人工智能在医学教育中的双语文献计量学分析:中英文研究热点与趋势比较

王文涛[1],潘俊颐[1],鲁子越[1],朱文彦[2],沈 健[3],李 响[3,4]

重庆医科大学:1.莱斯特大学联合学院;3.附属第一医院心血管内科;4.教务处,重庆 400016;
2.重庆医药高等专科学校科技产业发展中心,重庆 401331

【摘要】目的 比较 2015 — 2025 年人工智能医学教育中英文文献的热点与演化趋势。**方法** 检索 CNKI、万方、维普、Web of Science、PubMed 和 Scopus,纳入中文文献 2 890 篇、英文文献 4 131 篇;采用 CiteSpace、VOSviewer 和 BERTopic 分析发文趋势、合作网络、期刊分布、关键词及主题演化。**结果** 研究持续增长,热点由 VR/AR、模拟训练转向机器学习、自然语言处理、生成式 AI 和大语言模型;英文文献更关注教学应用、学习成效与评价框架,中文文献更聚焦教学改革和课程建设。**结论** 中英文文献在研究取向、知识载体和合作网络方面存在差异。我国应加强成效评价、伦理治理、教师数字素养和跨学科协作。

【关键词】人工智能;医学教育;文献计量学;研究热点;发展趋势

【中图分类号】R-4;G642.0 **【文献标志码】**A **【文章编号】**1004-5287(2026)05-0625-13

【DOI】10.13566/j.cnki.cmet.cn61-1317/g4.202605007

A bilingual bibliometric analysis of artificial intelligence in medical education: Comparing research hotspots and trends in Chinese and English literature

WANG Wentao[1], *PAN Junyi*[1], *LU Ziyue*[1], *ZHU Wenyan*[2], *SHEN Jian*[3], *LI Xiang*[3,4]

Chongqing Medical University: 1. *Chongqing Medical University-University of Leicester Joint Institute*;
3. *Department of Cardiovascular Medicine*, *the First Affiliated Hospital*; 4. *Academic Affairs Office*, *Chongqing* 400016, *China*;
2. *Technology and Industry Development Center*, *Chongqing Medical and Pharmaceutical College*, *Chongqing* 401331, *China*

【Abstract】Objective To compare the research hotspots and evolutionary trends of Chinese and English literature on artificial intelligence in medical education from 2015 to 2025. **Methods** CNKI, Wanfang, VIP, Web of Science, PubMed, and Scopus were searched, and 2 890 Chinese articles and 4 131 English articles were included. CiteSpace, VOSviewer, and BERTopic were used to analyze publication trends, collaboration networks, journal distribution, keywords, and thematic evolution from multiple perspectives. **Results** Research in this field showed continuous growth during the study period. The research hotspots shifted from VR/AR and simulation-based training to machine learning, natural language processing, generative AI, and large language models. English literature paid greater attention to teaching applications, learning outcomes, and assessment frameworks, whereas Chinese literature focused more on teaching reform and curriculum development. **Conclusion** Differences existed between Chinese and English literature in research orientation,

基金项目:重庆市高等教育教学改革研究项目(232047);重庆市高等教育学会高等教育科学研究课题(cqgj25zx10C);2025 年重庆医科大学校级教育教学改革研究项目重大项目(JY20250202);重庆医科大学附属第一医院研究生教育教学改革项目(CYYY-YJSGXM-202309)

收稿日期:2026-02-10

作者简介:王文涛,本科,研究方向为人工智能赋能医学教育、医学教育教学改革。

通信作者:李响,博士,副教授、副主任医师,硕士研究生导师,研究方向为新医科整合医学、教学方法改革和教学效果评估、人工智能赋能医学教育。E-mail:lzd9288@163.com

knowledge carriers, and collaboration networks. China should further strengthen outcome evaluation, ethical governance, teachers' digital literacy, and interdisciplinary collaboration in the field of artificial intelligence in medical education.

【**Keywords**】artificial intelligence; medical education; bibliometrics; research hotspots; development trends

近年来,人工智能(artificial intelligence, AI)在医学教育领域的应用迅速扩展,正在深刻改变医学人才培养的理念、教学方式与评价模式。从早期的虚拟现实(virtual reality,VR)与增强现实(augmented reality,AR)模拟训练[1-4],到学习分析(learning analytics, LA)与大语言模型(large language models, LLMs)支持的智能教学[5-9],AI已逐步渗透至临床技能训练、课程设计、学习支持、形成性评价与反馈等多个教学环节,成为推动医学教育数字化与智能化转型的重要力量。因此,系统研究人工智能在医学教育领域的研究热点与发展趋势,不仅有助于把握学术前沿,也有助于为医学教育实践、教学改革与技术应用提供参考。国家层面出台的《新一代人工智能发展规划》《教育信息化2.0行动计划》等政策文件持续推动人工智能与教育深度融合[10-11],"新医科"建设与"双一流"建设背景下的高等教育改革也进一步强调医学教育与信息技术、工程技术的交叉创新以及高水平人才培养[12-13]。在此背景下,人工智能如何更有效地融入医学教育实践、推动课程改革与评价优化,已成为值得系统关注的重要议题。

随着我国教育数字化战略持续推进,人工智能与高等教育的深度融合不断加快。通过文献分析系统梳理人工智能在医学教育中的研究格局,不仅有助于把握国际发展趋势,也有助于为我国医学教育技术发展、课程改革与教学评价体系完善提供参考。现有研究中,国外学者普遍基于科学引文索引数据库(Web of Science, WoS)、生物医学文献数据库(PubMed)和Scopus数据库等国际权威数据源,采用文献计量学方法探讨人工智能在医学教育中的演进。例如,Li等[14]通过对2000 — 2024年英文文献的计量分析发现,自2018年以来研究产出与引用量显著增长,研究重点已由医学影像与仿真训练转向智能模型与临床决策支持。Preiksaitis等[15]指出,生成式人工智能(generative artificial intelligence, GAI)的兴起正在重塑教学模式、评价体系与教师角色。国内研究则多依托中国知网(China National Knowledge Infrastructure, CNKI)、万方数据库(Wanfang Database)与维普数据库(VIP Database),聚焦智慧教学平台、虚拟仿真实验及生成式AI教学应用。封纯等[16]比较中外VR/AI教育研究后认为,国际研究在学术影响力与跨学科融合方面更具优势,而国内研究更强调教学改革与实践创新。总体来看,现有研究仍存在3个方面不足:第一,多数研究依赖单一数据库或单语种文献,缺乏多源融合与跨语种比较;第二,对中英文文献的研究热点、主题演化与结构差异的系统分析不足;第三,对2015 — 2025年生成式AI崛起背景下的研究演化缺乏连续性追踪。这些局限使学界尚难从整体视角把握人工智能医学教育研究的全球格局与演化特征。

基于此,本研究旨在系统揭示2015 — 2025年人工智能在医学教育领域的研究热点、合作格局与演化趋势,并在中英文文献中进行双语比较,以回答以下问题:第一,过去十年中人工智能医学教育研究的总体产出特征与合作网络如何演变;第二,中英文文献中的研究热点、关键词共现与主题聚类呈现出怎样的异同;第三,不同语言文献在研究取向、知识结构与主题演化上的差异,对我国医学教育技术发展、课程建设、教学评价与跨学科协作机制完善有何启示。为此,本研究整合CNKI、万方与维普三大中文数据库,以及Web of Science、PubMed和Scopus三大英文数据库,共获取7 000余篇文献(2015年1月至2025年10月),并从宏观、中观与微观三个层面构建综合分析框架。本研究重点聚焦生成式AI与LLMs在医学教育中的新兴应用,在中文与英文文献之间进行双语比较,分析两类文献在研究主题、知识结构与演化路径上的差异,并进一步探讨其对医学教育实践、课程改革、教学评价与教师发展机制的启示。

1 材料与方法

1.1 数据来源与检索策略

本研究基于3大中文数据库(CNKI、万方、维普)与3大英文数据库(Web of Science、PubMed、Scopus),系统检索2015年1月1日至2025年10月8日期间人工智能在医学教育领域的相关文献。为确保检索的全面性与可重复性,本研究构建统一检索框架,并根据不同数据库的字段规则进行适配。中文数据库以"医学教育""医学生"等为教育主题词,以"人工智能""AI""虚拟现实""增强现实""混合现实""自然语言处理""知识图谱""ChatGPT""大语言模型"等为技术主题词,在题名、关键词或主题字段中联合检索;

英文数据库以“medical education”“medical students”“health professions education”等为教育主题词，以“artificial intelligence”“AI”“virtual reality”“augmented reality”“mixed reality”“natural language processing”“knowledge graph”“ChatGPT”“large language model”“LLM”等为技术主题词，在题名、摘要、关键词或主题字段中进行组合检索。考虑到生成式人工智能相关术语在2023年后快速演化，部分研究逐渐以具体模型名称或新概念替代上位技术术语进行表述，本研究在检索策略中优先采用“人工智能”“大语言模型”等上位概念，并结合代表性相关术语进行检索，以提高检索框架的一致性与可重复性。

1.2　文献筛选与质量控制

各库初检结果经统一汇总后得到中文文献10 937篇，英文文献19 075篇。使用EndNote 2025.1与COOC 20.5进行自动去重（题名、DOI、作者字段匹配），并对疑似重复、缺失字段或主题不匹配的文献进行人工核查。为确保研究质量，本研究限定纳入同行评审期刊论文，排除会议摘要、评论、短讯、书评等非研究性文献，以及信息不完整或与医学教育无关的人工智能技术论文。经筛选后，最终纳入中文文献2 890篇、英文文献4 131篇（图1）。

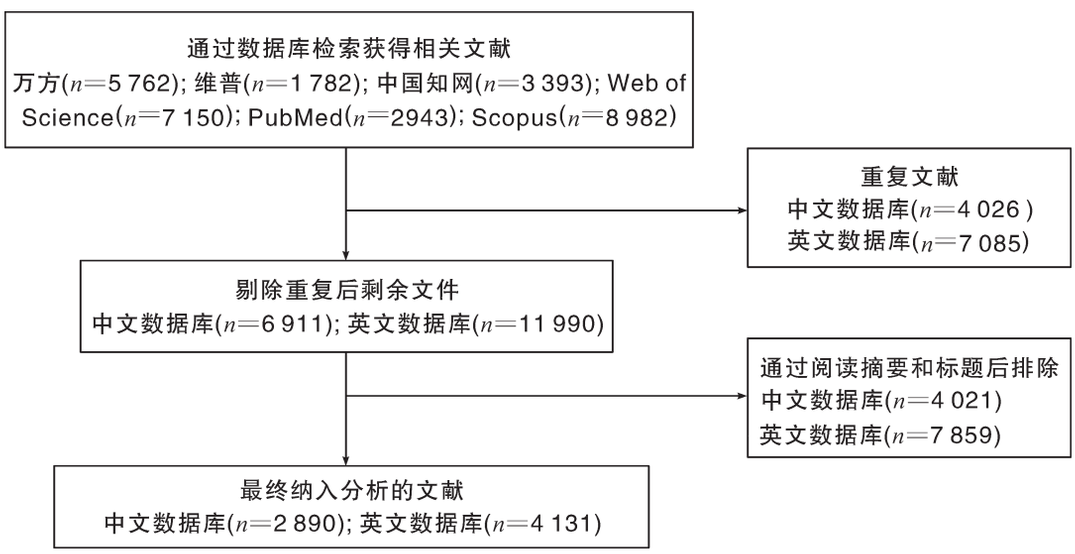


图1　文献筛选流程图

1.3　研究工具与分析方法

在数据分析方面，本研究使用Python 3.13对整理后的文献数据进行数据清洗、汇总统计与可视化处理，并结合VOSviewer 1.6.20、CiteSpace 6.3.R1、Pajek 6.01及COOC 20.5等工具，开展文献产出趋势、国家与机构合作网络、作者合作网络、关键词共现与突现分析及主题聚类分析[17-20]。其中，作者合作网络基于VOSviewer构建，并结合节点数、连边数、聚类数及总连接强度等指标进行分析。为确保多层次分析框架的可重复性与规范性，本研究依据现有文献计量学研究的宏观、中观、微观范式，明确各层面的指标选取方式：宏观层面包括国家/地区发文量、机构发文量及国家间合作网络等指标；中观层面包括期刊发文量、作者发文量及作者合作网络结构；微观层面包括关键词共现、关键词突现、聚类与主题演化等指标。以上指标体系参考了已有文献计量研究的分类方法，并结合本研究的数据库特征进行设定[21-22]。

此外，为揭示中文与英文文献中人工智能医学教育研究的主题演化，本研究采用自然语言处理技术构建双语主题模型。将文献题名、关键词和摘要拼接后的文本作为统一语料，使用BERTopic分别对中文与英文文献进行建模[23]。中文语料设置language=“chinese”，文本嵌入模型采用shibing624/text2vec-base-chinese；英文语料设置language=“english”，文本嵌入模型采用sentence-transformers/all-mpnet-base-v2，并使用CountVectorizer(stop_words=“english”, ngram_range=(1, 2), min_df=5)构建文本特征表示。中英文模型均设置calculate_probabilities=True，以输出文档-主题概率分布，用于后续分阶段主题汇总与演化分析。在文本预处理方面，中文语料采用分词与停用词过滤，英文语料进行统一清洗并去除停用词；中英文语料均仅保留中文或英文字母及数字，统一转换为小写，并过滤长度小于2的词项。

本研究以2015—2025年全部文献作为整体语料一次性训练主题模型，再依据年度文献量分布将研究期划分为4个连续阶段，即2015—2017年、2018—

2020年、2021 — 2023年和2024 — 2025年，以增强各阶段样本量的均衡性与跨阶段比较的可解释性。随后，基于文档-主题概率分布对各阶段主题进行汇总，并结合主题嵌入相似度构建主题演化关系，绘制主题演化桑基图。在主题演化结果聚合过程中，中文模型的跨阶段主题匹配相似度阈值设为0.60，英文模型设为0.55；中英文模型均设定每个源主题仅保留1个最优匹配主题，每阶段最多展示10个主题节点，并过滤占比低于0.08的细小连线。本研究未预设主题数量，由模型基于文本嵌入表示、UMAP降维和HDBSCAN聚类自动识别主题。除文中已说明的嵌入模型、向量化参数及主题演化聚合参数外，umap_model、hdbscan_model、min_topic_size和representation_model等均采用BERTopic默认配置。所得主题结果主要用于揭示总体主题结构及阶段性演化趋势，不作为主题边界的唯一划分依据。

CiteSpace设置时间范围为2015 — 2025年，时间切片为1年，提取题名、摘要和关键词（含keywords plus），并在cosine强度与within-slice范围内以g-index（k=25）进行节点选取。在期刊分析中，本研究依据布拉德福德定律，以文献计量学研究中常用的“累计发文量约占总量的1/3”作为核心区划分标准，对中英文数据库样本中的期刊进行排序与核心区识别。

2 人工智能在医学教育中的研究热点与前沿

2.1 宏观层面

2.1.1 文献产出分析

图2展示了2015年1月至2025年10月人工智能在医学教育领域的年度发文量及累计发文量。英文数据库的年度发文量在2015 — 2018年保持较低水平；2019年起逐步增加；2022年后增长幅度明显扩大；2023 — 2024年发文量达到较高水平；截至2025年10月，发文量已接近或超过2024年全年水平。中文数据库的年度发文量整体呈上升趋势，2015 — 2020年增长较为平稳；2021年以后增幅加快；2023 — 2025年持续上升。

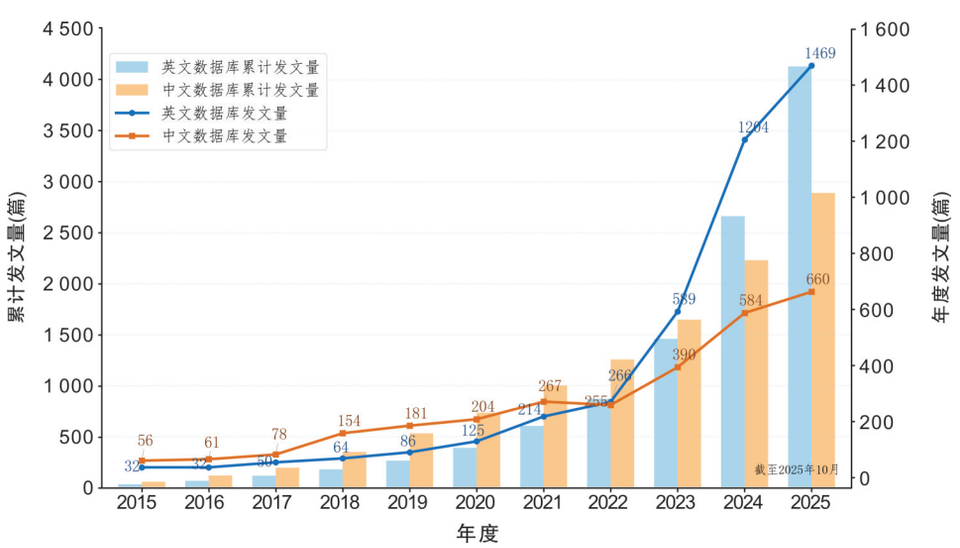


图2 2015—2025年人工智能在医学教育领域研究在不同数据库中的年度发文量与累计发文量统计图

2.1.2 国家合作分析

图3展示了2015 — 2025年人工智能在医学教育领域的国家科研产出分布及合作网络结构。结合表1来看，美国的发文量为1 375篇，占22.72%，在合作网络中的节点规模和连接度均较高；中国发文量为559篇，占9.24%，位居第2。英国、加拿大、德国、印度、澳大利亚等国家发文量也较高。合作网络中，美国、中国、英国、加拿大、德国、澳大利亚等国家节点规模较大，且与其他国家之间存在较多连线。沙特阿拉伯、阿联酋、新加坡、韩国、日本等国家也出现在合作网络中，并与部分国家形成合作联系。

表1 2015 — 2025年英文数据库中人工智能医学教育研究相关出版物数量排名前10的国家

排名	国家	出版物数量（篇）	占比（%）
1	United States	1 375	22.72
2	China	559	9.24
3	United Kingdom	410	6.77
4	Canada	309	5.11
5	Germany	264	4.36
6	India	205	3.39
7	Turkey	178	2.94
8	Saudi Arabia	172	2.84
9	Australia	172	2.84
10	Italy	128	2.12

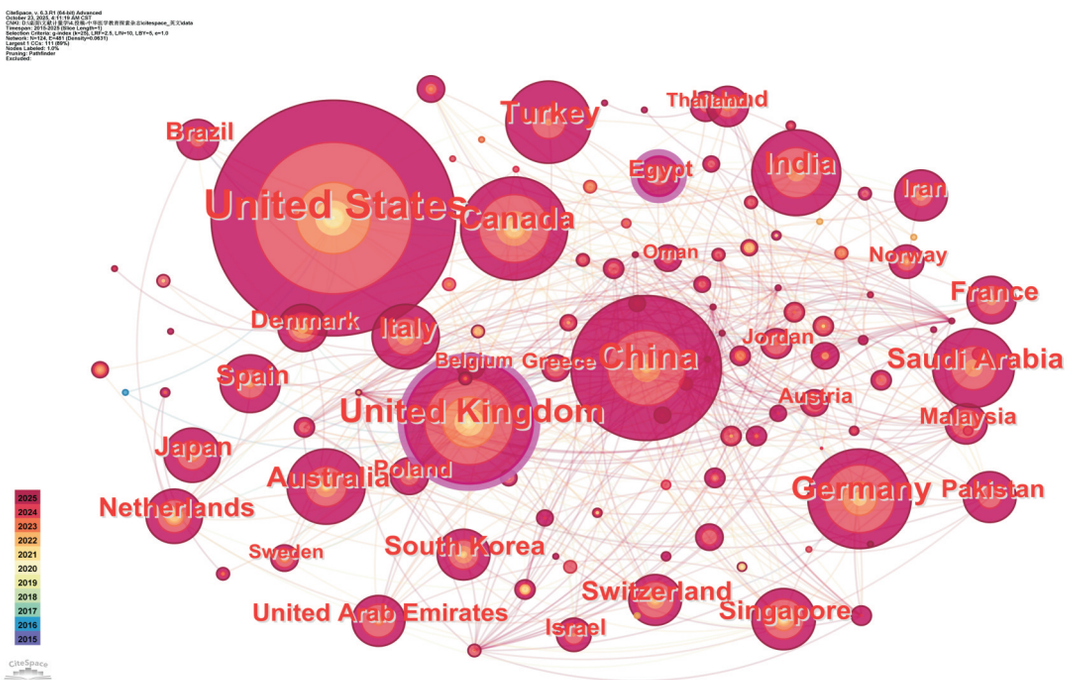


图 3　2015 — 2025 年各国在人工智能医学教育研究中的产出与合作交流图

2.1.3　机构发文量分析

如表 2 所示，英文数据库中哈佛大学发文量为 103 篇，多伦多大学、斯坦福大学等机构分列其后。中文数据库中，首都医科大学发文量为 110 篇，四川大学为 98 篇，均位于前列。中文数据库的前 10 机构整体发文量高于英文数据库中排名前 10 的机构。英文数据库发文量排名前 10 的机构均非中国高校。

表 2　2015—2025 年中英文数据库中人工智能医学教育研究相关出版物数量排名前 10 的机构

排名	发文机构（中文）	出版物数量（篇）	占比（%）	发文机构（英文）	出版物数量（篇）	占比（%）
1	首都医科大学	110	3.05	Harvard University	103	4.41
2	四川大学	98	2.71	University of Toronto	90	3.86
3	西安交通大学	89	2.46	Stanford University	66	2.83
4	北京大学	80	2.22	Imperial College London	61	2.61
5	中国医学科学院北京协和医学院	79	2.19	University of Michigan	59	2.53
6	海军军医大学	73	2.02	National University of Singapore	57	2.44
7	浙江大学	65	1.80	Mayo Clinic	57	2.44
8	上海交通大学	62	1.72	McGill University	53	2.27
9	空军军医大学	59	1.63	McMaster University	50	2.14
10	重庆医科大学	57	1.58	Johns Hopkins University	49	2.10

2.2　中观层面

2.2.1　期刊发文量分析

2015 — 2025 年人工智能在医学教育研究中，中英文数据库发文量排名前 10 的期刊分布情况，如表 3 所示。英文数据库中，*BMC Medical Education* 发文量最高，为 335 篇，占 13.13%；其次为 *JMIR Medical Education*，发文量为 165 篇，占 6.47%；*Cureus* 发文量为 154 篇，占 6.04%。此外，*Medical Teacher*、*Academic Medicine*、*Medical Science Educator* 等期刊发文量也位居前列。中文数据库中，《中国继续医学教育》的发文占比为 6.51%，排名第 1；《中国医学教育技术》占比为 5.61%，位列第 2；《中华医学教育探索杂志》《医学教育研究与实践》《中国高等医学教育》等期刊也位居发文量前列。

按照上述核心区划分标准，中文数据库中发文量排名前 10 位的期刊累计占比已接近核心区阈值，在进一步纳入《医学信息学杂志》和《中国毕业后医学教育》后，累计占比达到 33.85%，构成中文文献的核心期刊群。英文数据库中，发文量排名前 5 的期刊累计占比为 33.99%，对应的核心期刊包括 *BMC Medical Education*、*JMIR Medical Education*、*Cureus*、*Medical Teacher* 与 *Academic Medicine*。

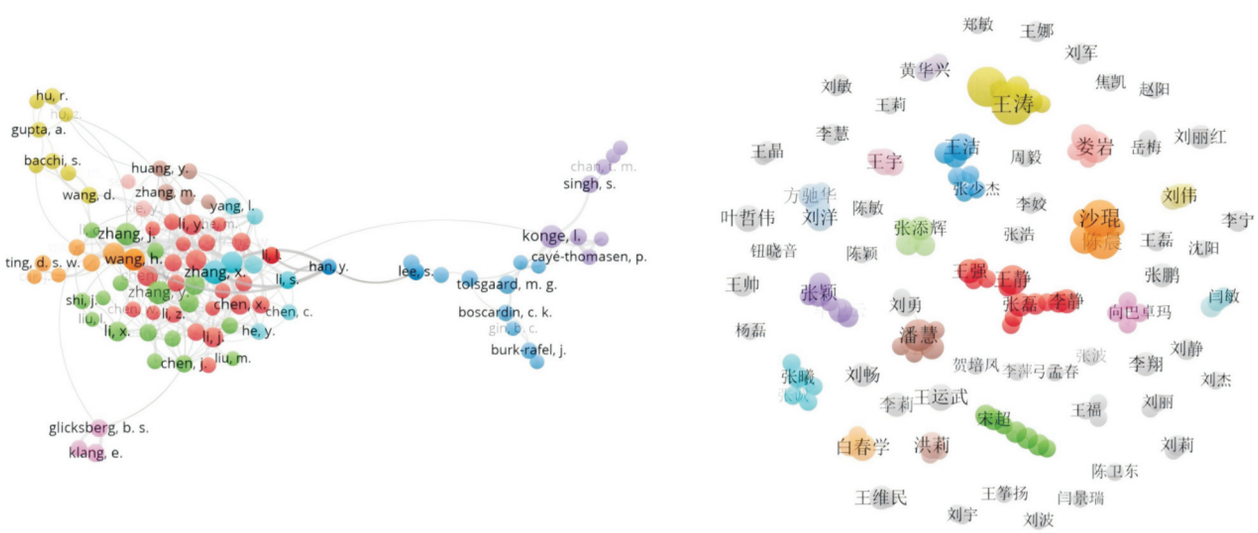
表 3 2015 — 2025 年中英文数据库中人工智能医学教育研究相关文献发文量排名前 10 的期刊

排名	中文数据库期刊	出版物数量（篇）	占比（%）	英文数据库期刊	出版物数量（篇）	占比（%）
1	中国继续医学教育	188	6.51	*BMC Medical Education*	335	13.13
2	中国医学教育技术	162	5.61	*JMIR Medical Education*	165	6.47
3	中华医学教育探索杂志	110	3.81	*Cureus*	154	6.04
4	医学教育研究与实践	81	2.80	*Medical Teacher*	135	5.29
5	中国高等医学教育	79	2.73	*Academic Medicine*	78	3.06
6	中华医学教育杂志	77	2.66	*Medical Science Educator*	60	2.35
7	继续医学教育	69	2.39	*Journal of Medical Internet Research*	59	2.31
8	基础医学教育	50	1.73	*Journal of Surgical Education*	58	2.27
9	卫生职业教育	49	1.70	*Frontiers in Medicine*	51	2.00
10	教育教学论坛	40	1.38	*Anatomical Sciences Education*	51	2.00

2.2.2 作者合作分析

图 4 展示了人工智能在医学教育领域的中英文作者合作网络结构。中文数据库中，王涛、沙琼、陈晨、娄岩等作者形成多个相对紧密的合作群体，部分作者位于不同集群的连接位置；参考普赖斯定律并结合作者发文分布特征，本研究在中文数据库中以 4 篇作为作者纳入阈值，共有 143 位作者进入合作网络，形成 117 条连边，总连接强度为 269，聚类数为 62。

英文数据库中，合作网络同样呈现多中心结构，其中 Zhang、Wang、Chen 等作者集群较为突出，并形成以 Konge、Tolsgaard 等为中心的多个合作子网络。参考普赖斯定律并结合作者发文分布特征，本研究在英文数据库中以 6 篇作为作者纳入阈值，共有 93 位作者进入合作网络，形成 466 条连边，总连接强度为 682，聚类数为 10。图 4 所示作者均为满足阈值条件并进入网络的高发文作者子集，相关结果主要反映核心作者层面的合作结构。



（a）英文数据库 （b）中文数据库

图 4 2015 — 2025 年中英文数据库中人工智能医学教育研究作者合作网络图

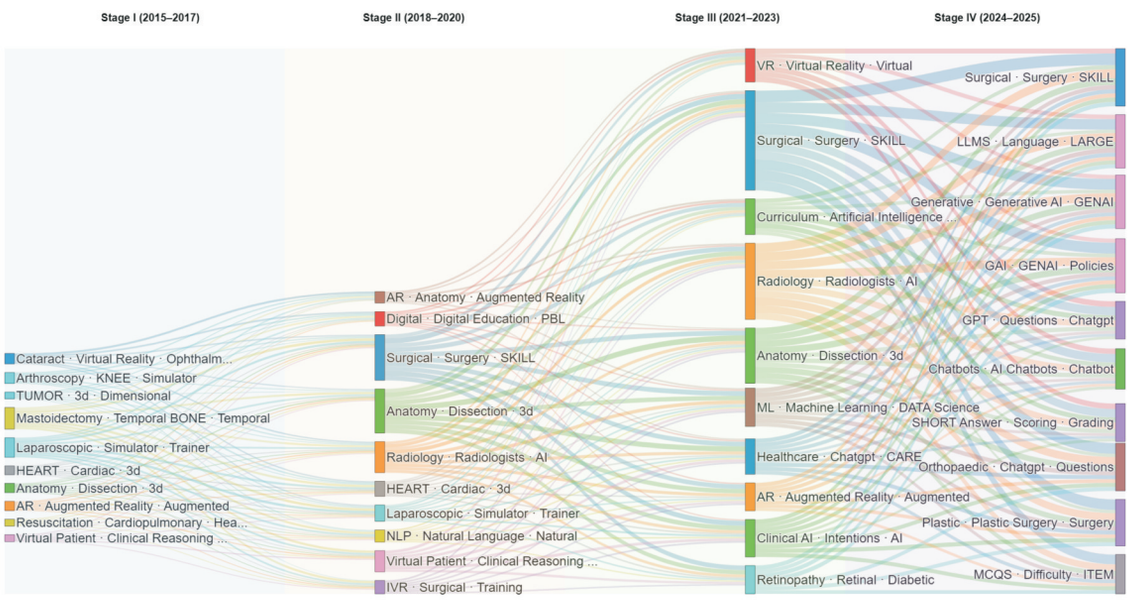
2.3 微观层面

2.3.1 研究主题演化分析

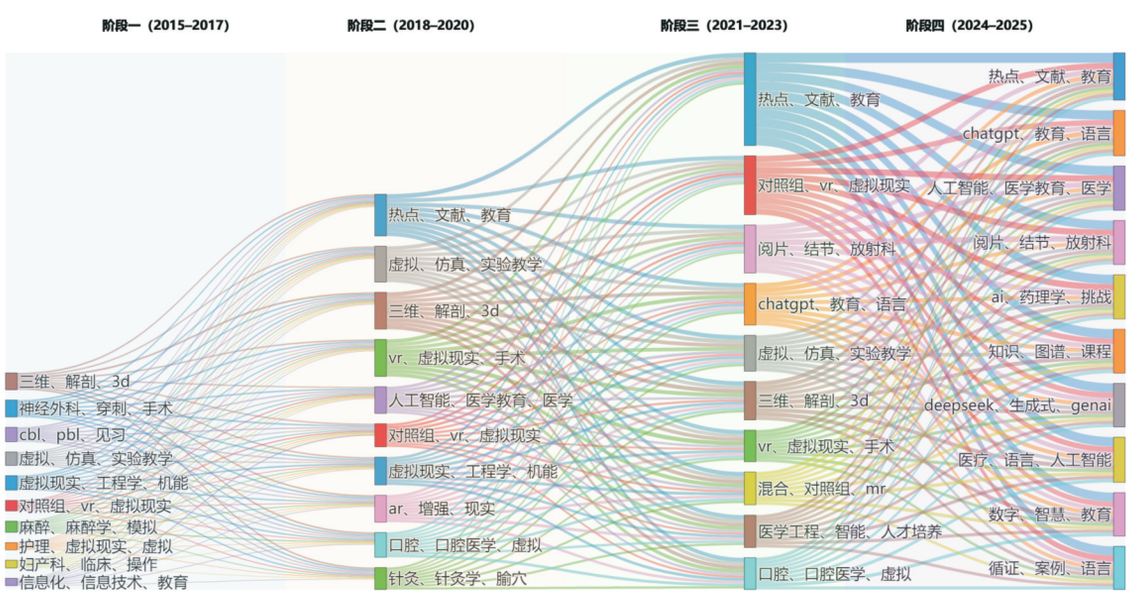
基于 BERTopic 对全语料进行统一建模，并结合文档-主题概率的分阶段汇总与主题嵌入相似度匹配，得到英文数据库的主题演化结果，如图 5(a)所示。阶段Ⅰ（2015 — 2017 年）主题主要集中于“虚拟现实（VR）”“模拟训练（simulation）”“手术技能（surgical skill）”；阶段Ⅱ（2018 — 2020 年）出现“增强现实（AR）”“机器学习（ML）”“自然语言处理（NLP）”等主题；阶段Ⅲ（2021 — 2023 年）扩展至机器学习、临床

AI、医疗照护等方向；阶段Ⅳ（2024 — 2025年）则新增“ChatGPT”“LLMs”“generative AI”“assessment”“question”“scoring”等主题。

中文数据库的主题演化结果如图5(b)所示。阶段Ⅰ（2015—2017年）的主要主题包括“虚拟现实”“三维教学”“实验训练”；阶段Ⅱ（2018—2020年）出现“人工智能”“增强现实”“教学改革”等主题；阶段Ⅲ（2021—2023年）进一步涉及“ChatGPT”“医学工程”“人才培养”等；阶段Ⅳ（2024—2025年）主要以“ChatGPT”“生成式AI”“教学设计”“课程改革”“医学影像”等主题为主，并出现“针灸”“针灸学”“腧穴”等医学专业相关主题。



（a）英文数据库



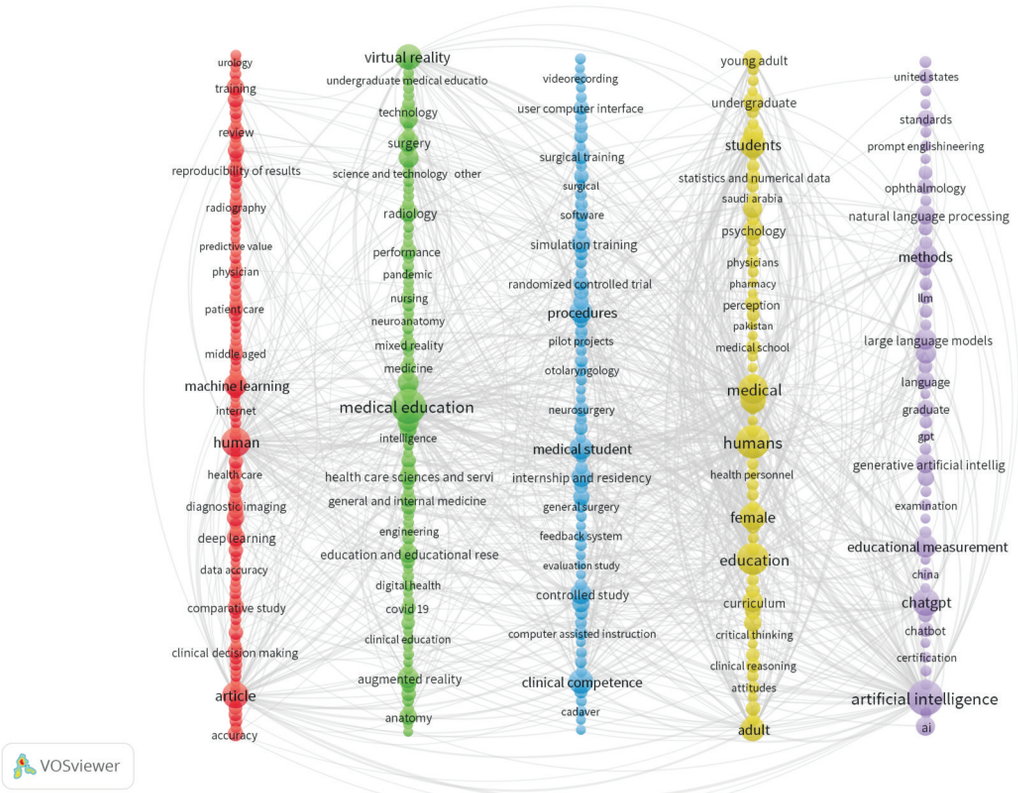
（b）中文数据库

图5　2015—2025年中英文数据库中人工智能医学教育研究主题演变图

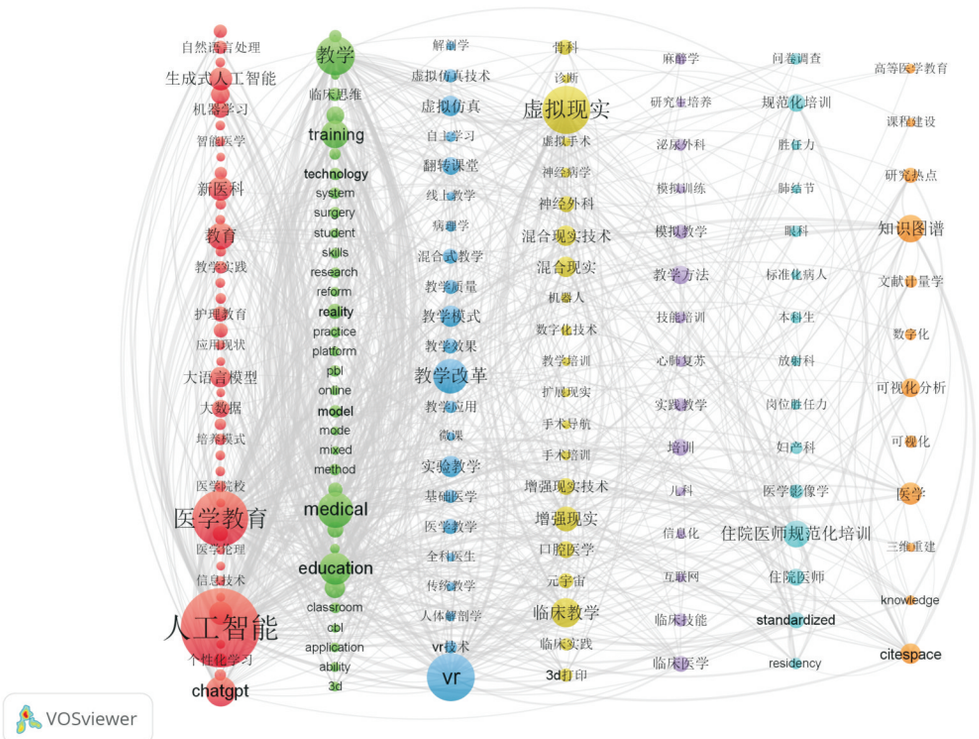
2.3.2　关键词共现与突现分析

英文数据库关键词共现网络如图6(a)所示。关键词网络以“artificial intelligence”“medical education”“virtual reality”“clinical competence”等为核心节点，形成多个紧密关联的主题集群。人工智能相关技术（如“machine learning”“deep learning”“ChatGPT”“large language models”）与医学教育主题高度共现。中文数据库关键词共现网络如图6(b)所示，主

要关键词包括“人工智能”“医学教育”“教学改革”“虚拟现实”“情景模拟”等。其中,“人工智能”与“医学教育”“教学改革”“虚拟现实”等关键词之间存在较多共现联系。



(a) 英文数据库



(b) 中文数据库

图 6 2015 — 2025 年中英文数据库中人工智能医学教育研究主要关键词共现网络

英文数据库共识别出 13 个聚类，主要包括“clinical competence”“augmented reality”“three dimensional”等主题，以及“diagnostic accuracy”“natural language processing”“digital health”“GPT-4”等关键词，如图 7(a)所示。中文数据库共识别出 12 个聚类，主要涉及“人工智能”“增强现实”“教学模式”“知识图谱”等主题，如图 7(b)所示。



（a）英文数据库　　（b）中文数据库

图 7　2015 — 2025 年不同数据库中人工智能医学教育研究主要关键词聚类可视化图

在关键词突现分析中，本研究使用 CiteSpace 6.3.R1 对 2015 — 2025 年间的文献进行突现检测。软件自动识别出 503 个关键词节点，并根据 Kleinberg 突现模型计算关键词的突现强度（Strength）、突现起始年份（Begin）与结束年份（End）。为突出研究主题演化的主要趋势，选取突现强度较高、持续时间较长或能够反映前沿转向的代表性关键词进行展示（表 4）。其中，突现强度最高的中文关键词为“虚拟现实”（Strength=16.35，2015 — 2018 年），英文关键词为“Virtual Reality”（Strength=77.55，2015 — 2022 年）。此外，“增强现实”（Strength=7.71，2017 — 2020 年）、“混合现实”（Strength=4.52，2018 — 2021 年）、“医学影像”（Strength=3.12，2020 — 2022 年）、“机器学习”（Strength=3.33，2022 — 2023 年）等词呈现出持续多年的突现特征。2022 年以后，“在线教学”“元宇宙”等关键词也进入突现结果。

表 4　2015 — 2025 年中英文数据库中人工智能医学教育研究相关文献的关键词突现表

序号	关键词	首次出现年份	突现强度	突现开始年份	突现结束年份
1	虚拟现实	2015	16.35	2015	2018
2	增强现实	2017	7.71	2017	2020
3	混合现实	2018	4.52	2018	2021
4	医学影像	2020	3.12	2020	2022
5	机器学习	2020	3.33	2022	2023
6	在线教学	2022	2.79	2022	2023
7	元宇宙	2022	3.74	2023	2025
8	virtual reality	2015	77.55	2015	2022
9	surgical training	2015	13.69	2015	2021
10	three dimensional imaging	2015	10.92	2015	2021
11	simulation training	2015	25.73	2016	2020
12	augmented reality	2015	26.85	2020	2023
13	mixed reality	2018	12.65	2020	2023
14	deep learning	2019	10.21	2020	2022
15	machine learning	2018	17.20	2021	2022

3 讨论

3.1 生成式AI成为近年重要前沿，关键词突现与主题演化共同提示研究热点转向

综合关键词突现与主题演化结果可见，2015—2020年前后，人工智能医学教育研究主要围绕沉浸式教学与模拟训练展开，“虚拟现实(virtual reality)”“模拟训练(simulation training)”“增强现实(augmented reality)”等关键词在较长时间窗内保持活跃。随后研究热点逐步拓展至混合现实、深度学习、机器学习等方向。2023年以来，生成式人工智能(generative AI)与大语言模型(large language models, LLMs)相关术语在主题演化中集中出现，如ChatGPT、LLMs、generative AI等(图5)；突现分析中也出现“元宇宙”等近年活跃关键词(表4)。上述结果提示，该领域研究重点已由早期沉浸式训练和模拟教学，进一步拓展至围绕大模型的学习支持、教学交互、评价反馈与智能辅助决策等方向。

这一变化与近年医学教育领域关于生成式AI应用场景、评价框架与证据现状的综述性或指南性研究基本一致[24-27]。生成式AI的快速发展，使其不再只是知识检索或文本生成工具，而是逐步进入病例讨论、临床推理训练、形成性评价、个性化反馈和学习支持等具体教学环节。对医学教育实践而言，生成式AI的价值不应停留于工具展示或单点试用，而应进一步嵌入真实教学情境，与课程目标、评价方式和教师反馈机制相结合，推动其由“辅助工具”向“教学支持系统”转化。

值得注意的是，生成式AI相关研究在不同语言文献中的术语使用和主题组织方式存在一定差异。英文文献中，相关表述逐渐从“ChatGPT”等单一模型名称扩展至“GPT-4”“generative AI”“LLMs”等更为多样化的术语；中文文献中，则可能同时出现基于具体教学场景、本土模型或课程改革语境的命名方式。这种术语层面的差异不仅反映技术演进，也可能影响关键词共现、主题识别与跨语种比较结果。因此，未来跨语种文献计量研究应更加关注术语演化、概念对齐和关键词归一化问题。

从关键词共现网络及主题演化结果来看，中英文文献在生成式AI医学教育应用中的关注重点也存在差异。英文文献更倾向于将大语言模型与学习成效、评价指标、反馈机制和应用证据联系起来；中文文献则更多将生成式AI嵌入教学改革、课程建设和系统整合等综合性议题之中。这说明中文研究更偏向教学改革与课程建设，英文研究则更强调成效评价与应用证据生成。对于我国医学教育技术发展而言，未来在继续推动AI课程建设和教学改革的同时，还应加强学习成效评价、伦理规范、教师数字素养提升以及人机协同机制建设，以推动生成式AI在医学教育中的规范化和证据化应用。

同时，生成式AI在医学教育中的快速扩展也带来了准确性、可解释性、隐私保护、学术诚信和教师角色变化等问题。因此，后续研究除继续关注其教学应用潜力外，还应进一步明确其应用边界，完善伦理治理框架，并探索教师、学生与人工智能系统之间的合理协作机制。

3.2 各期刊发文数量提示中英文样本的知识载体分布模式存在差异

期刊发文量分析显示，中英文数据库样本在期刊构成及分布上存在差异(表3)，这提示不同语言样本的知识载体分布模式并不完全相同。依据Bradford定律，并以“累计发文量约占总量的1/3”作为核心区划分标准，中文样本中按发文量排序的前10种期刊累计占比已接近核心区阈值；在进一步纳入《医学信息学杂志》和《中国毕业后医学教育》后，累计占比达到33.85%，由此可识别出中文样本的Bradford核心区期刊。相比之下，英文样本中发文量排名前5种期刊的累计占比即达到33.99%，说明英文样本中较少数量的高发文期刊即可达到核心区阈值。

在本研究样本中，中文文献主要发表于医学教育类期刊，Bradford核心区期刊相对分散；英文文献则更多分布于少数高发文期刊，并涉及医学教育、医学信息学及其他交叉领域期刊，呈现出一定的交叉载体特征。需要指出的是，本研究的差异比较建立在“中文文献集合”和“英文文献集合”之上，其结果不仅与研究主题有关，也可能受到传播语言、数据库覆盖范围、期刊收录机制及学科覆盖结构等因素影响[28-29]。英文作为国际科学交流的主要语言，往往对应更广泛的潜在读者与引用网络[30]；已有研究表明，在其他条件相近的情况下，英文论文的被引水平普遍高于非英文论文，提示语言可见度可能影响计量指标和知识传播效果[31]。同时，国际数据库对不同语言期刊的收录范围并不均衡，也会影响文献可见度和计量结果的可比性[28]。因此，本研究更适宜将上述结果解释为样本层面的知识载体分布差异，而不宜据此对中英文期刊生态作出超出数据支持范围的判断。

3.3 中英文作者合作网络均呈多中心集群结构，但合作结构特征存在差异

基于作者合作网络的可视化结果及结构指标(图

4），中英文文献样本均呈现多个合作集群并存的网络结构，但两类网络的组织形态存在差异。英文作者合作网络中，部分作者集群规模较大，集群之间存在一定合作联系。从结构指标来看，英文网络在满足阈值条件的作者子集中形成466条连边，总连接强度为682，聚类数为10，显示该作者子集在本研究样本中具有较多合作关联。相比之下，中文作者合作网络中，作者合作关系分布于多个相对独立的合作团簇。在满足阈值条件的作者子集中，共形成117条连边，总连接强度为269，聚类数为62，显示该作者子集在本研究样本中呈现较多小规模合作团簇并存的特征。

上述结果提示，人工智能医学教育研究已形成一定的作者合作基础，但不同语言样本中的合作结构有所差异。对于我国人工智能医学教育研究而言，未来可进一步加强跨院校、跨学科和跨平台协作，推动研究从单中心教学实践逐步走向多中心证据积累。此外，作者同名消歧及计数方法差异等问题亦可能影响合作强度估计，后续研究有必要进一步规范作者识别、机构归并和计数方法[32-33]。需要指出的是，中英文网络在作者最小发文量阈值设定上存在差异，中文为4篇，英文为6篇，该差异可能在一定程度上影响两类网络结构指标的直接可比性。因此，上述结果主要反映满足阈值条件的高发文作者子集的合作结构特征。

3.4　中英文研究取向差异提示我国人工智能医学教育需进一步加强证据生成与评价研究

本研究显示，中文文献更多聚焦教学改革、课程建设与系统整合，英文文献则更关注具体教学场景中的应用效果评估、反馈机制与评价框架构建。同时，英文文献的期刊分布涉及医学教育、医学信息学及教育技术等交叉领域，作者合作网络在满足阈值条件的高发文作者子集中也呈现出更多合作联系。上述结果表明，中英文文献差异不仅体现在研究热点层面，也体现在知识传播载体与学术协作方式层面。

这一发现与既有研究的总体判断基本一致[15-16,24-27]。已有研究指出，国内相关研究更多强调教学改革与实践创新，而国际研究则更关注应用成效、评价证据与规范框架[16]。与既往研究相比，本研究基于中英文多数据库样本，进一步从主题演化、期刊分布与作者合作网络等多个维度呈现了这一差异。造成上述差异的原因，可能与研究发展阶段、发表生态、数据库覆盖范围及协作方式等因素有关。我国人工智能医学教育研究近年来发展迅速，但部分研究仍集中于平台建设、课堂试点和教学模式探索等应用拓展层面。

这种差异在一定程度上也与我国医学教育改革的政策导向有关。“新医科”建设强调医学教育与信息技术、工程技术等领域的交叉融合，推动课程改革、教学创新与人才培养模式优化[12]；“双一流”建设背景下人工智能相关人才培养政策则进一步强调学科交叉融合与高水平人才培养[13]；相较之下，英文研究更早进入应用效果验证与评价框架构建阶段，这也提示我国未来有必要在持续推进教学改革的同时，进一步加强人工智能教育效果评估研究，形成更加充分的本土证据。此外，中文研究主要集中发表于医学教育类期刊，而英文研究更多分布于医学教育、医学信息学和教育技术等交叉领域期刊，也更有利于形成对学习成效与证据质量的持续关注。

因此，在“新医科”建设与教育数字化持续推进的背景下，我国人工智能医学教育下一步应在持续推进教学应用的同时，进一步加强课程整合、效果评价与多中心证据生成[10-11]。未来有必要在持续推进AI教学应用的同时，更加重视学习质量、能力形成、评价信度、教学公平性及潜在风险的系统研究，并同步完善教师培训、伦理治理与跨学科协作机制，以推动人工智能医学教育由应用导向进一步走向证据导向。

3.5　研究局限性

本研究在数据获取与分析过程中仍存在若干局限，需要在解释结果时予以注意。第一，尽管本研究覆盖了中英文主流数据库，并采用较为全面的主题词体系，但生成式人工智能相关术语在2023年后快速演化，部分研究逐渐以具体模型名称或新兴概念替代上位技术术语进行表述，如GPT-4、Gemini、Claude、多模态AI、AI Agent及部分中文大模型名称等。基于检索策略稳定性与可重复性的考虑，本研究优先采用“生成式人工智能”“大语言模型”等上位概念进行检索，而未对具体模型名称进行穷举式列举。因此，若部分文献未同时包含上述上位概念关键词，仍可能存在少量遗漏。结合关键词共现、聚类及主题演化结果来看，ChatGPT、GPT-4、LLMs、generative AI等新兴术语已在样本中得到体现，因此该问题对总体趋势判断的影响预计有限。第二，本研究按照语言属性将文献划分为中文与英文两类用于双语比较，但语言并不等同于研究语境、研究者背景或数据来源，因此基于语言的分析难以充分反映真实研究体系的复杂性。第三，在作者合作网络分析中，本研究结合了可视化图谱及VOSviewer输出的结构指标，如节点数、聚类数、连边数和总连接强度，以降低单纯依赖视觉

观察带来的主观性。然而，本研究尚未进一步引入网络密度、平均路径长度、聚类系数和模块化指数等拓扑学指标进行更细致的定量验证，因此有关网络结构差异的解释仍应保持谨慎。第四，在跨语言文献整合过程中，由于术语体系、关键词表达和概念边界存在差异，关键词归一化与主题聚类可能受到语义不完全对齐的影响，从而影响部分主题的聚合精度。第五，BERTopic 主题建模虽有助于从大规模文本中识别潜在主题，但主题结构仍可能受到文本预处理方式、嵌入模型选择及默认聚类配置的影响。本研究未对 UMAP、HDBSCAN 及最小主题规模等参数进行单独调优，而采用 BERTopic 默认设置，因此所得结果更适用于揭示总体主题演化趋势，而不宜被理解为对主题边界的唯一划分方式。第六，本研究的文献筛选虽结合自动去重与人工核查，但在处理 7 000 余条记录时难以完全避免边缘化文献的纳入或人工判断的主观偏差。第七，CiteSpace、VOSviewer 等文献计量工具主要用于揭示宏观结构与趋势，难以深入呈现单篇文献的技术细节、教育情境及机制解释。因此，本研究结论主要反映人工智能医学教育研究的整体趋势。未来研究可结合系统综述、内容分析及专家访谈等方法，进一步提升分析的深度与解释力。

4 结束语

本研究基于中英文数据库，对 2015 — 2025 年人工智能在医学教育领域的相关文献进行了系统计量分析，揭示了该领域研究规模的持续增长及主题结构的不断扩展。结果显示，生成式人工智能已成为近年来人工智能医学教育研究中的重要新兴主题。双语文献比较表明，在本研究纳入的中英文数据库样本中，不同语言文献在研究主题、知识载体分布及合作网络形态上呈现出一定差异。上述发现可为人工智能医学教育领域的后续研究、课程建设与教学评价提供参考。未来研究可重点围绕多模态 AI 在临床教学中的情境化应用、生成式 AI 支持下的课程设计与评价优化、人工智能教育伦理与治理机制、教师数字素养提升及教师角色重塑等方向开展更深入的实证研究。同时，应进一步加强多中心研究设计、教育成效评价和跨学科协作机制建设，从而推动人工智能更规范、更有效地服务医学教育高质量发展。

参考文献

[1] 石琬若，韩锡斌. 生成式人工智能对学习分析研究的影响：现状与前瞻：2024 年学习分析与知识国际会议（LAK24）述评[J]. 电化教育研究，2024，45(12)：113-120.

[2] 谢浩然，陈协玲，郑国城，等. 人工智能赋能个性化学习：E-learning 推荐系统研究热点与展望[J]. 现代远程教育研究，2022，34(3)：15-23.

[3] 岳家贤，商庆尧，刘佳祥，等. 虚拟现实和增强现实技术在医学教育中的应用[J]. 中国医学装备，2025，22(7)：172-176.

[4] 李琨，石晓明，贾立伟. 基于 VR 技术的《人体解剖学》虚拟教学系统的设计与实现[J]. 电脑编程技巧与维护，2024(1)：149-150.

[5] Salimi S，Asgari Z，Mohammadnejad A，et al. Efficacy of virtual reality and augmented reality in anatomy education：A systematic review and meta-analysis[J]. Anatomical Sciences Education，2024，17(9)：1668-1685.

[6] 刘明，吴忠明，廖剑，等. 大语言模型的教育应用：原理、现状与挑战：从轻量级 BERT 到对话式 ChatGPT[J]. 现代教育技术，2023，33(8)：19-28.

[7] Kasneci E，Sessler K，Küchemann S，et al. ChatGPT for good? On opportunities and challenges of large language models for education[J]. Learning and Individual Differences，2023，103：102274.

[8] Lo C K. What is the impact of ChatGPT on education? A rapid review of the literature[J]. Education Sciences，2023，13(4)：410.

[9] Alafnan M A，Dishari S，Jovic M，et al. ChatGPT as an educational tool：Opportunities，challenges，and recommendations for communication，business writing，and composition courses[J]. Journal of Artificial Intelligence and Technology，2023，3(2)：60-68.

[10] 国务院. 国务院关于印发《新一代人工智能发展规划》的通知[EB/OL]. (2017-07-08)[2026-03-07]. https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.

[11] 教育部. 教育部关于印发《教育信息化 2.0 行动计划》的通知[EB/OL]. (2018-04-13)[2026-03-07]. http://www.moe.gov.cn/srcsite/A16/s3342/201804/t20180425_334188.html.

[12] 国务院办公厅. 国务院办公厅关于加快医学教育创新发展的指导意见[EB/OL]. (2020-09-17)[2026-03-07]. https://www.gov.cn/zhengce/content/2020-09/23/content_5546373.htm.

[13] 教育部，国家发展改革委，财政部. 关于“双一流”建设高校促进学科融合 加快人工智能领域研究生培养的若干意见[EB/OL]. (2020-01-21)[2026-03-07]. http://www.moe.gov.cn/srcsite/A22/moe_826/202003/t20200303_426801.html.

[14] Li R，Wu T. Evolution of artificial intelligence in medi-

cal education from 2000 to 2024: Bibliometric analysis[J]. Interactive Journal of Medical Research, 2025, 14: e63775.

[15] Preiksaitis C, Rose C. Opportunities, challenges, and future directions of generative artificial intelligence in medical education: Scoping review[J]. JMIR Medical Education, 2023, 9: e48785.

[16] 封纯，吕萍萍，金敏．基于文献计量学的虚拟现实和人工智能在医学教育中的应用[J]. 医学教育研究与实践，2023，31(3): 278-287.

[17] Chen C M. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377.

[18] 学术点滴，文献计量．COOC一款用于文献计量和知识图谱绘制的软件[CP/OL]. (2020-08-16)[2025-10-13]. https://gitee. com/academic_2088904822/academic-drip.

[19] Batagelj V, Mrvar A. Pajek: Analysis and visualization of large networks[M]//JÜNGER M, MUTZEL P. Graph drawing software. Berlin, Heidelberg: Springer, 2004: 77-103.

[20] van Eck N J, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping[J]. Scientometrics, 2010, 84(2): 523-538.

[21] 梁茂云，张帆，孙晶晶．可视化管理研究领域文献计量分析[J]. 科学观察，2022，17(3): 47-57.

[22] 傅隆生，家博，刘小娟，等．基于文献计量学的农业工程技术发展动态分析[J]. 农业机械学报，2025，56(10): 249-276.

[23] Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure[PP/OL]. Arxiv (2022-03-11) [2025-11-30]. https://arxiv. org/abs/2203. 05794.

[24] Masters K, Herrmann-Werner A, Festl-Wietek T, et al. Preparing for artificial general intelligence (AGI) in health professions education: AMEE guide No. 172[J]. Medical Teacher, 2024, 46(10): 1258-1271.

[25] Masters K, Macneil H, Benjamin J, et al. Artificial intelligence in health professions education assessment: AMEE guide No. 178[J]. Medical Teacher, 2025, 47(9): 1410-1424.

[26] Thomae A V, Witt C M, Barth J. Integration of ChatGPT into a course for medical students: Explorative study on teaching scenarios, students' perception, and applications[J]. JMIR Medical Education, 2024, 10: e50545.

[27] Li J, Yin K Y, Wang Y D, et al. Effectiveness of generative artificial intelligence-based teaching versus traditional teaching methods in medical education: A meta-analysis of randomized controlled trials[J]. BMC Medical Education, 2025, 25(1): 1175.

[28] Mongeon P, Paul-Hus A. The journal coverage of Web of Science and Scopus: A comparative analysis[J]. Scientometrics, 2016, 106(1): 213-228.

[29] Aksnes D W, Sivertsen G. A criteria-based assessment of the coverage of Scopus and Web of Science[J]. Journal of Data and Information Science, 2019, 4(1): 1-21.

[30] Bahji A, Acion L, Laslett A M, et al. Exclusion of the non-English-speaking world from the scientific literature: Recommendations for change for addiction journals and publishers[J]. Nordic Studies on Alcohol and Drugs, 2023, 40(1): 6-13.

[31] Di Bitetti M S, Ferreras J A. Publish (in English) or perish: The effect on citation rate of using languages other than English in scientific publications[J]. Ambio, 2017, 46(1): 121-127.

[32] Torvik V I, Smalheiser N R. Author name disambiguation in MEDLINE[J]. ACM Transactions on Knowledge Discovery from Data, 2009, 3(3): 11.

[33] Waltman L, van Eck N J. Field-normalized citation impact indicators and the choice of an appropriate counting method[J]. Journal of Informetrics, 2015, 9(4): 872-894.