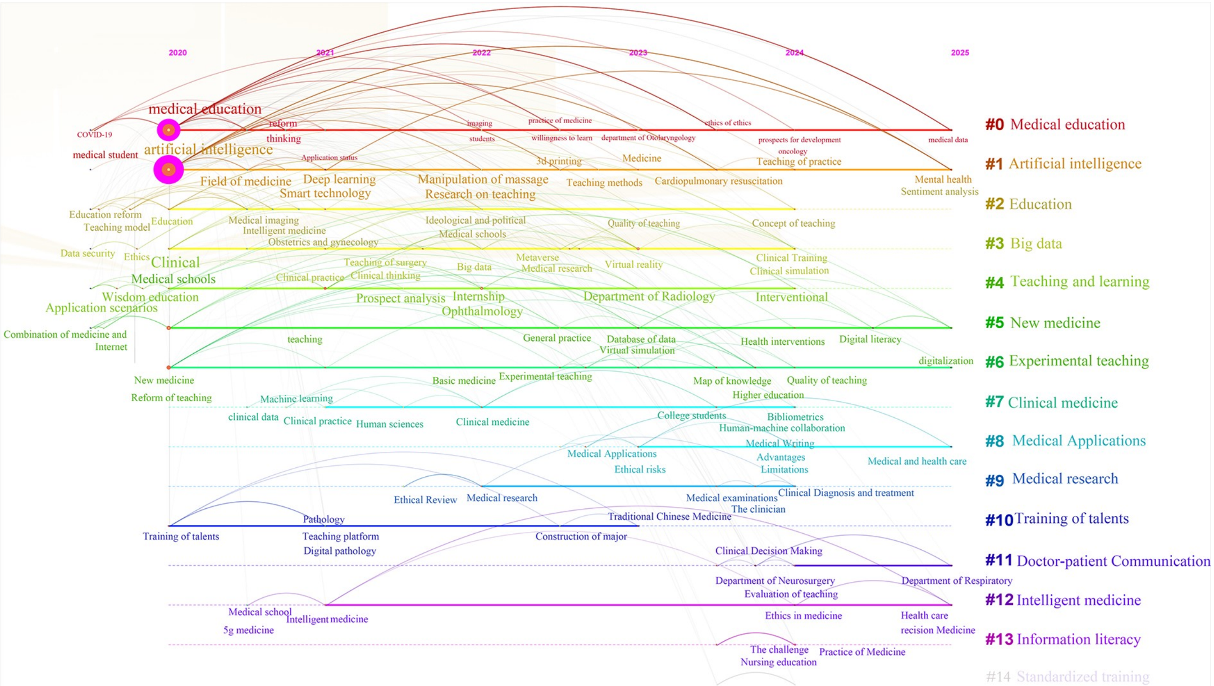


Figure 10 Timeline visualization of the Chinese literature from 2004 to 2025 (with CiteSpace)

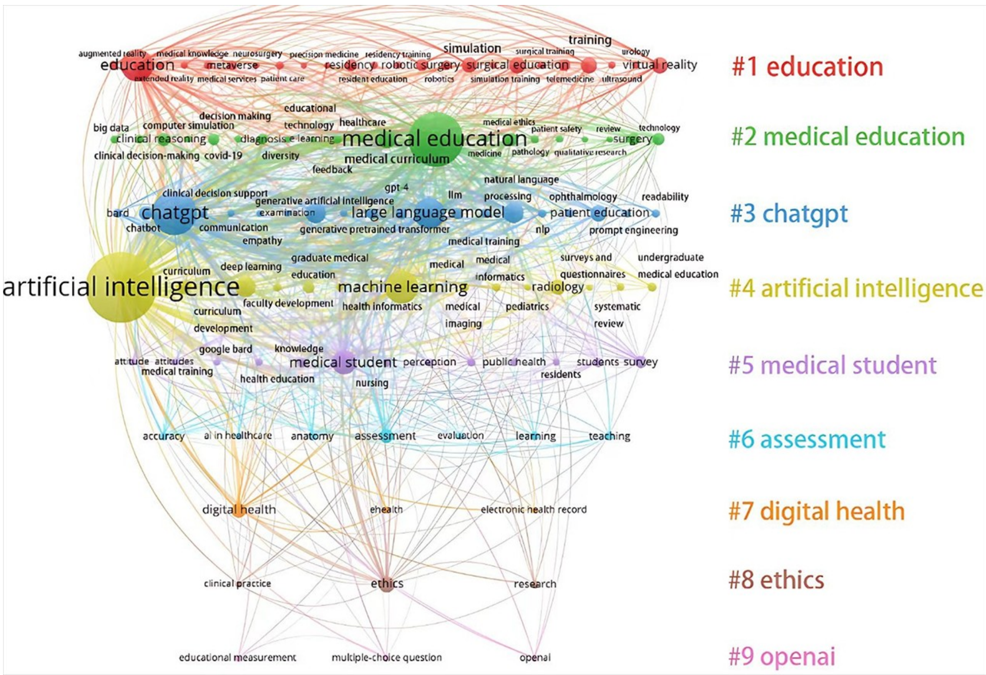


Figure 11 Timeline visualization of the English literature from 2004 to 2025 (with VOSviewer)


清华大学出版社 Tsinghua University Press | SciOpen

ing indicates that AI is widely used in medical education globally. The association between #4 (artificial intelligence), #6 (assessment), and #8 (ethics) suggests that the medical community is concerned about the ethical issues of AI applications and requires further reviews of its use in the field.

The analysis of the timeline chart of Chinese literature generated using VOSviewer (Figure 12) reveals that the correlations between #1 (artificial intelligence), #3 (chatgpt), and #4 (teaching reform) suggest that the medical education field strongly recommends introducing new technological means to promote and improve teaching effectiveness. Moreover, integrating AI technology into education can effectively transform traditional teaching. The connection between #1 (artificial intelligence), #3 (chatgpt), and #5 (medical education) indicates that practices in China align with the global trend, focusing on applying AI in medical education and using it to provide medical students with simulation-based scenario training and personalized learning resources.

## 3.7. Keywords highlight (research frontier)

Keyword bursting is the sudden or repeated appearance of keywords within a defined timeframe. This phenomenon indicates an increase in attention or research activity around a topic or research area related to a keyword during a specific period. “Burst” (sudden emergence) is a key metric detected by the Kleinberg algorithm and used to identify sudden, explosive growth in the frequency of a keyword, reference, author, or term over a specific period. It reveals the “turning point” or “starting point” of research interest.

### 3.7.1. English keywords highlight research frontiers

Keyword occurrences were roughly divided into three different stages based on their initial appearances (Figure 13). Phase A lasted from 2020 to 2021, during which deep learning and clinical data were central, reflecting the urgent need for the digital transformation of medical education during the early stages of the COVID-19 pandemic. The primary research areas during this period included machine learning, deep learning, robotic surgery, electronic health records, precision medicine, residency training, convolutional neural networks, and medical imaging. The continuous popularity of machine learning and deep learning stems from the teaching revolution in structural biology driven by AlphaFold2. The emergence of the keyword *robotic surgery* coincided with the popularization of the Da Vinci surgical system teaching module.

Phase B began in 2022. As metaverse technology advanced, the research focus shifted to immersive teaching and curriculum reconstruction. The main research fields during this period included VR, medical curriculum, medical education and training, qualitative research, AR, and MR. The emergence of the keywords *VR* and *MR* has transformed anatomy teaching models. Furthermore, the popularity of keyword qualitative research has increased, reflecting the transformation of AI education effect evaluation from a quantitative indicator to a mixed research paradigm.

Phase C, from 2023 to 2024, focused on popularizing generative AI to rebalance educational ethics and humanistic qualities. In this period, the main research areas included resident education, medical school, chat groups, doctor-patient relationships, neurosurgical training, and clinical practice. The emergence of the keyword *chatgpt* was accompanied by a crisis in academic integrity. The co-occurrence of the keywords *doctor-patient relationship* and *neurosurgical training* reveals a new path for cultivating clinical empathy empowered by technology. The keyword *clinical practice* continued to emerge until 2025. This marked a paradigm shift in AI education from knowledge transmission to the cultivation of clinical competence.

### 3.7.2. Chinese keywords research frontiers

Phase A, which began in 2020, mainly focused on research topics such as educational reform, talent cultivation, the Internet, and cultivation models (Figure 14). In this period, the focus was on deep learning and clinical dativization, reflecting the teaching response to the digital transformation of education during the early stages of the COVID-19 pandemic. For instance, AlphaFold2 has revolutionized teaching in structural biology, and robotic surgery training has reduced the incidence of complications among resident physicians.

Since phase B in 2021, as metaverse technology has advanced, research has shifted toward immersive teaching and curriculum reconstruction. Keywords include *Intelligent medicine*, *application status*, *education, reform*, *thinking*, *medical imaging*, and *clinical practice*. This period transformed anatomical teaching models, such as MR holographic models, enhancing cognitive efficiency and facilitating the integration of AI tools into medical courses, including the application of AI case generators at Harvard Medical School.

Starting in 2022, phase C highlights that AI in medical education is now

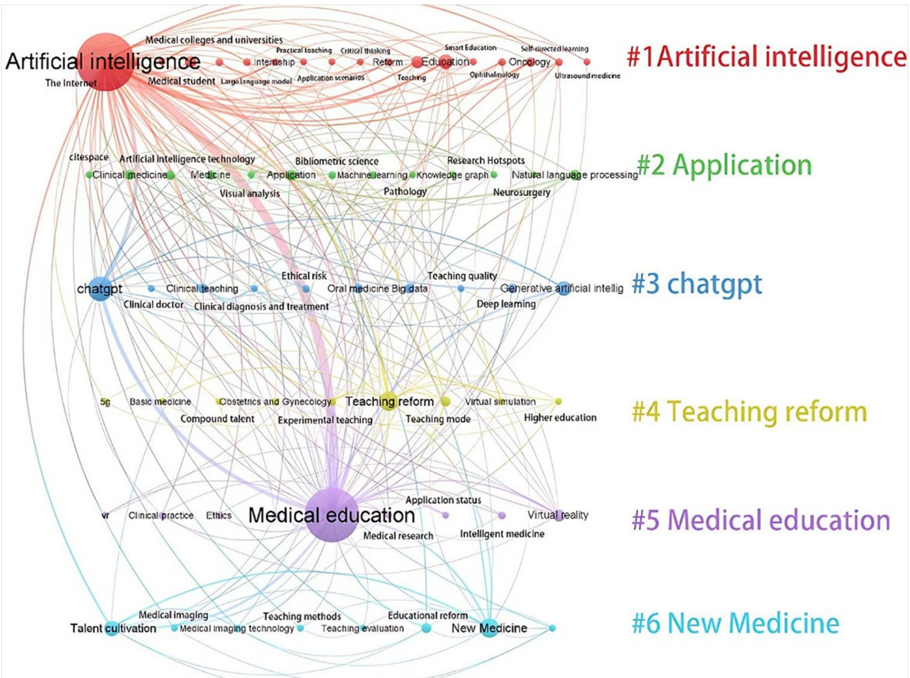


**Figure 12** **Timeline visualization of the Chinese literature from 2004 to 2025 (with VOSviewer)**

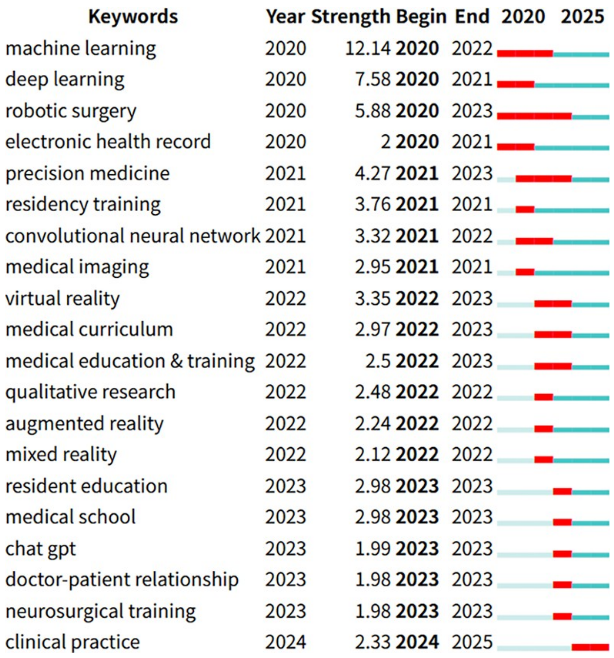


**Figure 13 Top 20 keywords with the strongest citation bursts in English literature from 2004 to 2025**

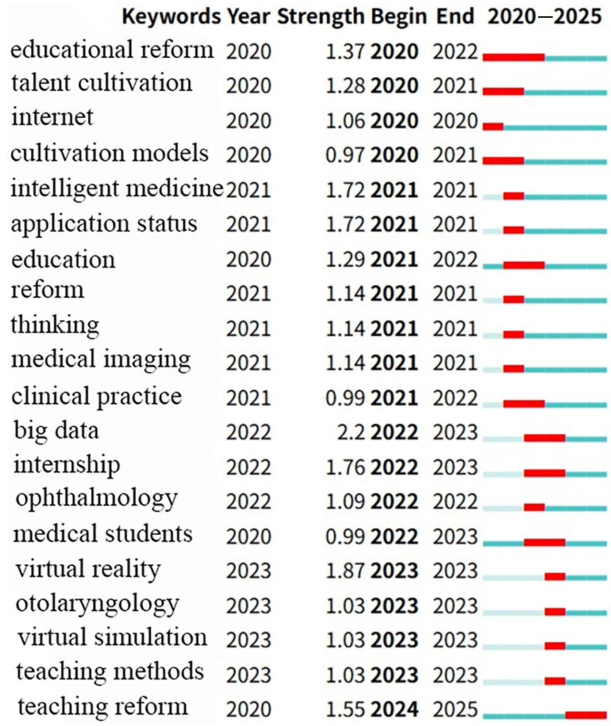


**Figure 14 Top 20 keywords with the strongest citation bursts in Chinese literature from 2004 to 2025**

being implemented in specific hospital departments. The keywords used were *big data*, *internship*, *ophthalmology*, *medical students*, *VR*, *otolaryngology*, *virtual simulation*, *teaching methods*, and *teaching reform*. During this period, following the previous period of AI application in education and teaching reform, it began to penetrate various hospital departments and was applied based on local conditions to transform education from knowledge-based to practical ability training.

### 3.7.3. Comparison and cross-trend between Chinese and English

Chinese and English literature reflect a transition from methodological reform to the deep integration of technologies, and big data and machine learning are the common core of both. The English keyword *surgical education* and that of the Chinese keyword *internship* emphasized technical support for clinical practice.

Cutting-edge technologies, such as deep learning and robotic surgery, were explored in the English literature, whereas the Chinese literature focused on technical adaptability, such as the reform of teaching models. In the Chinese literature, the education system and basic research are emphasized, whereas English literature focuses on clinical scenarios and curriculum reconstruction, including the use of VR.

# 4. Discussion

We discussed the literature and keywords in AI for medical education. The analysis was divided into four stages: research hotspots identified through keywords, cluster analysis topics, research development in the field of timeline analysis, and research frontiers in the application of AI in medical education identified through keywords.

## 4.1. Research hotspots identified through keywords

*AI*, *medical education*, and *machine learning*, these English keywords clearly demonstrate the remarkable progress of artificial intelligence and machine learning. Furthermore, applications include the use of ChatGPT in the automatic scoring of medical examinations, as well as employing multimodal LLMs (such as GPT-4V) to assist in teaching imaging diagnosis and other cutting-edge scenarios.[6] The keywords *VR* and *surgical education* had high co-occurrence rates, suggesting a close connection between VR technology and surgical education.[7] Representative research projects include verifying the effectiveness of VR surgical simulators in resident training and applying AR technology in real-time guidance of minimally invasive surgery.[8,9] The keywords *patient education* and *digital health* frequently appeared in the literature retrieval, suggesting emerging directions for AI technology application. For example, AI is used to generate personalized patient education materials, such as diabetes management animation, and to support doctor–patient communication through scenario-based simulation training using data from wearable devices.[10]

## 4.2. Cluster analysis topics

The English clusters #0 (machine learning) and #4 (generative ai) demonstrate that deep learning architectures and LLMs are not merely technical tools but also fundamental forces reshaping educational possibilities. The core content of this research includes developing intelligent question-answering systems that enhance clinical reasoning training for medical students and applications such as ChatGPT to optimize instructional content generation.[11-13] For #6 (virtual reality) and #7 (digital health), the technological foundation is extended to immersive and digital health domains. Research on #6 (virtual reality) focuses on virtual and mixed reality (MR) applications in anatomy education and surgical simulation, including 3D anatomical models and robotic surgical training, demonstrating how interactive modalities improve practical skill acquisition while reducing the resource dependencies associated with cadaver dissection and traditional apprenticeships.[14,15] #7 (digital health) encompasses digital health technologies that integrate AI into broader healthcare delivery systems. The significance of this technological category lies in its demonstration that AI fundamentally expands the pedagogical toolkit, enabling personalized, scalable, and immersive learning experiences that were previously impossible. The second category comprises education-focused clusters, which represent the primary application domains of these technologies. In #1 (medical education), #3 (medical students), #5 (graduate medical education), and #8 (surgical education), the sections collectively address the continuum of undergraduate medical training through graduate and specialized surgical education. The core content reveals how AI technologies are deployed across this continuum; AI-driven clinical scenario simulations and intelligent tutoring systems (ITSs) are used to train diagnostic logic and decision-making pathways.[16,17] Deep learning-based medical image analysis, such as orthopedic imaging reading and tumor detection, has revolutionized radiology training by providing automated interpretation and immediate feedback.[18,19] #2 (patient education) represents a distinctive application domain in which AI technologies are deployed beyond traditional medical training to support patient learning and engagement. The field significance of this category is that it demonstrates that AI integration in medical education is not monolithic, but differentiated across training levels and contexts, with technologies being adapted to meet the specific needs of learners at different stages. The third category encompasses the ethical and governance clusters that address the societal implications of AI integration. #9 (medical ethics) represents research examining the ethical challenges associated with AI adoption. The core content focuses on implicit technical ethics issues such as algorithm bias in diagnostic AI, patient data privacy in electronic health record-based learning, and the transparency of clinical decision support systems.[20,21] In this study, the coexisting privacy and data-sharing mechanisms for electronic health records that balance educational value with ethical risks were explored. The significance of this category lies in its essential role as a corrective and guiding force, ensuring that technological enthusiasm is tempered by ethical reflection and that educational innovations are implemented responsibly. Collectively, these three categories form a comprehensive framework for understanding how AI is reshaping medical education within the English language research tradition.

The Chinese clusters #0 (medical education), #2 (education), #3 (big data), #6 (experimental teaching), #7 (clinical medicine), and #14 (standardized training) collectively address the structural reorganization of medical education in response to national policy directives. The core content reveals how these clusters operationalize policy goals such as the New Medicine construction strategy, which mandates the deep integration of AI and big data into medical education. Research in #3 (big data) and #6 (experimental teaching) demonstrates that VR/MR technology has become a core tool for experimental teaching, particularly in fields such as stomatology.[22] Coxe et al.[23] showed that VR simulation training improved outcomes compared with those of controls, resulting in shorter procedure times, higher rates of correct step completion, and more accurate implant placement.

## 4.3. Research development in the field of timeline analysis

The data show that research on ML-based clinical decision support systems has increased significantly since 2020, Mohammad et al.[24] indicated that the assistance that AI case generators provide for medical education.

The research density of #8 (medical applications) and #9 (medical research) increased in 2024, which aligns with the goal of achieving a coverage rate of intelligent medical courses of ≥60% proposed in the Outline of China's Medical Education Reform (2023). Zheng et al.[25] conducted a study indicating that the AI-enabled scenario simulation-based teaching mode aided in enhancing the clinical thinking and skills of medical undergraduates. Notably, the VR War Trauma Rescue Training System of the Army Medical University (Chongqing, China) has significantly shortened the time required for trainees to master skills and is recognized as a virtual simulation experimental teaching project by the Ministry of Education. The emergence of #11 (doctor-patient communication) in 2024 is directly related to the Physicians' Law, which emphasizes strengthening the communication between doctors and patients. AI provides personalized learning resources and interaction methods tailored to students' different situations, and solves specific medical problems students encounter in a targeted manner. However, traditional clinical training often faces insufficient resources. A study conducted in the College of Nursing at Sahmyook University (Seoul, Republic of Korea) indicated that AI can create virtual environments, facilitate clinical skill training, provide medical students with practical opportunities similar to real diagnosis and treatment scenarios, and supplement their limited clinical experience.[26]

English research focuses more on the underlying technological innovations (such as #4 generative ai), whereas Chinese literature emphasizes the localized application of technologies (such as #1 medical education and #12 intelligent medicine). In the English domain, as early as 2020, there was a concentrated discussion on question #9 (medical ethics). In Chinese literature, research on ethical review mechanisms emerged only in Cluster 8 (Medical Applications) in 2023, reflecting the phased differences in the construction of the governance framework. Cluster #5 (graduate medical education) in English research emphasized the infiltration of AI into specialist physicians' continuous education. Meanwhile, #5 (New Medicine) in Chinese research indicates that it focuses on integrating and reconstructing intelligent medicine and traditional clinical courses.

## 4.4. Research frontiers in the application of AI in medical education identified through keywords

The co-occurrence of electronic health record (EHR) and precision medicine indicates the initial application of FHIR standards in medical education. Milano et al.[27] proposed that the EHR simulation system can provide a safe, practical environment for learners and greatly reduce the error rate of medical consultation data entry.

The emergence of the keywords *VR* and *MR* has transformed anatomy teaching models. Harvard Medical School (Cambridge, MA) was the first to integrate an AI case generator into its core curriculum, covering most of the preclinical teaching links. Furthermore, the popularity of keyword qualitative research has increased, reflecting the transformation of AI education effect evaluation from a quantitative indicator to a mixed research paradigm. A study published in JAMA (2022) showed that qualitative analysis methods combined with NLP could enhance the reliability of evaluating the effectiveness of teaching interventions. The emergence of the keyword *chatgpt* was accompanied by a crisis in academic integrity. A study published in *The New England Journal of Medicine* in 2023 revealed that many medical students will use generative AI to write papers, with most of these papers carrying the risk of fabricating references.[28] The co-occurrence of the keywords *doctor-patient relationship* and *neurosurgical training* reveals a new path for cultivating clinical empathy empowered by technology. Endres et al.[29] found that medical students improved their ability to recognize patients' microexpressions after training with a microexpression training tool.

## 4.5. Preliminary summary and limitations

In this study, bibliometric and knowledge-graph methods were employed. VOSviewer and CiteSpace were used to analyze 2763 studies on the application of AI to medical education published between 2004 and 2025. The findings revealed three core research themes: technology-driven educational innovation, intelligent clinical training transformation, and ethical risk governance.

Building on the comparative findings, the distinct clustering pattern-specific AI tools (LLMs and robotic surgery) in the English literature versus the New Medicine policy and curriculum reform in the Chinese literature can be explained through established psychological frameworks of technology acceptance and motivation. The focus of English literature on discrete, high-fidelity tools aligns with competence-and autonomy-driven motivation, which is often prevalent in individualistic academic cultures. These tools provide immediate, tangible mastery experiences that enhance self-efficacy and satisfy the need for achievement and precision in specialized tasks. Conversely, the strong link between policy and curriculum reform in China is psychologically underpinned by institutional support and resource allocation, which are critical external motivators. Top-down policies such as the New Medicine do not merely mandate change; they provide essential resources, training, and systemic legitimacy that directly boost the perceived collective efficacy of educators and the sense of belonging within a reformed ecosystem.[30] This approach reduces the anxiety and perceived risk associated with innovation, thereby increasing the perceived usefulness and ease of AI integration at the institutional level, which, in turn, fosters the emotional engagement required for a large-scale curricular overhaul. Therefore, clustering reflects topical and fundamental differences in the primary motivational drivers and support structures that facilitate AI acceptance and implementation across academic spheres.

Wang et al.[31] investigated the factors influencing students' acceptance of AI-driven hybrid learning in business higher education by integrating the technology acceptance model (TAM) with self-determination theory. Self-efficacy, emotional engagement, and university support significantly improved acceptance. University support was a key moderator, strengthening the impact of self-efficacy on acceptance attitudes.

In China, the application of AI in medical education is embedded in the New Medicine initiative and related national reforms, in which policy guidance and institutional capacity play essential roles.[32] Topic clusters around the "New Medicine" and "teaching reform" in Chinese visualization analyses emphasize this system orientation. The findings show that self-efficacy improves perceived ease of use but is not always useful. Together with multimodal literacy and institutional support, these factors shape how students assess the value of AI.[33] In Chinese medical ITS, perceived enjoyment is the strongest antecedent of usefulness and ease of use.[34] In AI‑hybrid learning, self-efficacy, perceived playfulness, and emotional engagement interact with university support to increase acceptance and satisfaction.

The analysis revealed three cores, interlocking thematic pillars that define the field: technology-driven educational transformation, cultivation of clinical capabilities and skills, and governance of ethical risks. These themes were not derived arbitrarily but emerged from clustering high-frequency keywords and their co-occurrence networks, which represent the dominant conceptual frameworks of scholarly discourse. The distinction between the themes follows a logical progression from technological enablers (Theme 1) to their primary application domain in skill development (Theme 2) and finally to the necessary sociotechnical constraints and frameworks (Theme 3) required for sustainable integration. This structure aligns with established models for analyzing technology integration in education, such as the Technological Pedagogical Content Knowledge framework, which emphasizes the intersection of technology, pedagogy, and content. It also reflects the "AI in Education" sociotechnical critique that necessitates examining technical capabilities alongside ethical and governance challenges.[35]

Theme 1: technology-driven medical and educational innovations. This theme encapsulates research that focuses on the development and adoption of specific AI technologies as disruptive tools. Our analysis confirms its centrality in the English-language literature, characterized by deep, sustained investigation into subdivision technologies such as deep learning, LLMs, and robotic surgery. The prominence of keywords like "*machine learning*" and "*generative ai*" is empirically supported by their strong citation bursts after 2020, coinciding with milestones such as the release of AlphaFold2 and ChatGPT. This focus is consistent with global priorities in computational medicine and aligns with funding and publication trends in major Western journals that often prioritize novel algorithmic contributions.

In contrast, while these technologies are used in the Chinese literature, they are consistently framed within a broader policy-led systemic integration strategy. The high frequency and centrality of keywords such as "*New Medicine*" and "*teaching reform*" are direct reflections of top-down national policies. The New Medicine construction strategy, formally launched by China's Ministry of Education, explicitly mandates the deep integration of AI, big data, and other smart technologies into medical education to develop "medical+" composite talents. Subsequent guiding documents, such as the Outline of China's Medical Education Reform (2023), further specify goals, including achieving≥60% coverage of intelligent medical courses.

Theme 2: the intelligent transformation of clinical ability training. This theme represents the convergence of technological innovation and pedagogical goals, focusing on competency development. Both academic spheres exhibited a strong synergy, particularly regarding VR/AR/MR. The co-occurrence of "VR" with "surgical education" in English literature and its presence in Chinese clusters (#3 big data, #6 experimental teaching) underscores a shared recognition of the value of immersive simulation. Empirical studies support this relationship: VR surgical simulators significantly improve resident performance metrics,[8] and in stomatology training, VR simulation increases clinical competency assessment pass rates to 89.4%.[22,23] The shared emphasis on keywords such as "*clinical reasoning*" (English) and "*training of talents*" (Chinese) reveals a common pedagogical shift from knowledge transmission to the cultivation of higher-order cognitive skills. AI-driven clinical scenario simulations and ITSs have been deployed to train diagnostic reasoning and decision-making pathways.[16] This convergence suggests the evolution of medical education as a discipline; core competencies are increasingly defined in terms of cognitive processes rather than mastery of content. The practical implication for curriculum designers is to embed opportunities for deliberate practice of clinical reasoning throughout training, supported by technologies that can provide scalable, standardized, and realistic patient encounters. We also found that the performance of AI tools in the field of traditional Chinese medicine (TCM) education still requires improvement. One factor lies in the challenges associated with translation and cultural context. Certain medical terms have different meanings within the Chinese and Western medical systems, and AI systems often default to interpretations based on the Western framework.[36] Moreover, the linear algorithms commonly used in AI are limited in their ability to incorporate holistic thinking, making it difficult to fully understand the non-linear causal relationships in the diagnostic reasoning process of TCM.[37] Therefore, the integration of AI into TCM education may bring unique challenges, and its pedagogical effectiveness warrants further investigation.

Theme 3: ethical risk and construction of governance systems. This theme addresses the critical sociotechnical challenges associated with AI integration. A comparative timeline analysis revealed a notable phase difference in the intensification of ethical discourse. In the English literature, concentrated discussions on "medical ethics" (cluster #9) emerged as a significant node within 2020–2022, focusing on implicit technical ethics such as algorithm bias in diagnostic AI, patient data privacy in EHR-based learning, and transparency of clinical decision support systems.[20,21] This finding aligns with earlier global ethical guidelines, such as those of the World Health Organization on the ethics and governance of AI for health (2021).

Chinese literature shows a later but rapidly intensifying focus, with explicit keywords such as "*ethical risks*" and clusters like #9 (medical research/ethics) gaining prominence between 2023 and 2024. This period closely aligns with the maturation of China's AI governance framework. Foundational principles were established in key policy documents such as the Ethical Norms for New Generation Artificial Intelligence (2021) and the Management Measures for Algorithmic Recommendation in Internet Information Services (2022). Their application in education requires establishing dynamic ethical review mechanisms for AI tools in academic settings. Thus, the Chinese discourse is characterized by a focus on explicit policy risks and institutional governance mechanisms, such as the role of regional ethics committees in reviewing AI-assisted research by medical students, reflecting a responsive approach to national regulatory developments.

The comparison elucidates the dynamic interplay between global technological discourse (predominant in English-language literature) and localized policy-driven systemic adaptation (predominant in the Chinese literature). Rather than being contradictory, these forces complement each other in shaping the field. The English corpus, which often serves as the primary conduit of global technological innovation, provides a rich repository of technical prototypes and early ethical foresight. The Chinese language corpus, deeply embedded in its national policy context, demonstrates how these technologies are strategically harnessed and institutionalized within a large-scale, reform-oriented education system.

This analysis moves beyond superficial observations to offer a causal explanation; research priorities in each sphere are shaped by distinct ecosystem drivers. English research is heavily influenced by the academic incentive structures of leading journals and the innovation cycles of Silicon Valley and global tech hubs, which favor granular technical exploration.[38] Meanwhile, Chinese research is strongly guided by national strategic plans such as "Healthy China 2030" and the "New Generation Artificial Intelligence Development Plan" (The State Council of China, 2017), which explicitly link AI development to educational modernization. Therefore, the finding that "English research prioritizes specific technologies while Chinese research emphasizes policy-driven integration" is not merely descriptive but empirically grounded in the differing institutional and policy logics of these research


清華大学出版社 Tsinghua University Press | SciOpen

systems.

This perspective explains why cutting-edge technological clusters, such as generative AI, are increasingly linked with discussions on assessment and pedagogy. The focus has shifted from AI's capability to deliver content to its role in creating technology-supported learning environments that are cognitively efficient, engaging, and conducive to the active construction of expertise.

This study underscores a critical and underexplored dimension: the humanization of AI and the cultivation of trust. Beyond functional integration, successful adoption also depends on how medical students and educators perceive, accept, and trust AI systems. The process involves designing AI systems with intuitive, explainable interfaces that foster a sense of partnership rather than opaque automation. Therefore, the psychological and sociological factors influencing trust, such as algorithmic transparency, system reliability, and the alignment of AI decision-making with clinical pedagogy, should be investigated in future research.[39]

Future research can be expanded as follows: First, the data sources can be expanded by broadening the scope of literature collection and conducting multi-language and cross-database mining to create an AI-based comparative medical education system on a global scale. Second, combining qualitative and quantitative methods and linking social network analysis to explore the interaction processes among stakeholders in technology adoption. Third, to address emerging concerns about generative AI, including the erosion of research integrity and critical thinking, we conducted longitudinal studies to establish an empirical basis for adaptive ethics oversight. Interdisciplinary and multiscale research approaches will help shift the AI paradigm of medical education from technology-driven to value-driven, ultimately achieving learner-centered educational innovation.

The findings provide researchers with insights into technical and educational practice dimensions. At the technological research and development level, researchers can focus on optimizing cross-language and cross-cultural algorithmic applications and other technologies, such as training NLP models on Chinese medical texts. At an educational level, the research focus should be on combining technological tools and theories.

In this study, we established a relatively complete graph analysis; nevertheless, some limitations remain. First, the literature sources include Chinese- and English-language publications but do not cover research results related to regions in Asia where medical education development is relatively active, such as Japan and Republic of Korea, which may lead to an inaccurate analysis of regional characteristics. Second, only qualitative data analysis from VOSviewer and CiteSpace was conducted, without including typical case studies (such as in-depth interviews with developers of AI education products), making it impossible to obtain the hidden challenges faced by AI applications. Moreover, the clustering algorithm is automatically filtered and generated, but some marginal topics (such as how AI can be appropriately applied to traditional Chinese medicine) may be overlooked. Finally, this methodology has inherent constraints as a bibliometric analysis; it relies on published literature indexed in selected databases, which introduces an inherent publication bias toward positive results and established research trends, potentially underrepresenting negative findings or innovative but non-mainstream approaches.

## 5. Conclusion

Through a visual analysis, our findings provide a clear knowledge context and cutting-edge guidance for researchers in AI in medical education. In response to the primary aim, our findings empirically clarify how context-specific policy drivers and the global technological discourse interact to shape the evolution of this field. The analysis revealed that these forces operate in a dynamic, complementary relationship, producing convergence and divergence across linguistic contexts. This study provides valuable guidance, serves as a reference for further research, and offers practical applications of AI in medical education. The integration of AI into TCM education is hindered by two factors: the cultural mistranslation of specialized terms, and the fundamental incompatibility between algorithmic, linear logic and TCM's holistic, non-linear diagnostic approach. This inherent tension renders the convergence of these two paradigms a uniquely complex challenge. Therefore, it also highlights the need for more interdisciplinary and cross-cultural collaborative research to address the new challenges posed by technological change.

## Funding

This work was supported by grants from the High Level Traditional Chinese Medicine Key Subject Construction Project of the National Administration of Traditional Chinese Medicine—Evidence-Based Traditional Chinese Medicine (grant no.: zyyzdxk-2023249); the National Administration of Traditional Chinese Medicine (grant no.: YGZXKT2024046); the Hebei Administration of Traditional Chinese Medicine (grant no.: B2025019); and the National Health Commission (grant no.: YLZLXZ24G033).

## Data Availability

Data will be made available by the corresponding author upon reasonable request.

## Author Contribution Statement

**Yu-qing Zhang:** Data curation, methodology, formal analysis, and writing – original draft. **Xuan Wang:** Data curation, software, resources, visualization, validation, and writing – original draft. **Ya-ping He:** Methodology, writing – original draft, and writing – review & editing. **Hong-guo Rong:** Conceptualization, supervision, project administration, funding acquisition, and writing – review & editing. **Yan-yan Meng:** Conceptualization, supervision, and writing – review & editing.

## Use of AI Statement

None.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this study.

## Electronic Supplementary Material

The online version contains supplementary material available at https://doi.org/10.26599/eCMTA.2026.9570027.

### References

[1] Song W, Hou X, Li S, et al. An intelligent virtual standard patient for medical students training based on oral knowledge graph. *IEEE Trans Multimed*. 2023;25:6132−6145.
[2] Kennedy GAL, Pedram S, Sanzone S. Improving safety outcomes through medical error reduction *via* virtual reality-based clinical skills training. *Saf Sci*. 2023;165:106200.
[3] Sunar AS, Khalid MS. Natural language processing of student’s feedback to instructors: a systematic review. *IEEE Trans Learn Technol*. 2024;17:741−753.
[4] Wang J, Li Y, Chen X, Wang C. Before and after: how did the pandemic reshape virtual laboratory research? *Sage Open*. 2025;15(2):21582440251339961.
[5] Du M, Li H, Guo H, Zhang X, Rong H, Hao X. Bibliometric analysis and key messages of integrating Chinese and Western Medicine for COVID-19. *Heliyon*. 2024;10(6):e27293.
[6] Agbareia R, Omar M, Soffer S, Glicksberg BS, Nadkarni GN, Klang E. Visual-textual integration in LLMs for medical diagnosis: a preliminary quantitative analysis. *Comput Struct Biotechnol J*. 2025;27:184−189.
[7] Popov V, Mateju N, Jeske C, Lewis KO. Metaverse-based simulation: a scoping review of charting medical education over the last two decades in the lens of the ‘marvelous medical education machine’. *Ann Med*. 2024;56(1):2424450.
[8] Mao RQ, Lan L, Kay J, et al. Immersive virtual reality for surgical training: a systematic review. *J Surg Res*. 2021;268:40−58.
[9] Basdogan C, Sedef M, Harders M, Wesarg S. VR-based simulators for training in minimally invasive surgery. *IEEE Comput Graph Appl*. 2007;27(2):54−66.
[10] Soroudi D, Gozali A, Knox JA, et al. Comparing provider and ChatGPT responses to breast reconstruction patient questions in the electronic health record. *Ann Plast Surg*.

2024;93(5):541−545.
[11] Lee H. The rise of ChatGPT: exploring its potential in medical education. *Anat Sci Educ*. 2024;17(5):926−931.
[12] Yalamanchili A, Sengupta B, Song J, et al. Quality of large language model responses to radiation oncology patient care questions. *JAMA Netw Open*. 2024;7(4):e244630.
[13] Zhang J, Yu H, Zheng X, et al. Deep-learning-based survival prediction of patients with lower limb melanoma. *Discov Oncol*. 2023;14(1):218.
[14] Fairén M, Moyés J, Insa E. VR4Health: personalized teaching and learning anatomy using VR. *J Med Syst*. 2020;44(5):94.
[15] Zhou Y, Hou J, Liu Q, et al. VR/AR technology in human anatomy teaching and operation training. *J Healthc Eng*. 2021;2021:9998427.
[16] Derevianko A, Pizzoli SFM, Pesapane F, et al. The use of artificial intelligence (AI) in the radiology field: what is the state of doctor-patient communication in cancer diagnosis? *Cancers*. 2023;15(2):470.
[17] Li Q, Qin Y. AI in medical education: medical student perception, curriculum recommendations and design suggestions. *BMC Med Educ*. 2023;23(1):852.
[18] Melazzini L, Bortolotto C, Brizzi L, et al. AI for image quality and patient safety in CT and MRI. *Eur Radiol Exp*. 2025;9(1):28.
[19] Park C, You SC, Jeon H, Jeong CW, Choi JW, Park RW. Development and validation of the radiology common data model (R-CDM) for the international standardization of medical imaging data. *Yonsei Med J*. 2022;63(Suppl):S74−S83.
[20] Alanazi B, Butler-Henderson K, Alanazi M. Perceptions of healthcare professionals about the adoption and use of EHR in Gulf Cooperation Council countries: a systematic review. *BMJ Health Care Inform*. 2020;27(1):e100099.
[21] Basil NN, Ambe S, Ekhator C, Fonkem E. Health records database and inherent security concerns: a review of the literature. *Cureus*. 2022;14(10):e30168.
[22] Huang TK, Yang CH, Hsieh YH, Wang JC, Hung CC. Augmented reality (AR) and virtual reality (VR) applied in dentistry. *Kaohsiung J Med Sci*. 2018;34(4):243−248.
[23] Coxe FR, Stauffer TP, Ast MP. Virtual reality simulation in orthopedic surgery education improves immediate procedural skill and knowledge acquisition, but evidence on cost-effectiveness and skill retention remains lacking. *Curr Rev Musculoskelet Med*. 2025;18(10):363−378.
[24] Mohammad B, Supti T, Alzubaidi M, et al. The pros and cons of using ChatGPT in medical education: a scoping review. *Stud Health Technol Inform*. 2023;305:644−647.
[25] Zheng K, Shen Z, Chen Z, Che C, Zhu H. Application of AI-empowered scenario-based simulation teaching mode in cardiovascular disease education. *BMC Med Educ*. 2024;24(1):1003.
[26] KIM HY. Development and usability assessment of virtual reality- and haptic technology-based educational content for perioperative nursing education. *Healthcare (Basel)*. 2024;12(19):1947.
[27] Milano CE, Hardman JA, Plesiu A, Rdesinski RE, Biagioli FE. Simulated electronic health record (Sim-EHR) curriculum: teaching EHR skills and use of the EHR for disease management and prevention. *Acad Med*. 2014;89(3):399−403.
[28] Cooper A, Rodman A. AI and medical education: a 21st-century *Pandora*'s box. *N Engl J Med*. 2023;389(5):385−387.
[29] Endres J, Laidlaw A. Micro-expression recognition training in medical students: a pilot study. *BMC Med Educ*. 2009;9:47.
[30] Wang J, Li Y, Chen X, Wang C. Before and after: how did the pandemic reshape virtual laboratory research? *SAGE Open*. 2025;15(2):1–23.
[31] Wang S, Sun Z, Wang H, Yang D, Zhang H. Enhancing student acceptance of artificial intelligence-driven hybrid learning in business education: interaction between self-efficacy, playfulness, emotional engagement, and university support. *Int J Manag Educ*. 2025;23(2):101184.
[32] Peng J, Zhang H, Tu X, et al. Effectiveness of AI-assisted medical education for Chinese undergraduate medical students: a meta-analysis. *BMC Med Educ*. 2025;25(1):1207.
[33] Yan Y, Wu B, Pi J, Zhang X. Perceptions of AI in higher education: insights from students at a top-tier Chinese University. *Educ Sci*. 2025;15(6):735.
[34] Han FF. Factors affecting Chinese undergraduate medical students' behavioural intention and actual use of intelligent tutoring systems. *Australas J Educ Technol*. 2024;40(2):76−93.
[35] Han S, Liu M, Pan Z, Cai Y, Shao P. Making FAQ chatbots more inclusive: an examination of non-native English users' interactions with new technology in massive open online courses. *Int J Artif Intell Educ*. 2023;33(3):752−780.
[36] Tseng LW, Lu YC, Tseng LC, Chen YC, Chen HY. Performance of ChatGPT-4 on Taiwanese traditional Chinese medicine licensing examinations: cross-sectional study. *JMIR Med Educ*. 2025;11:e58897.
[37] Han C, Yang G, Li H, Zhu L, Feng M. Tuning and clinical application of large language models in Traditional Chinese Medicine: scoping review. *Chin Med*. 2026;21(1):71.
[38] Ejaz H, McGrath H, Wong BL, Guise A, Vercauteren T, Shapey J. Artificial intelligence and medical education: a global mixed-methods study of medical students' perspectives. *Digit Health*. 2022;8: 20552076221089099.
[39] Jung I, Sasaki T, Latchem C. A framework for assessing fitness for purpose in open educational resources. *Int J Educ Technol High Educ*. 2016;13(1):3.