DOI:10.13193/j.issn.1673-7717.2025.02.004

大语言模型在中医药领域的应用展望思考

杨乐乐[1],王哲[2],姚克宇[1],刘丽红[1],朱彦[1]
(1.中国中医科学院中医药信息研究所,北京 100700;2.莱比锡大学医学信息统计学暨流行病学研究所,德国 莱比锡 04103)

摘要:中医药作为我国传统医学的重要组成部分,具有丰富的理论基础和宝贵的实践经验,但其智能化发展仍需要多种工具支持。随着自然语言处理技术和人工智能的快速发展,大语言模型作为其重要技术在医疗健康领域的应用受到广泛关注。基于大语言模型在生物医学领域的研究,对其在中医药领域的应用研究进行展望思考,以期为中医药领域的智能化研究和实践工作提供参考。为了探讨大语言模型在中医药领域的应用前景,通过对生物医学领域大语言模型相关文献进行分析研究,首先概述了医学大语言模型的发展进程,介绍了大语言模型的概念、发展历程及国内外医学大语言模型的研究现状。其次,基于生物医学领域已开展的大语言模型相关研究,思考并展望了大语言模型在中医药领域的应用前景,并深入思考了构建及应用中医大语言模型可能面临的风险与挑战,针对这些问题提出了相应的建议措施。最后,总结并展望了大语言模型在中医药领域的未来发展方向。

关键词:大语言模型;中医药;应用展望
中图分类号:R2-03 **文献标志码**:A **文章编号**:1673-7717(2025)02-0016-09

Prospective Reflections on Application of Large Language Models in Field of Traditional Chinese Medicine

YANG Lele[1], WANG Zhe[2], YAO Keyu[1], LIU Lihong[1], ZHU Yan[1]
(1. Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China;
2. Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Leipzig 04103, Germany)

Abstract: Traditional Chinese medicine (TCM) is an integral part of traditional medicine in China, with a rich theoretical basis and valuable practical experience. However, the advancement of TCM necessitates diverse tools to support its intelligent development. With the rapid advancements in natural language processing technology and artificial intelligence, the application of large language models (LLMs) has garnered widespread attention in the field of medical health. This paper focusedon exploring the potential application of LLMs in the field of TCM, aiming to provide insights for intelligent research and practice in this domain. Firstly, it provided an overview of the development of LLMs in medicine, introduced the concept and development history of such models, and presented the current research status in the biomedical field. Subsequently, the paper delved into the potential application of large language models in TCM, considered associated risks and challenges, and provided recommendations to address these issues. Finally, it outlined the future development direction of large language models in the field of TCM.

Keywords: large language models; traditional Chinese medicine; application perspectives

作为最具代表性的大语言模型之一,OpenAI开发的ChatGPT(chat.openai.com)展现出了强大的自然语言处理能力,在无监督生成文本方面具备独特优势。这种技术为生物医学领域的智能化发展带来了崭新的机遇。例如,它可以用于辅助临床决策[1]、自动生成临床文本[2],还可以搭建生成式智能问答系统[3]。中医学作为我国传统医学的重要组成部分,具有独特的理论基础和宝贵的临床经验,但目前大语言模型相关研究较少,其智能化发展仍面临诸多挑战。在如今人工智能广泛应用于医学领域的背景下,探索大语言模型在中医药领域的多种应用模式,对中医药领域的智能化发展具有重要意义。

中医学与生物医学同属于医学领域,有一定的相似性和可迁移性,可以通过借鉴生物医学领域大语言模型的研究经验,展望大语言模型在中医药领域的应用发展方向。本文通过对医学大语言模型相关文献进行分析研究,综合医学大语言模型的发展进程和国内外研究现状,并从不同维度展望大语言模型在中医药领域的应用,包括临床、科研、医学教育等场景。同

基金项目:国家自然科学基金面上项目(82174534,82274684);国家中医药管理局中医师资格认证中心/全国中医药教育发展中心项目(TB2023008);中国中医科学院基本科研业务费自主选题项目(ZZ170326)

作者简介:杨乐乐(1999-),女,安徽阜阳人,硕士在读,研究方向:中医药领域本体构建。
通讯作者:刘丽红(1979-),女(满族),河北遵化人,副研究员,硕士研究生导师,硕士,研究方向:中医药知识组织。E-mail:llhong_2008@163.com。
朱彦(1983-),男,湖南长沙人,副研究员,硕士研究生导师,博士,研究方向:中医药本体研究。E-mail:zhuyan166@126.com。

时，思考大语言模型开发和应用中可能面临的风险和挑战，提出了相关建议，以期为中医药领域大语言模型的研究与实践提供参考。

1 医学大语言模型的发展概述

1.1 大语言模型的概述及发展进程 大语言模型是通过在极大规模语料（通常为上百亿词或更多）上进行无监督预训练所得的生成式语言模型。它可以学习语言的统计规律与知识，生成自然流畅的语言序列，并在下游任务中实现零样本（zero - shot）或少样本（few - shot）的学习[1]。近年来，大语言模型随着深度学习和计算能力的发展而快速崛起。伴随着自然语言处理技术的发展，大语言模型的产生和发展经历了以下几个阶段。

第一阶段是早期的统计语言模型阶段，这一阶段采用统计学方法从大规模语料中学习词频、语法结构等知识，实现了从依赖人工构建知识库到机器自动大规模学习语言知识的转变。主流技术包括支持向量机（support vector machine，SVM）、隐马尔可夫模型（hidden Markov model，HMM）[4]、最大熵模型（maximum entropy model，MaxEnt）、条件随机场（conditional random field，CRF）等，依靠人工标注语料集进行监督学习，数据量级在百万级左右[5-6]。

第二阶段是深度学习语言模型阶段，这一阶段引入了词向量和各种深度神经网络模型，提高了语言的连续性和流畅性，依靠大量人工标注语料进行训练。时代背景是大数据与计算能力的提升。典型技术栈包括 Encoder - Decoder、长短期记忆（long short - term memory，LSTM）、Attention、Embedding 等，依靠人工标注的训练数据也提升到了千万量级，这为后来的预训练语言模型奠定了技术基础[7-8]。

第三阶段是预训练语言模型阶段，采用了自监督学习策略进行模型预训练，大幅减少对人工标注语料的依赖，语言表达和理解能力取得飞跃。预训练语言模型系统通常分为两步：无监督预训练和有监督微调，该阶段将预训练数据量扩大 3 ~ 5 倍，典型技术包括 Encoder - Decoder、Transformer、Attention 等。预训练模型将语言表达和理解的能力提升到全新的高度[9-10]。

当前处于第四阶段的超大规模语言模型阶段，进一步扩大模型规模，注重融合不同领域知识，增强人机交互能力，从被动服务向主动对话演进，使语言模型得以服务于更加开放和普适的用途。这些模型可以视为在通用预训练基础上进行的知识增强和迁移学习[11]。

从统计模型到深度学习模型，再到预训练模型，自然语言处理经历了方法论的变革和技术的飞跃。这主要得益于计算能力、数据量和算法的提升。当前出现的超大规模预训练语言模型标志着自然语言处理发展进入了一个新的阶段，应用前景广阔[11]。

1.2 国内外医学大语言模型现状 近年来，大语言模型在医学领域的研究探索取得了显著进展。通过学习和理解大量的医学文本数据，大语言模型能够获取医学领域的背景知识，并在医学自然语言理解和生成等方面展现出优异的表现。尤其是医学大语言模型在医学信息提取、问答系统、知识图谱等任务上的广泛研究和出色表现，已经在国内外得到了广泛的关注和应用。如今，国内外的医学大语言模型研究和应用已经初具规模，各种医学大语言模型相继涌现。见表 1。

表 1 医学领域的大语言模型

国内外	模型名称	模型尺寸	基础模型	获取途径	简介	是否开源
国内	MedGPT[12]	100 B	GPT	https://medgpt.co/home/zh	通过命名实体识别和链接工具（即 MedCAT）来构建和组织 EHR 的自由文本部分，并预测一系列未来的医学事件	/
	岐黄问道大语言模型	/	/	https://dajingtcm.com	中医大语言模型，由大经中医出品。共有 3 个子模型：基于已确诊疾病的临床诊疗大语言模型；仅基于症状体征的临床诊疗大语言模型；中医养生调理大语言模型	否
	TCMLLM - PR	10 M	ChatGLM	https://github.com/2020MEAI/TCMLLM	由北京交通大学计算机与信息技术学院医学智能团队开发，目前针对中医临床智能诊疗问题中的处方推荐任务	是
	IvyGPT	/	LLaMA	https://github.com/WangRongsheng/IvyGPT	IvyGPT 通过 RLHF 可以提供更丰富、更接近人类的诊断和治疗答案，并且实验结果表明其性能优于其他医学 GPT 模型，有效改善医学诊断和治疗的准确性和效率	是
	XrayGLM[13]（首个会看胸部 X 光片的中文多模态医学大语言模型）	/	ChatGPT、VisualGLM - 6B	https://github.com/WangRongsheng/XrayGLM	是一种创新的对话式医疗视觉语言模型，专门用于分析胸部 X 光片。其核心概念围绕着对齐医学视觉和文本表示，在医学影像诊断和多轮交互对话上显示出了非凡的潜力	是
	砭石大语言模型	/	ChatGPT	https://www.wjx.cn/vm/tz4UXzB.aspx	是支持多模态（问诊文本数据、医疗影像数据、用户的面部特征视频数据、用户睡眠音频数据）的大语言模型。实现智能问诊、辅助阅片、面诊舌诊、生理指标预测、睡眠监测等多样化的任务	/
	ShenNong - TCM - LLM - 神农中医药大语言模型	/	LlaMA - 7bLoRA	https://github.com/michael - wzhu/ShenNong - TCM - LLM	与 ChatMed 相伴生，旨在推动 LLM 在中医药领域的发展和落地，提升 LLM 在中医药方面的知识与回答医学咨询的能力	是
	OphGLM[14]	/	/	https://github.com/ML - AILab/OphGLM /	基于疾病相关知识数据和公开的真实医疗对话，在大型语言模型中引入了视觉能力，是一个眼科大型语言和视觉助手	/
	DoctorGLM[15]	/	ChatGLM - 6B	https://github.com/xionghonglin/DoctorGLM	中文问诊模型，通过中文医疗对话数据集进行微调	是
	BenTsao[16]	/	LLaMA	https://github.com/SCIR - HI/Huatuo - Llama - Med - Chinese	基于医学知识图谱微调的医疗领域模型	/
	BianQue	/	ClueAI/ChatYuan - large - v2	https://github.com/scutcyr/BianQue	由华南理工大学开发的一个经过指令与多轮问询对话联合微调的医疗对话大语言模型	/

续表 1 医学领域的大语言模型

国内外	模型名称	模型尺寸	基础模型	获取途径	简介	是否开源
	HuatuoGPT[17]	/	GPT-like	https://github.com/FreedomIntelligence/HuatuoGPT	经过中文医学指令精调/指令微调(Instruct-tuning)的一个 GPT-like 模型	是
	孙思邈中文医疗大语言模型(Sunsimiao)	/	Baichuan	https://modelscope.cn/models/AI-ModelScope/Sunsimiao/summary	是由 baichuan-7B 及书生·浦语 InternLM 在十万级高质量的中文医疗数据中基于 LoRa 微调而得	/
	Med-ChatGLM	7 B	ChatGLM	https://github.com/SCIR-HI/Med-ChatGLM	基于中文医学知识的 ChatGLM 模型微调,微调数据与 BenTsao 相同	是
	QiZhenGPT	/	LLaMA-7B	https://github.com/CMKRG/QiZhenGPT	首先针对药品知识问答发布了评测数据集,后续计划优化疾病、手术、检验等方面的问答效果,并针对医患问答、病历自动生成等应用展开拓展	/
	ChatMed-Consult	/	LlaMA-7b	https://github.com/michael-wzhu/ChatMed	以中文医疗在线问诊数据集 ChatMed_Consult_Dataset 作为微调训练集。赋能中医药传承	是
	PULSE 中文医疗大语言模型	/	OpenChineseLLaMA 13B	https://github.com/openmedlab/PULSE	支持医学领域的各种自然语 言处理任务,包括健康教育、医师考试问题、报告解读、医疗记录结构化 以及模拟诊断和治疗	/
	MING	/	ChatGLM-6B LoRA 16-bit	https://github.com/MediaBrain-SJTU/MING	该模型提供了两种功能:①医疗问答:对医疗问题进行解答,对案例进行分析;②智能问诊:多轮问诊后给出诊断结果和建议	/
国外	BioBERT[18]	/	BERT	https://github.com/dmis-lab/biobert	是一种生物医学语言表示模型,专为生物医学文本挖掘任务(如生物医学命名实体识别、关系提取、问题解答等)而设计	是
	BioMedLM	2.7 B	GPT-2	https://github.com/stanford-crfm/BioMedLM	斯坦福大学基础模型研究中心(Center for Research on Foundation Models,CRFM)和 MosaicML 宣布发布,是一种专门构建的 AI 模型,经过训练可以解释生物医学语言	/
	Visual Med-Alpaca[14]	7 B 和 13 B	LLaMa-7B	https://github.com/cambridgeltl/visual-med-alpaca	建立在 LLaMa-7B 基础上的一个专门为生物医学领域设计的开源、多模态的基础模型	是
	BlueBERT	1 B	BERT	https://github.com/ncbi-nlp/BlueBERT	基于 BERT 预训练的生物医学文本理解模型,包含领域特定的词汇和医学标注	是
	ClinicalBERT[19]	/	BERT	https://github.com/EmilyAlsentzer/clinicalBERT	是专门针对临床医学领域而设计的语言模型,可以应用于临床文本挖掘、疾病诊断等任务	/
	BioMedGPT[20]	10 B	/	/	是清华大学聂再清教授带领团队着手构建的多模态生物医药领域基础模型,旨在将生物世界分子、文本与知识进行统一表示学习以达到在各项下游任务上能力的整体提升	是
	PubMedGPT	31 B	BERT	/	由斯坦福基础模型研究中心(CRFM)和 MosaicML 联合开发的一种经过训练可以解释生物医学语言的大型语言模型	是
	PMC-LLaMA[21]	7 B	/	https://github.com/chaoyi-wu/PMC-LLaMA	它在 480 万篇生物医学学术论文上进行了微调,以增强其在医学领域的能力。在微调后,该模型展现出更好的医学概念理解能力,并在 QA 基准上实现了高性能	是
	Med-PaLM 2	/	/	/	用于促进丰富、信息丰富的讨论,回答复杂的医学问题,并在复杂和非结构化的医学文本中找到见解	否
	Med-PaLM	/	/	/	谷歌联手 DeepMind 宣布发布的一个在更窄的参数范围内运行的,专门用于回答医疗保健相关问题的大型医用语言模型	否
	MedicalGPT	/	GPT	https://github.com/shibing624/MedicalGPT	MedicalGPT 训练医疗大语言模型,基于 ChatGPT 训练模式,实现包括二次预训练、有监督微调、奖励建模、强化学习训练	是
	ChatDoctor	7 B	LLaMA	https://chatdoc-china.com/	使用医学领域知识在大型语言模型 LLaMA 上进行微调的医疗大语言模型	是
	MedAlpaca	7 B 和 13 B	Stanford Alpaca、AlpacaLoRA	https://github.com/kbressem/medAlpaca	专为医学问答和对话应用优化的大型语言模型,采用多种医学文本训练,为医疗机器人提供开源的语言处理技术	是

表 1 梳理了近年来具有代表性的医学大语言模型的相关信息,包括模型名称、模型尺寸、基础模型、获取途径、简介以及是否开源等。这些模型在医学领域的研究应用中发挥着重要的作用,它们以其优秀的性能和广泛的适用性,为医学文本理解、临床特征提取和回答医学相关问题等方面提供了有力的支持。

在东方医院与其他公司合作的中医药大语言模型项目中[22],已初步实现医疗问答和辅助诊疗等功能。与西医相比,中医注重整体观察和诊断,强调病因与病机的调节,中医大语言模型基于中医临床诊疗特色和辨证特征,探索通用人工智能(artificial intelligence,AI)在中医领域的应用。

2 大语言模型在中医药领域的应用展望

大语言模型是一种基于深度学习的自然语言处理模型,作为医生的知识型辅助工具,在医学临床、科研和教育等方面都

展现出巨大的应用潜力。鉴于中医学拥有独特的理论体系和丰富的经验知识,建立面向中医领域的大语言模型,将为推动中医智能化和信息化进程提供重要机遇。基于对现有大语言模型在生物医学领域应用的分析,我们对大语言模型在中医药领域的应用进行思考,从应用场景和数据处理两个维度分别论述,见表 2。

2.1 临床应用 在现代医学的临床应用方面,医学大语言模型在辅助临床决策、优化医疗质量、实现个性化诊疗、提升患者体验、平均医疗资源以及临床电子病历文本处理等多方面都体现出了巨大的潜力[23],这为大语言模型在传统中医临床的应用提供了宝贵的借鉴意义。

(1)辅助支持临床决策,优化医疗质量:大语言模型可以通过分析患者历史数据和当前症状,提供诊断假设和检查建议,通过归纳推导决策规律,提高临床诊断决策的准确性,以辅助医生制定诊疗方案[24-25]。RAO A 等[26]的研究表明,ChatGPT 可以对乳腺疼痛和癌症筛查提出合理的影像学检查建议,验证了大语言模型进行医学影像决策的可行性。另外,多模态大语言模型还可以分析患者的症状、药物信息和医学图像,以辅助提高诊断的准确率、优化治疗方案以及监测医疗质量[27]。在中医临床的应用场景中,大语言模型则可以在辨证论治的过程中,借助人工智能模拟中医诊疗思维[28],提供对病因、辨证类型、治疗方案等方面的决策支持和知识咨询。目前有少量大语言模型产品发布,宣称其在部分诊断场景下的准确率可达 90%[29],但具体效果还有待进一步验证。

(2)提供个性化诊疗建议、提升患者体验:医学数据的复杂性要求医生需要根据不同患者的情况提供定制化的诊疗方案[30],大语言模型可以通过分析患者的个人健康信息和其他相关数据,设计个性化的诊疗方案。同时进行智能健康知识问答,来提供日常健康管理建议[31],模型可以在诊疗过程中缓解患者的紧张情绪并提供安慰[11,32-34],以提供更加人性化的交互体验。AYERS J W 等[35]的研究显示,大语言模型在回答医学问题时体现出的同理心,超越了医生的表现。辨证论治是中医临床的特色与优势[36]。大语言模型可以根据不同体质特点提供个性化的中医日常健康管理建议,以促进中医药的个体化诊疗。在治疗方面,大语言模型可以帮助优化中医临床的辨证分型和药方匹配,实现更精准有效的治疗。此外,大语言模型可以在中医诊疗过程中,通过模拟人类语言的情感关怀交流方式,提升患者的就医体验。

(3)平均医疗资源:JAMA 期刊征集了一批 AI 应用于提高医疗资源匮乏地区的医疗水平的研究报告,发现在医疗资源有限的地区,大语言模型可以通过学习全国范围内的数据,驱动资源分配的策略研发,弥补医疗资源的不足,提高这些地区的医疗诊断水平[23]。我国存在区域医疗资源分布不均衡的问题[37],在中医领域更加突出,大语言模型可以通过对中医资源进行优化配置,帮助医疗资源匮乏的地区获取更好的中医诊疗条件。

(4)临床电子病历文本的高效处理:多模态大语言模型具有强大的信息集成能力,可以实现不同数据源之间的跨模式融合[37-39],这为医疗文书的智能生成提供了可能。有研究显示[40-41],大语言模型可以克服传统方式下出院小结常常遗漏细节的问题,快速生成标准化的出院小结,减轻医生的文书负担,节省出更多时间照顾患者[30]。大语言模型还可以通过提问患者症状等信息快速收集和整合临床资料[42-43],结合语音识别技术,有望实现临床文书的全流程自动编写[44-45]。同样,国家中医药管理局印发的《"十四五"中医药人才发展规划》中提到中医药人才总体规模不够即中医医生更短缺,利用大语言模型自动生成中医电子病历等临床文书,可以提高中医医生的书写效率,减轻他们的文书工作负担。

2.2 科研应用 大语言模型是医学知识管理和发现的强大新工具,在构建医学知识图谱、继承和挖掘医学文献知识、药物研发、辅助科研实验设计等科研场景同样展现出了巨大潜力。对于中医药领域而言,借鉴这些大语言模型的应用经验和技术手段,有望在中医古籍的整理、中药配伍规律的解析等方面取得突破性进展。

表 2 大语言模型在中医药领域的应用展望

数据维度	数据收集	数据整理	数据分析	数据可视化
临床	电子病历的文本处理 收集中医临床病案等大量非结构化文本数据	将病历结构化,实现症状、诊断、治疗等信息的自动提取 构建临床数据库	使用工具对建立的数据库进行分析,根据需要进行应用	临床知识问答系统 多模态信息交互展示
科研	文献的文本收集	从大规模医案中抽取症状、治法等知识 提取中医文献中蕴含的传统知识和经验 学术论文实体识别 使用语言模型进行实验报告的自动解析,提取关键实验数据 构建包含方剂、药材、疾病等多方面知识的中医知识图谱 建立覆盖中药成分、靶点、药理活性的结构化知识库	使用工具分析从各个渠道提取/抽取的数据,根据需要进行应用 分析古籍中的方剂组成、功效特点以及经典医案	中医问答系统 借助大语言模型提供准确的中医药文献检索和推荐功能 助力临床试验研究 中药相关研究 古籍检索问答系统
教学	医学教材、讲义、习题等教学资料文本知识收集	构建结构化教学知识库	分析知识库中的教学组织内容 利用知识库中医药领域的教学资料和课程内容	提供中医药学习者自主学习的交互式工具,教学问答系统,支持学生进行问题自主检索 开发评测医学生能力-考试的功能

(1)临床数据处理:胡振生等[46]通过总结现有文章发现,大语言模型能够高效识别中文医学命名等实体、对其进行关系抽取和医学文本语义匹配,以实现对临床数据的快速分类和处理。大语言模型可以自动抽取并深度理解来自医学文献、电子病历、临床试验等多源异构医学数据[47-48],有助于构建更加全面、准确的医学知识图谱。在中医药领域,大语言模型也可以构建中医知识图谱,并从中医典籍中提取隐含的知识关系,帮助中医研究人员从繁杂的数据中快速筛选出有价值的信息。

(2)文献传承和知识挖掘:大语言模型可以利用收集大规模医学文献的数据,实现文献分类、术语标准化、关系提取、知识图谱构建等功能,为医学研究提供支持和指导[49]。中医古籍包含了丰富的理论体系、临床经验和医案,这些古籍是中医发展重要的活水之源,大语言模型在中医古籍领域具有广阔的应用前景,可以利用其文本挖掘和知识提取的自动化能力,对中医古籍文献进行整理、构建古籍数据库,实现对中医文献的整理、索引和概括;结合知识图谱和问答系统等可视化应用,可以实现对古籍知识的挖掘和利用,弥补中医药古籍中部分内容碎片化的不足;还可以利用编码生成和代码管理能力,构建具有智能问答和知识推理功能的中医药古籍智能系统,推动古籍的数字化和智能化发展。

(3)助力药物研发:大语言模型可以通过对海量药物文献和实验数据进行文本挖掘,发现药物成分的新用途、隐含的药物相互作用和药效机制[50],也可以处理基因、蛋白质等非语言数据,推动相关跨学科研究的发展。例如,ProtGPT2[51]可以在几秒钟内为球状蛋白设计结构,且无须进行额外的模型训练。同时它通过预训练调节,可以生成针对特定蛋白家族、功能或结构的设计。在中医药领域,可利用大语言模型分析中药成分与人体蛋白质之间的作用关系,同时可以通过挖掘中药文献,发现中药成分的新用途和更多药效机制。

(4)辅助实验设计:大语言模型可以从实验报告中抽取可用的数据并生成初始报告文本,以此来辅助设计实验方案,提高研究效率。INAGAKI T 等[52]的研究表明,GPT-4(generative pre-trained transformer 4)可以通过提示指挥机器人自动生成实验指令,推动实验进行。同样,大语言模型可以辅助设计中医研究的实验方案和中医研究报告的生成。

2.3 医学教育应用 大语言模型在医学教育领域的作用主要体现在测试医学生的知识水平[53]、帮助教师进行个性化教学和提升教学效率等方面。在中医药教育领域,大语言模型也可以带来新的考评工具和教学模式,推动中医药教育向个性化、智能化的方向发展。

(1)测试医学生的知识:最近,Med-PaLM 2 在美国医学考试问题上的表现持续达到“专家”医生水平,其得分率达到 85%,比之前的 Med-PaLM 模型提高了 19%[33]。这表明大语言模型可以作为测试学生对所学知识掌握程度的有效工具。同时,大语言模型还可以对学生提出的医学问题进行智能化问答,并提供个性化的学习建议[54-55],以提高学习效率。这也为中医药教育带来了新的考评工具,推动中医教育向智能化发展。

(2)帮助教师进行个性化教学、提升教学效率:KASNECI E 等[56]认为大语言模型可以通过分析学生的成绩数据,提出针对性的干预措施和推荐特定资源,以弥补学生的学习差距,从而提高教学效率[57-59]。教师可以据此制定针对性的教学计划,实现中医教学的个性化,即可以在大语言模型中输入学生的优势、劣势、学习目标和偏好等个人信息,生成适合学生特定需求的个性化学习计划,以提升教学效率。

医学领域涉及到复杂的数据和微妙的推理技巧,需要医生在决策过程中进行更综合的分析与判断。当前的工具大多针对单一任务,大语言模型可以帮助医生整合复杂的数据,提供基于广泛知识的见解,从而更高效地支持医学教育、研究和临床决策,同时改善诊断的准确性和效果,尤其对多模态数据的处理具有潜在优势[60]。

3 风险、挑战及建议

虽然医学大语言模型在医学临床、科研和教育等领域具有巨大的潜力,但在一定程度上也面临着一些风险挑战。其中包括:数据可能并非近期更新,导致模型训练数据的时效性问题;模型输出结果的准确性和一致性可能存在问题;模型缺乏透明度和可解释性以及测试基准的有限性等。此外,还存在模型的现实应用性能问题和与之相关的法律风险。这些挑战也可能延伸到中医药领域的大语言模型应用中。本文将现阶段构建及应用医学大语言模型,尤其是中医大语言模型可能面临的风险与挑战进行汇总分析,并给出一些相应的建议,见表 3。

3.1 新近性 西医的训练数据大多为临床试验和研究数据,相对容易获取,而中医的相关数据则分散在不同的古籍和文献中,其数据新近性问题可能对模型影响较小,但仍存在。大语言模型的训练数据集往往存在时间截止点,未能涵盖最近时期的语料。例如,GPT3.5、GPT4 的训练数据集仅截止于 2021 年

表 3 构建应用中医大语言模型的局限性及未来发展建议

局限性	描述	建议
新近性	中医模型的训练数据不具有实时性;未接入实时上网连接	收集最新的训练数据;实时互联网访问(例如,Bing AI,Sparrow 和 Blender Bot 3)
准确性和一致性	模型输出的准确率参差不齐且未达到完全正确;模型未掌握真正的语义,输出结果会自相矛盾(中医各流派的标准不统一)	训练数据的验证;不确定性指标;微调以优化医疗准确性;通过智能提示(例如,思维链);进行自我改进
透明度和可解释性	作为黑箱模型,不能准确展示其推理机制/过程;中医疾病诊断必需的推断依据,大语言模型的逻辑思维不够严格;中医各流派的学术思想错综复杂,加大解释难度	要求输出引用数据集的哪些部分有助于模型的答案;“可解释的”人工智能研究和开发
应用性能	模型构建成本高,会阻碍中医模型构建的多样性;中医模型的实用性能测试不太理想;古籍与现代融合的理解能力	通过反馈改进具体性能问题;降低数据共享的门槛
法律	模型使用的中医医疗数据隐私;及推荐建议的合法性;输出内容的知识产权等问题	保护医疗数据隐私;制定相关法律

9 月前,导致其对最近两年的新内容理解存在局限[45]。另外,在不接入互联网或使用插件的情况下,大语言模型的知识更新仍然存在时效性问题。为此,谷歌称其开发的模型可以联网搜索信息,但仍存在一定时间差[61]。针对这些问题,我们有如下建议。

(1)提高大语言模型对新近领域知识的敏感度:利用新近的中医学讲座、病例研讨音视频等多模态数据,有助于模型更好地理解和表达新近领域的知识。例如,从最新发表的文献中采集训练数据;扩展收集最新的中医音视频资料,构建多模态的中医大语言模型。

(2)接入实时互联网,获得实时语料:JIN Q 等[62]的研究显示,GPT 模型连接美国国家生物技术信息中心(National Center for Biotechnology Information,NCBI)数据库应用程序编程接口(Application Programming Interface,API),可以高效获取最新信息并准确表达生物学知识。为实现医学大语言模型覆盖内容的持续更新,可以规划借助不同的 API 外部工具[63]。Bing AI 和 Google Bard 已经具备了这一功能[64]。同样,中医大语言模型可通过接入实时互联网语料,获取最新领域知识,提高对新兴研究进展的敏感度和理解能力。

3.2 准确性和一致性 尽管生物医学及中医大语言模型的输出准确度有所提高,但模型输出存在的偏差或"幻觉"还是必须警惕的风险[11]。目前这些模型都是在无监督的大规模语料上通过词语联想进行预训练的,并没有真正掌握语义知识。因此,大语言模型的输出可能存在不一致或错误的内容,这是由其训练机制决定的。我们通过对近期多个医学大语言模型进行测评,发现其对中医领域知识的理解可能存在诸多歧义,测试结果并不如预期的准确性高。在未来的中医大语言模型研发中,需要继续提升模型对语义知识的理解能力,增强其生成内容的可信度。

(1)提高模型准确性:熊明辉等[65]和 HARSHA 等[2]都提出,可以使用高质量的数据持续进行预训练和微调,并引入医学专家进行监督评估来提高大语言模型的准确性。在中医领域的应用中,需要继续优化模型的推理机制,增加结果监测,以最大限度提高模型的准确性。同时由于中医领域术语的特殊性和复杂性,将大语言模型与中医药领域的知识表达技术结合,通过微调来适应在中医药领域的专业应用,使其具备对中医专业术语进行专业分析的能力,应对中医药术语的多义性和概念模糊性等问题,辅助大语言模型对中医药内涵的准确理解。

(2)应用后继续积累数据:在中医大语言模型应用后,可以继续通过任务特定的数据集对模型进行微调,以减少特定领域的不一致输出。例如,可以使用标准的中医病历文本来微调模型,优化其对中医诊断语言的一致理解。

3.3 透明度和可解释性 周学峰[66]认为当前的大语言模型作为黑箱模型,其内部工作机制难以解释[67],无法对其输出内容提供合理的推理解释,所以很难判断其输出的准确性。在医疗保健领域,模型的可解释性至关重要[68],它可以解释患者症状变化的证据和时间相关性,解决系统与专家之间的分歧,也是知情同意的需要[69]。在中医药领域应用也存在这样的黑箱模型风险,其推理过程和结果的准确性及可靠性难以被评估和理解,加上中医流派众多,各学派思想的交叉融合使其解释更加错综复杂。因此,我们需要采用知识驱动与数据驱动相结合的方法,使模型输出更具逻辑性和可解释性,才能推动可解释的中医大语言模型的开发与应用。

(1)利用知识图谱增强模型,引入特定的领域知识:相比大语言模型,知识图谱(Knowledge Graph,KG)可以明确存储丰富的事实知识[67]。JIN Q 等[70]认为引入特定任务相关的知识图谱可以为大语言模型提供外部知识,增强其推理和可解释性[71-72]。在中医药领域,利用中医知识图谱来增强中医大语言模型,可以提高模型输出的逻辑性和可解释性,更好地解释其推理过程,从而提高中医大语言模型的可靠性。

(2)从数据入手提高可解释性:高远等[73]提出建立医疗数据共享模式和元数据标准,以改善模型数据质量,并组跨学科团队根据中医知识优化模型。不同研究团队需要基于中医诊断知识进行推理,用大数据支持调参优化算法,构建融合知识与数据驱动方法的智能诊断系统,以此来增强模型结果的逻辑性和可解释性。

3.4 测试基准的有限性 谷歌表示,缺乏高质量的测试基准是阻碍通用生物医学 AI 发展的关键瓶颈之一[33,74]。目前存在一些单任务的生物医学人工智能数据集,但将它们统一起来构建通用基准的尝试还非常有限。TU T 等[74]阐述 Google Brain 团队开发的 MultiMedBench 是朝多模态医学数据集建设方向迈出了一步,但该基准仍存在数据集规模小、任务和模态多样性不足的问题,如缺乏转录组学和蛋白质组学等生命科学数据。同样,中医药领域也没有一个统一的中医大语言模型测试基准。应扩大模型涵盖的医学和科学领域范围。

当前的医学模型主要基于专家编写的理想化情况,而临床实践中的复杂性需要结合患者信息进行综合评估[75]。因此,需要开发新的基准任务,以反映真实的临床工作流程。在中医领域,应结合专家经验知识,拓展中医问答模型的测试范围使其涵盖中医学的更多方面,以构建质量高、多样性强且规模充分的通用中医模型测试基准。

3.5 现实应用性能 大语言模型在实际医疗场景中的应用存在一些困难[11,74]。一方面,构建医疗领域的大语言模型需要收集大量高质量的行业数据并进行模型微调,这需要大量的人力和财力投入。支振锋[76]提出现有大语言模型存在算力受限、训练不足、研发和运作成本高等问题[29]。另一方面,将这些大型模型部署和运行所需的计算资源对许多医疗机构来说是个障碍,特别是在资源有限的环境下。在中医领域,要推动大语言模型的实际应用,还需要解决这些资源和成本方面的限制,否则模型的实际表现可能无法达到预期。

(1)探索更高效的模型设计,多部门合作:KARABACAK M 等[77]提出了一种能够开发计算效率更高的模型的方法,该方法在保证性能的同时减少了计算资源的需求,从而有助于降低大语言模型的实施成本和资源门槛,并使不同规模的医疗机构都能够受益于这些前沿技术。在中医领域,我们也可以探索更为轻量化和高效的模型设计,并鼓励中医药部门之间的合作,集合资源以构建适用于实际应用的中医大语言模型。

(2)提高中医大语言模型的实用性能:参考 Med-PaLM 2 等成功案例,通过在多个场景中实际应用并获取大量用户反馈,可以不断优化模型,以更好地满足实际中医业务的需求。如对中医古籍进行分析理解时,可以考虑与现代思想融合,使

其更加通俗易懂。同时,在模型的实际应用中必须考虑安全性和公平性的问题。只有当中医大语言模型的性能达到或超过专业的单任务模型,并通过了临床验证,才能真正应用到医疗实践中[73]。

3.6 法律与伦理 大语言模型在医疗领域的应用面临多方面的法律和伦理风险,需要全面审视并采取措施规避[78]。首先是医疗数据隐私泄露风险[30,79-83],KARABACAK M等[77]论述了大语言模型训练需要大量医疗数据,如果处理不当可能导致患者隐私被泄露,违反相关法规。其次是诊断建议和药物推荐的法律合规性风险。日前,北京市卫健委发布的办法中提出,严格限制自动生成处方等行为[84]。再者是RAY P P E[85]和LIEBRENZ M等[86]认为在模型训练和应用中可能会引起知识产权争议。最后,当前相关法规不完善,使大语言模型在决策过程中涉及伦理和道德问题,例如决策的公平性、偏见和责任分配等[66,78,83]。在中医领域开发应用大语言模型时,同样需要关注这些风险。

(1)从医疗数据处理上规避风险:为了严格保护患者隐私,可以采用技术手段加强措施,以避免泄露其真实身份信息。在将模型实际应用于临床前,通过小规模的样本验证对模型的有效性和安全性进行评估,并建立问答系统的监管标准。在模型的输出结果中,应增加警示声明,明确告知用户应以医生的诊断为准,不可过度依赖模型结果。同时,对训练语料的版权审核要加强,以确保生成的内容不侵犯他人权益。此外,可以在模型输出结果中补充数据来源声明,以与外部信息进行区分。最后需要加强对大语言模型的伦理规制,将伦理道德纳入大语言模型研发和应用的全过程,增强伦理意识和行为自觉。

(2)从法律层面规避风险:为了规避技术创新和伦理风险,并保护各方的合法权益,国家应尽快制定相关的法律法规,明确大语言模型的知识产权归属和责任分配。如通过建立伦理委员会或专家团队、制定可追溯可解释的模型设计和决策规则,以及用户申诉和监督机制,解决大语言模型在生物医学领域应用中的伦理和问责问题,从而促进该领域的可持续发展。同时,医学行业的监管部门也应及时制定大语言模型的行业标准,为各参与方提供明确的指引[65]。GATES B[87]研究表明,对于大语言模型需要进行严格的质量评估,并建立明确的治理结构,以合理分配责任,保障患者的安全性。在中医领域开发大语言模型时,应参考这些研究观点,从法律和伦理层面全面考虑可能存在的风险,并建立完善的评估和监管机制,以避免潜在问题的发生,保护公共利益。

4 总结与展望

大语言模型作为一种新兴的人工智能技术,在中医药领域展现出广阔的应用前景。本文通过总结分析大语言模型在生物医学领域的临床、医学研究和医学教育等方面的研究,展望了其在中医药领域多个维度的潜在应用价值,并参考已有的生物医学大语言模型,思考列出了构建应用中医大语言模型这个过程中可能存在的局限性和潜在风险,如数据新近性、输出准确性、可解释性、测试基准、实际应用性能和法律等,同时结合现有研究给出相应的改进建议。针对中医药领域数据的特征和特异性,也可能面临建立全面、精确的中医药知识库、处理中医药术语的多义性和概念模糊性等挑战,应针对性解决这些挑战,如完善知识库、处理术语歧义、个性化医疗优化、提供解释模型决策的机制,同时加强数据管控、加强隐私保护,以推动大语言模型在中医药领域的应用和发展。

总之,大语言模型可以为中医领域的智能化发展带来新的可能性。通过充分利用大语言模型的技术优势,结合中医药领域的丰富知识和数据资源,我们可以期待中医药在知识传承、临床实践、个性化治疗和药物研发等方面取得更加显著的进展。随着相关研究的不断深入,大语言模型将为中医药的科学性和现代化发展注入新的动力。我们期待未来的研究和实践能够进一步探索大语言模型在中医药领域的应用研究,为中医药传统智慧与现代科技的融合带来新的突破,实现中医药领域的智能化发展。

参考文献

[1] NORI H, KING N, MCKINNEY S, et al. Capabilities of GPT-4 on medical challenge problems[J]. 2023. arXiv:2303.13375 [cs.CL].

[2] LIU Y H, HAN T L, MA S Y, et al. Summary of ChatGPT-related research and perspective towards the future of large language models[J]. 2023: arXiv: 2304.01852.

[3] TIAN S B, JIN Q, YEGANOVA L, et al. Opportunities and challenges for ChatGPT and large language models in biomedicine and health[J]. ArXiv, 2023: 2306.10070v2.

[4] KOEHN P, OCH F J, MARCU D. Statistical phrase-based translation[C]//Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1. 2003:N03-1017.

[5] VOGEL S , NEY H . HMM BasedWord Alignment in Statistical Machine Translation[J]. coling, 1996. DOI:10.1007/978-3-642-14883-5-17.

[6] MANNING C D, HINRICH SCHÜTZE. Foundations of Statistical Natural Language Processing[M]. MIT Press, 1999. DOI:10.1162/coli.2000.26.2.277.

[7] WU Y H, SCHUSTER M, CHEN Z F, et al. Google's neural machine translation system: bridging the gap between human and machine translation[J]. arXiv E Prints, 2016: arXiv: 1609.08144.

[8] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. Computer Science, 2013. DOI:https://doi.org/10.48550/arXiv.1301.3781.

[9] LIN Z H, PAN X, WANG M X, et al. Pre-training multilingual neural machine translation by leveraging alignment information[J]. 2020. DOI:https://doi.org/10.48550/arXiv.2010.03142.

[10] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training [EB/OL]. 2018. https://paperswithcode.com/paper/improving-language-understanding-by.

[11] WAYNE X Z, ZHOU K, LI J, et al. A Survey of Large Language Models. [EB/OL]. [2023-08-25]. DOI: https://doi.org/10.48550/arXiv.2303.18223.

[12] KRALJEVIC Z, SHEK A, BEAN D, et al. MedGPT: Medical Concept Prediction from Clinical Narratives[J]. 2021. DOI: https://doi.org/10.48550/arXiv.2107.03134.

[13] THAWKAR O, SHAKER A, SHAJI MULLAPPILLY S, et al. XrayGPT: Chest Radiographs Summarization using Medical Vision-Language Models[EB/OL]. [2023-08-25]. DOI:https://doi.

org/10.48550/arXiv.2306.07971

[14] GAO W H, DENG Z, NIU Z Y, et al. OphGLM: training an ophthalmology large language - and - vision assistant based on instructions and dialogue[EB/OL]. 2023: arXiv: 2306.12174. http://arxiv.org/abs/2306.12174

[15] XIONG H, WANG S, ZHU Y, et al. DoctorGLM: Fine - tuning your Chinese Doctor is not a Herculean Task[EB/OL]. [2023 - 08 - 25]. DOI: https://doi.org/10.48550/arXiv.2304.01097.

[16] TAN Y, LI M, HUANG Z, et al. MedChatZH: a Better Medical Adviser Learns from Better Instructions[EB/OL]. [2023 - 08 - 25]. DOI:https://arxiv.org/abs/2309.01114.

[17] WANG H, LIU C, XI N, et al. Huatuo: Tuning llama model with chinese medical knowledge[EB/OL]. [2023 - 08 - 25]. DOI: https://doi.org/10.48550/arXiv.2304.06975.

[18] LEE J, YOON W, KIM S, et al. BioBERT: a pre - trained biomedical language representation model for biomedical text mining [J]. Bioinformatics, 2020, 36(4): 1234 - 1240.

[19] HUANG K, ALTOSAAR J, RANGANATH R. ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission[J]. 2019. DOI: https://doi.org/10.48550/arXiv.1904.05342.

[20] LUO Y, ZHANG J, FAAN S, et al. BioMedGPT: Open Multimodal Generative Pre - trained Transformer for BioMedicine [M]. Ithaca; Cornell University Library, arXiv.org. 2023. DOI: https://doi.org/10.48550/arXiv.2308.09442.

[21] WU C Y, LIN W X, ZHANG X M, et al. PMC - LLaMA: towards building open - source language models for medicine[EB/OL]. 2023: arXiv: 2304.14454. http://arxiv.org/abs/2304.14454

[22] 胡凯文团队. 数字中医大模型入选北京市首批行业大模型典型应用案例[EB/OL]. (2023 - 06 - 28)[2023 - 09 - 18]. https://www.dongfangyy.com.cn/Html/News/Articles/9432.html.

[23] KHERA R, BUTTE A J, BERKWITS M, et al. AI in medicine - JAMA's focus on clinical outcomes, patient - centered care, quality, and equity[J]. JAMA, 2023, 330(9): 818 - 820.

[24] FATANI B. ChatGPT for future medical and dental research[J]. Cureus, 2023, 15(4): e37285.

[25] BROWN C, NAZEER R, GIBBS A, et al. Breaking bias: the role of artificial intelligence in improving clinical decision - making[J]. Cureus, 2023, 15(3): e36415.

[26] RAO A, KIM J, KAMINENI M, et al. Evaluating ChatGPT as an Adjunct for Radiologic Decision - Making [J]. medRxiv, 2023. DOI: 10.1101/2023.02.02.23285399.

[27] SALLAM M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns[J]. Healthcare, 2023, 11(6): 887.

[28] 郭仪, 许斌, 胡楠. 人工智能在辅助中医临床领域的研究现状与展望[J]. 中华中医药学刊, 2021, 39(6): 76 - 78.

[29] 中医大模型来了,发布现场训练过程大公开,超 1100 万条知识图谱数据[DB/OL]. (2023 - 08 - 02)[2023 - 08 - 12]. https://www.thepaper.cn/newsDetail_forward_24068128.

[30] 马武仁, 弓孟春, 戴辉, 等. 以 ChatGPT 为代表的大语言模型在临床医学中的应用综述[J]. 医学信息学杂志, 2023, 44(7): 9 - 17.

[31] 孔祥勇. ChatGPT 在医疗行业的创新机遇与挑战 [J]. 张江科技评论,2023(2): 68 - 71.

[32] ESHGHIE M. ChatGPT as a Therapist Assistant: A Suitability Study[EB/OL]. [2023 - 08 - 25]. DOI: https://doi.org/10.48550/arXiv.2304.09873.

[33] SINGHAL K, AZIZI S, TU T, et al. Large language models encode clinical knowledge[J]. Nature, 2023, 620(7972): 172 - 180.

[34] SINGHAL K, TU T, GOTTWEIS J, et al. Towards Expert - Level Medical Question Answering with Large Language Models [EB/OL]. [2023 - 08 - 25]. DOI: https://doi.org/10.48550/arXiv.2305.09617.

[35] AYERS J W, POLIAK A, DREDZE M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum[J]. JAMA Intern Med, 2023, 183(6): 589 - 596.

[36] 张柯欣, 孙艳秋. 中医人工智能模型的设计要素研究[J]. 中华中医药学刊, 2020, 38(12): 116 - 118.

[37] 吕柯, 马英鹏. 当前我国医疗卫生资源分配中存在的主要问题[J]. 中华疾病控制杂志, 2008, 12(6): 626 - 629.

[38] OPEN A I, ACHIAM J, ADLER S, et al. GPT - 4 technical report [EB/OL]. 2023: arXiv: 2303.08774. http://arxiv.org/abs/2303.08774

[39] SPATARO J. Introducing Microsoft 365 Copilot—your copilot for work. The Official Microsoft Blog [EB/OL]. 2023. https://blogs.microsoft.com/blog/2023/03/16/introducing - microsoft - 365 - copilot - your - copilot - for - work/.

[40] GHAHRAMANI, Z. Introducing PaLM 2. [EB/OL]. [2023 - 09 - 20]. Google, 2023. https://blog.google/technology/ai/google - palm - 2 - ai - large - language - model/.

[41] PATEL S B, LAM K. ChatGPT: the future of discharge summaries? [J]. Lancet Digit Health, 2023, 5(3): e107 - e108.

[42] TOUVRON H, LAVRIL T, IZACARD G, et al. LLaMA: Open and Efficient Foundation Language Models[EB/OL]. [2023 - 08 - 25]. DOI: https://doi.org/10.48550/arXiv.2302.13971.

[43] BROWN T, MANN B, RYDER N, et al. Language Models are Few - Shot Learners[J]. 2020. DOI: https://doi.org/10.48550/arXiv.2005.14165.

[44] VAN ES B, RETEIG L C, TAN S C, et al. Negation detection in Dutch clinical texts: an evaluation of rule - based and machine learning methods[J]. BMC Bioinformatics, 2023, 24(1): 10.

[45] THIRUNAVUKARASU A J, TING DSJ, ELANGOVAN K, et al. Large language models in medicine[J]. Nat Med, 2023, 29(8): 1930 - 1940.

[46] 胡振生, 杨瑞, 朱嘉豪, 等. 大语言模型在医学领域的研究与应用发展[J]. 人工智能, 2023, 9(4): 10 - 19.

[47] YANG X, CHEN A K, POURNEJATIAN N, et al. A large language model for electronic health records[J]. NPJ Digit Med, 2022, 5(1): 194.

[48] AGRAWAL M, HEGSELMANN S, LANG H, et al. Large Language Models are Few - Shot Clinical Information Extractors[EB/OL]. [2023 - 08 - 25]. DOI: https://doi.org/10.48550/arXiv.2205.12689.

[49] ARIF T B, MUNAF U, UL - HAQUE I. The future of medical education and research: is ChatGPT a blessing or blight in disguise? [J]. Med Educ Online, 2023, 28(1): 2181052.

[50] LEE P, BUBECK S, PETRO J. Benefits, limits, and risks of

GPT-4 as an AI chatbot for medicine[J]. N Engl J Med, 2023, 388(13): 1233-1239.

[51] FERRUZ N, SCHMIDT S, HÖCKER B. ProtGPT2 is a deep unsupervised language model for protein design[J]. Nat Commun, 2022, 13: 4348.

[52] INAGAKI T, KATO A, TAKAHASHI K, et al. LLMs can generate robotic scripts from goal-oriented instructions in biological laboratory automation[EB/OL]. [2023-08-25]. DOI: https://doi.org/10.48550/arXiv.2304.10267.

[53] 袁欢欢, 全祉悦, 林祎, 等. 当医学课程考核遇上生成式人工智能: 价值、挑战与未来展望[J]. 中国医学教育技术, 2023, 37(3): 254-259.

[54] 瞿星, 杨金铭, 陈滔, 等. ChatGPT 对医学教育模式改变的思考[J]. 四川大学学报(医学版), 2023, 54(5): 937-940.

[55] ABD-ALRAZAQ A, ALSAAD R, ALHUWAIL D, et al. Large language models in medical education: opportunities, challenges, and future directions[J]. JMIR Med Educ, 2023, 9: e48291.

[56] 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(10): 227-240.

[57] KUNG T H, CHEATHAM M, MEDENILLA A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models[J]. PLoS Digit Health, 2023, 2(2): e0000198.

[58] LEE H. The rise of ChatGPT: exploring its potential in medical education[J]. Anat Sci Educ, 2024, 17(5): 926-931.

[59] MASTERS K. Response to: aye, AI! ChatGPT passes multiple-choice family medicine exam[J]. Med Teach, 2023, 45(6): 666.

[60] ZHOU H, GU B, ZOU X, et al. A Survey of Large Language Models in Medicine: Progress, Application, and Challenge[J]. ArXiv, 2023, abs/2311.05112.

[61] RUDINGER R, MAY C, VAN DURME B. Social bias in elicited natural language inferences[C]//Proceedings of the First ACL Workshop on Ethics in Natural Language Processing. Valencia, Spain. Stroudsburg, PA, USA: Association for Computational Linguistics, 2017: W17-1609.

[62] JIN Q, YANG Y F, CHEN Q Y, et al. GeneGPT: augmenting large language models with domain tools for improved access to biomedical information[J]. ArXiv, 2023: 2304.09667v3.

[63] 肖仰华, 徐一丹. 大规模生成式语言模型在医疗领域的应用: 机遇与挑战[J]. 医学信息学杂志, 2023, 44(9): 1-11.

[64] SHUSTER K, KOMEILI M, ADOLPHS L, et al. Language Models that Seek for Knowledge: Modular Search & Generation for Dialogue and Prompt Completion[J]. 2022. DOI: https://doi.org/10.48550/arXiv.2203.13224.

[65] 熊明辉, 池骁. 论生成式大语言模型应用的安全性——以 ChatGPT 为例[J]. 山东社会科学, 2023(5): 79-90.

[66] 周学峰. 生成式人工智能侵权责任探析[J]. 比较法研究, 2023(4): 117-131.

[67] PAN S R, LUO L H, WANG Y F, et al. Unifying large language models and knowledge graphs: a roadmap[J]. IEEE Trans Knowl Data Eng, 2024, 36(7): 3580-3599.

[68] HOLZINGER A, LANGS G, DENK H, et al. Causability and explainability of artificial intelligence in medicine[J]. Wiley Interdiscip Rev Data Min Knowl Discov, 2019, 9(4): e1312.

[69] BOMMASANI R, HUDSON D A, ADELI E, et al. On the Opportunities and Risks of Foundation Models[J]. 2021. DOI: https://doi.org/10.48550/arXiv.2108.07258.

[70] JIN Q, YUAN Z, XIONG G Z, et al. Biomedical question answering: a survey of approaches and challenges[J]. ACM Comput Surv, 2023, 55(2): 1-36.

[71] XU J, KIM S, SONG M, et al. Building a PubMed knowledge graph[J]. Sci Data, 2020, 7(1): 205.

[72] ROBERTS A, RAFFEL C, SHAZEER N. How Much Knowledge Can You Pack Into the Parameters of a Language Model?[J]. 2020. DOI: https://doi.org/10.48550/arXiv.2002.08910.

[73] 高远, 张仕娜, 盛博洋, 等. 由 ChatGPT 引发中医智能诊断研究中数据问题的思考[J]. 湖南中医药大学学报, 2023, 43(7): 1320-1324.

[74] TU T, AZIZI S, DRIESS D, et al. Towards Generalist Biomedical AI[EB/OL]. [2023-08-25]. DOI: https://doi.org/10.48550/arXiv.2307.14334.

[75] 颜见智, 何雨鑫, 骆子烨, 等. 生成式大语言模型在医疗领域的潜在典型应用与面临的挑战[J]. 医学信息学杂志, 2023, 44(9): 23-31.

[76] 支振锋. 生成式人工智能大模型的信息内容治理[J]. 政法论坛, 2023, 41(4): 34-48.

[77] KARABACAK M, MARGETIS K. Embracing large language models for medical applications: opportunities and challenges[J]. Cureus, 2023, 15(5): e39305.

[78] 李紫菡, 周双双, 唐国恒, 等. ChatGPT 概述及应用研究[J]. 债券, 2023(6): 70-75.

[79] 蔡士林, 杨磊. ChatGPT 智能机器人应用的风险与协同治理研究[J]. 情报理论与实践, 2023, 46(5): 14-22.

[80] 郭天赐, 陈继鑫, 余伟杰, 等. 人工智能在骨关节炎诊疗中的应用进展[J]. 中国全科医学, 2023, 26(19): 2428-2433.

[81] 郝国强, 李星莹, 杨琴. 通用人工智能(AGI)的技术、应用及安全问题: 以 ChatGPT 为例[J]. 传承, 2023(1): 100-110.

[82] 郭一帆. 医疗领域人工智能的应用, 风险及对策[J]. 西华大学学报(哲学社会科学版), 2021, 40(1): 44-51.

[83] 谭佐财. ChatGPT 的法律风险与治理路径[J]. 湖南科技大学学报(社会科学版), 2023, 26(3): 117-125.

[84] 关于印发互联网诊疗监管细则(试行)的通知[DB/OL]. 医政司, (2022-03-15)[2023-08-21]. http://www.nhc.gov.cn/yzygj/s3594q/202203/fa87807fa6e1411e9afeb82a4211f287.shtml.

[85] RAY P P E. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope[J]. Internet Things Cyber Phys Syst, 2023, 3: 121-154.

[86] LIEBRENZ M, SCHLEIFER R, BUADZE A, et al. Generating scholarly content with ChatGPT: ethical challenges for medical publishing[J]. Lancet Digit Health, 2023, 5(3): e105-e106.

[87] GATES B. Will ChatGPT transform healthcare?[J]. Nat Med, 2023, 29(3): 505-506.