

大语言模型对话中医及其潜在的局限性*

李花¹, 王玉曼¹, 刘晓雨¹, 孙润雪², 王绍坡², 杨倩^{2,3}, 杜艳茹^{2,4**}

(1. 河北中医药大学中医学院 石家庄 050020; 2. 河北省中医院 石家庄 050011; 3. 河北省浊毒证重点实验室 石家庄 050031; 4. 河北中西医结合胃肠病重点实验室 石家庄 050011)

摘要:生成式预训练转换器(Generative pre-training transformer, GPT)和大型语言模型(Large language models, LLMs)的出现为医学领域带来了变革性应用。在中医药(Traditional Chinese medicine, TCM)领域, LLMs为解决临床工作流程效率低下和改善患者体验提供了独特的机会。然而,这些前景也伴随着挑战,包括数据质量、安全和隐私、虚假信息、伦理以及其他各种问题。本文系统阐述了LLMs在中医领域的应用,以及其在改善中医诊疗服务、提升教学效能、优化医疗管理流程等方面的应用价值,并深入分析了实际落地过程中可能遇到的技术与伦理障碍。为确保LLMs在中医临床中的安全、公正应用,必须建立完善的保障机制。我们呼吁临床医师与研究人员共同应对挑战,在优化模型性能的同时有效管控风险,这将为中医药文化的传承与创新注入强劲动力。

关键词:大语言模型 中医 患者体验 工作流程 挑战

DOI: 10.11842/wst.20250212001 CSTR: 32150.14.wst.20250212001 中图分类号: TP181 文献标识码: A

1 前言

2022年, OpenAI推出的对话型人工智能ChatGPT引发了全球关注,该模型通过模拟人类对话方式可实现多领域深度交流,其核心技术源于基于海量文本数据训练的大语言模型(Large language models, LLMs)^[1-3]。在现代医学领域, LLMs正掀起一场技术革命。BioMedLM是由斯坦福大学基础建模中心与MosaicML联合研发的医学专用LLMs,旨在处理生物医学领域的语言任务,该模型采用27亿参数规模(与ChatGPT基础版30亿参数相当),其训练数据完全源自PubMed数据库的摘要及全文文献^[4-7]。除了生物医学领域, LLMs也已经在医疗保健^[8]、心理学^[9]和神经学^[10]等领域显示出潜在应用价值。中医学与现代医学同属人类健康科学体系,二者在认知范式与方法论上具有互补性。借鉴现代医学大语言模型的研究成果,

可为中医药领域的人工智能应用提供重要启示。

聚焦中医领域,人工智能技术在自动化舌诊系统^[11]、体质辨识^[12]、疾病预测^[13-16]、处方推荐^[17]等方面得到了重要发展。然而,面对中医数据的复杂性和临床诊疗的多样性,智能诊疗模型仍需提升适应能力和诊断精度^[18-19]。随着中医药大数据平台建设的不断深入,医院信息系统及文献管理平台已积累了大量结构化诊疗文本、经典医案及科研数据,这些具有领域特异性的资源为LLMs的适应性微调提供了关键的训练数据,通过构建中医药领域特定的预训练-微调范式, LLMs能够有效改善患者就诊体验、提升教学效率、优化服务者管理等多种应用场景的需求。当然,技术创新往往伴随挑战,本文最后将探讨LLMs应用部署中的关键难题及其应对策略。

收稿日期:2025-02-12

修回日期:2025-07-08

* 国家中医药管理局项目(GZY-KJS-2023-013):《中医脾胃病临床标准化、数字化、工程化与工程化研究》负责人:杜艳茹;河北省重点研发计划(23377701D):《慢性萎缩性胃炎中医病证诊断数字化研究》负责人:杜艳茹。

** 通讯作者:杜艳茹(ORCID:0000-0002-4877-4959),主要从事方向:中医脾胃病的临床与基础研究。

2 中医大语言模型研究现状

LLMs 凭借其在复杂文本处理与语义理解方面的卓越能力,已在医学人工智能领域引起广泛关注^[20-21]。中国工程院院士、国医大师张伯礼在接受专访时表明,以“数智本草”和“天河灵枢”为代表的大模型为中医药产业的创新和发展提供了新的动力^[22]。近年来,大语言模型在中医药领域的应用呈现蓬勃发展态势,已形成涵盖智能问答、辅助诊疗、病历生成等多元化的应用场景^[23]。中医药领域已相继研发出多款专业大语言模型,包括:专攻中医执业考试与诊断推理的TCM-GPT-7B^[24]、在中医专业测试中表现突出的HuatuoGPT-II^[25]、专注于中医临床问答的ZhongJing^[26]、针对疫病防治场景的EpidemicCHAT^[27]、以及在中草药处方推荐方面表现卓越的Lingdan^[28]。值得一提的是,Huang-Di大模型开创性地实现了中医古籍知识问答功能,突破了传统数字化研究的局限^[29],而岐黄问道LLMs^[30]与CPMI-ChatGLM^[31]则分别在中医知识问答与药物配伍建议方面取得重要突破。

在中医药现代化转型进程中,大语言模型的应用正展现出独特的学术价值与实践潜力。潘晨浩等^[32]研究发现,ChatGPT可以对中医药范畴的基本性问题给出快速准确的回答,对临床诊疗问题给出符合中医思维的答案并给出合理药方。吕关成^[33]研究发现,智慧中医线上“四诊”等与人工智能技术结合的智慧中医诊疗新模式可以提高患者就诊体验及满意度。Li等^[34]比较了预训练的GPT-3.5、微调后的GPT-3.5、预训练的GPT-4和预训练的Llama3四种模型在腧穴位置关系提取任务中的应用效果,结果发现微调后的GPT-3.5模型展示了出色的性能。张童等^[35]采用文献研究结合实验研究的方法,选取9个不同类型的大语言模型,设计专业Prompt测试系统评估了大语言模型的中医药专业能力,结果表明,通用大语言模型在泛化性、易读性、安全性方面具有优势。张默等人^[36]基于350例门诊患者舌象数据,对舌象的初步描述(舌苔、舌质、舌形)进行了预处理、特征地提取以及模式识别,应用阿里达摩院开源的通义千问VL模型,通过人工智能技术分析验证多模态模型在中医舌诊辨识与疾病诊断(HP感染)中的效能,结果表明,该模型具有快速学习的优势,能通过分析舌象特征来检测幽门螺杆菌感染。

目前,LLMs是通过大量的文本数据进行训练,中医领域LLMs主要应用于辨证论治、处方推荐和古籍

解析等文本处理任务。当前的中医四诊多模态研究还处在初起阶段,与中医认知思维有一定差距,模型性能欠佳。林树元等^[37]从认知神经科学的全局工作空间理论出发,提出了“统一表征与模型内融合”的技术发展方向,该研究思路为中医四诊多模态研究奠定了新的理论与方法学基础,然而关于更多模态大语言模型的实践性研究有待进一步发展。

3 优化患者就医体验

3.1 促进分流

LLMs的核心优势在于其能够通过动态双向交互机制,整合个体化诊疗情境、自适应医学知识更新及类人认知推理能力,可基于患者特异性数据及实时更新的临床信息进行动态调适,进而生成针对性诊疗优化方案,并有效支撑患者分级诊疗与资源协调配置的科学化决策。以眼科领域的研究为例:LLMs作为远程分流患者的工具能够提供从常见症状(如视力模糊、眼痛发红)到特定疾病(青光眼、白内障)等方面的信息并回答相关问题;在实际应用方面,其可通过程序界面和医院信息系统连接,帮助患者提供主诉、其他症状和病史,与医疗服务提供者进行远程协调,并帮助患者预约^[38]。

3.2 改善患者就医体验

在医疗流程管理中,患者分流与协调通常是门诊接待的重要前置环节。精准分流能快速引导患者至合适的科室或医生,从而避免过度拥挤的门诊环境并提高诊疗效率。然而,现有通信系统仍高度依赖人工操作,兼具劳动密集和知识密集的双重特性。面对日益多元的患者需求和日趋复杂的医疗服务,传统人工沟通模式已难以满足高效精准的服务要求。最近的研究表明^[39],LLMs不仅可以为患者提供丰富的移情反应,同时可与护士合作有效协助患者与护士之间的沟通,显著提高患者对咨询过程的满意度。这表明LLMs在临床环境中的应用可有效弥补传统沟通模式的不足:在问诊过程中提供富有共情心的指导,增强患者在咨询过程中的情感体验。更重要的是,LLMs的应用能显著缩短患者候诊时间,改善就医体验,从而优化门诊流程,提升整体诊疗效率和质量。患者分流协调是中西医共有的服务模式,LLMs在现代医学临床流程优化方面的潜力,在中医药体系中同样具有广阔的应用。近期^[40],中国中医科学院团队基于

YAYI 2-30B 微调中医大语言模型,采用 Deep- Speed ZeRO-3 分布式训练策略,整理中医书籍和临床病案中的关系数据,建立智能体辨证推理模型,利用智能体协作架构实现中医大语言模型与辨证推理模型交互,结果发现,基于大语言模型的中医智能预问诊系统可以在一定程度上模拟临床中医医生的辨证思维,提高医生诊疗效率及患者就医体验。

4 提高医学教学效率

当代医学教育范式已突破传统课堂讲授与纸质教材主导的知识传递模式,在数字技术革命的协同驱动下,教学形态正经历系统性变革。目前,许多学生已将教科书换成了第三方资源^[41-42]。最近,以 ChatGPT 为代表的 LLMs 凭借其生成式人工智能技术与自然语言处理能力,正逐步演变为医学教育生态系统中的新型认知工具。医学知识体系固有的复杂性与知识体量的指数级增长使学习者普遍面临认知负荷过载的困境,LLMs 通过对话式交互界面实现了知识解构与重构的智能化支持,这些工具可作为“远程辅导员”,帮助学生总结概念、提炼要点、生成练习题,并作为高效参考资料^[43]。

技术驱动的教育变革正在重塑现代医学教学体系,为中医药教育的数字化转型开辟全新路径。随着生成式人工智能的普及,医学生的信息获取方式正从传统的搜索引擎依赖转向智能对话交互。将 LLMs 引入临床教学,使学生能够更专注于真实医疗场景中的患者诊疗。在此过程中,培养学生批判性评估 AI 应答的能力尤为重要,这不仅能强化医疗决策中的伦理意识,更能塑造智能化时代的数字公民素养^[44]。在中医药教育领域,大语言模型对于培养兼具中医思维和现代医学知识的复合型人才至关重要。结合中医药基准测试数据集训练大语言模型,能够为中医教学提供更标准的考核系统,有利于减轻教学评测工作的压力,提高中医知识考核的效率和准确性。同时,利用自然语言处理技术,将大语言模型应用于中医药教学中,可以快速从大量中医药文献中提取核心知识,帮助学生更加深入地理解传统中医学理论。

尽管如此,生成式大语言模型本质上还是一种基于数据处理的“聊天机器人”,本身存在缺乏理解能力、知识立场不坚定、语言是生成而非创造、难以满足个别化学习需求等困难。因此,无论从知识传授还是

情感培养方面看,生成式大语言模型自身均因受到特定价值取向影响、缺乏策略性引导,难以做到因材施教^[45]。尤其在中医领域,生成式人工智能对中医教育的影响仍需置于其哲学根基与文化语境中审慎考量。基于《周易》、河图洛书等诞生的象数思维、阴阳五行等概念是中医理论体系形成的哲学基础,中医讲究“道术相合”。然而,中医学“道、法、术”与人工智能“理、法、数”之间仍然存在桎梏。LLMs 为中医药教育可能带来重要影响,但目前仍然无法取代传统的师徒模式的精神内涵,师承首先在于“精神传承”,有了内在精神的一脉相传,才能得先师“真传”、续圣哲“余脉”^[46]。目前包括 LLMs 在内的人工智能技术虽能进行逻辑推理,但尚无法完全模拟医师对患者全方位的四诊,尤其是中医最看重的“神”以及对“道”的传授。

5 优化医疗服务提供者管理

5.1 优化临床实践

2021 年全球疾病负担 (Global burden of disease, GBD) 数据显示,医疗系统超负荷运转成为导致疾病负担上升的关键因素。这一发现凸显了全球社会亟需采取跨学科协作策略,以应对日益严峻的老龄化挑战^[47]。中医药作为国际替代医学的一部分,在慢性疾病的诊疗及预防方面独具特色和优势,构建中医 LLMs 有助于为老龄化人口提供更为全面的健康管理,将 LLMs 部署到临床工作中可以在每次就诊前对患者信息进行快速摘要和筛选,不仅可以减轻医疗系统的压力同时有助于提升临床决策的精准度。南京中医药大学团队^[48]基于 ChatGLM 模型与 Neo4j 图数据库构建了中医妇科知识图谱,相较传统 Bi-LSTM-CRF 模型在知识抽取上表现更优,显著提升了各项指标,助力临床决策,提高诊疗效率与准确性,展现出较高实用价值。

出院指导作为医院护理与家庭康复的关键纽带,在信息传递和提升患者治疗依从性方面发挥着重要作用。提高患者对出院指导的理解和出院后的自我管理及改善健康状况与降低再入院率相关^[49-50]。研究表明,GPT 生成的出院指导与标准出院指导相比得到了患者更高的评价,同时对患者而言更具解释性和理解性^[51]。在一项单盲观察研究中,LLMs 生成的高质量皮肤科病例报告与人类创建的病例报告几乎没有区别^[52]。目前看来,LLMs 生成的出院指导和病例报告

可作为当前临床工作流程的辅助工具,以增强患者对关键信息的理解和可读性,同时帮助医生完成病例撰写等重复性任务,从而减轻临床医生的疲劳。

5.2 驱动科研革新

LLMs 正在为中医药科研带来新的变革。最近,多项研究表明^[53-56],ChatGPT 在排除标准、筛选文献、评估出版物、生成 Meta 分析以及协助撰写全文等方面可以发挥重要作用,如确定和完善研究问题、进行文献综述和设计研究^[57]。LLMs 还可用于写作任务,如起草提纲、压缩文件以满足字数或字符限制。值得注意的是,学术出版界已着手建立 LLMs 应用的规范体系,众多会议和期刊通过在投稿时使用人工智能检测软件,扩大了有关使用 LLMs 生成文本的政策范围。部分期刊要求对 LLMs 生成内容进行授权,另一些则强制披露 AI 使用情况^[58]。使用 LLMs 生成文本在提高写作效率的同时也面临学术诚信的挑战,并可能引发潜在的知识产权相关法律问题。研究者需保持警惕,避免因使用 LLMs 而产生无意识的抄袭行为。

6 风险、挑战和建议

技术革命在萌芽阶段往往会遇到阻力,即使在科学领域也是如此,这些技术在中医药发展带来可观前景的同时也带来了一些潜在的挑战和威胁。

6.1 数据质量

大数据是人工智能未来发展的基础。LLMs 需要海量训练数据来应对自然语言的复杂性,同时必须持续更新数据并监控潜在偏差。电子病历(Electronic medical record, EHR)正日益成为人工智能技术的重要训练数据集。与许多其他信息系统一样,EHR 数据的质量在完整性、正确性、一致性、可信度和及时性方面参差不齐^[59]。因此,严格的数据清洗和筛选流程至关重要,必须建立完善的质量控制机制以避免“垃圾输入,垃圾输出”现象,确保模型性能的可靠性。中医大数据涵盖 EHR、名家医案、古籍文献等多种类型,其中大量为非结构化文本数据。这些数据普遍存在质量不均或缺失等问题,若预处理不当,很容易导致模型错误或性能表现低下。此外,由于客观化的四诊信息采集普及程度不高,中医药信息系统互联互通仍需加强,“信息孤岛”依然存在,加之中医“望闻问切”数据类型多样、数据结构复杂、数据标准不健全,以及中医药知识产权保护等多方面原因,导致中医领域难以形

成开放优质的大规模中医标注数据集^[60]。

在未来的中医大语言模型研发中,我们建议制定统一的中医数据采集与存储规范,搭建基于区块链的国家级共享平台,通过智能合约实现数据确权与安全流通,激励机构以脱敏数据换取资源或分成可能有助于打破目前中医药“信息孤岛”的问题。关于数据质量不一致的问题,可以通过将 LLMs 与中医药领域的知识表达技术结合,通过微调来适应 LLMs 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,辅助 LLMs 对中医药内涵的准确理解,在其后续应用中,持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。

6.2 虚假信息 and 幻觉

LLMs 有可能提供指令和代码入侵包含敏感医疗信息的计算机系统或生成虚假信息,例如,ChatGPT 能够基于虚构数据生成极具说服力的科学摘要^[61],这一现象已引起学术界的广泛关注^[62-63]。更为严峻的是,面向患者的 GPT 应用可能从根本上动摇医患关系的信任基础。在最近的一项调查中,Martinho A 等人^[64]报告说,尽管意识到人工智能这一新技术的重要性和好处,但医疗服务提供者对人工智能系统和人工智能行业产生了不信任。当 LLMs 应用于医疗场景时,其可能产生的错误医疗建议和诊断决策,一旦涉及公众健康和生命安全,必将引发医患双方的强烈抵触情绪。这种信任危机可能造成不可逆的负面影响,严重阻碍 LLMs 的实际应用,并带来巨大的经济损失。

为确保人工智能生成内容的科学性与可靠性,有必要建立严格的质量控制与验证机制。人工智能生成的医学信息应由独立研究人员与医学专家进行系统监测及评估,并采用随机对照试验及其他严谨的科学试验方法验证其准确性与临床价值。有专家建议,应对人工智能生成内容进行明确标注,制定基于 AI 生成内容的安全管理规范,以防范潜在的学术不端行为,并有效识别伪造或误导性科学内容^[65]。为了降低虚假信息传播的风险,未来的医学 LLMs 应基于准确性与透明度原则构建其训练数据集。进入训练集的数据应经过严格筛选和独立验证,并确保其来源公开透明,以便于外部专家和研究机构进行审查^[62]。在中医领域开发大语言模型时,应参考这些研究观点。

6.3 可解释性

可解释性是确保模型可信度并被医生和患者接受

的关键因素^[66]。然而,当前的LLMs作为黑箱模型,其内部工作机制难以解释,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^[67]。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,加之中医术语的复杂性和多样性,以及中医经验的个体性和主观性,语言模型通常难以完全掌握经典理论的深层含义。此外,中医领域中医流派众多,各学派思想的交叉融合使其解释更加错综复杂。因此,有专家提出,采用知识驱动与数据驱动相结合的方法可以使模型输出更具逻辑性和可解释性。Pan S R等人^[68]认为利用知识图谱增强模型,可以明确存储丰富的事实知识,Xu J等人^[69]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。在中医药领域,利用中医知识图谱来增强中医大语言模型,有助于提高模型输出的逻辑性和可解释性。

6.4 多模态四诊融合

大语言模型在智能问答、辅助诊疗、病历生成、中医执业考试、中草药处方推荐等多场景应用中表现出较好性能。同时也面临缺乏“整体观念”的问题,中医辨证体系是四诊合参,谨守病机,综合面色、舌色、舌苔、脉象的变化,可以揭示人体整体的生理和病理变化。面诊仪、舌诊仪、脉诊仪、目诊仪、声诊采集系统、电子鼻等在各自领域朝智能化、大数据化、迷你化可穿戴方向蓬勃发展,但四诊合参的智能化无论从硬件集成系统,还是四诊融合的智能算法、大数据校验均处在初级发展阶段^[70]。距离构建集望诊、闻诊、问诊、脉诊等信息于一体的多模态中医LLMs尚有一段距离。“望神”作为中医学的重要概念,涉及情感、精神状态、整体气色等复杂信息的全面感知,“望神”属于中医医师个性化的复杂感官判断,目前来看,AI仍然无法完全模仿人类经验和直觉。

6.5 伦理与安全

随着技术突飞猛进,医疗失误可能引发的伦理问题不容忽视。科学界已经广泛讨论了与LLMs相关的伦理性问题^[71-73],与LLMs相关的许多生物学方面的伦理问题反映了现有人工智能在医学应用中的常见问题,如偏见问题^[74]。数据收集或使用过程中的不道德行为以及从业人员的偏见倾向可能会给模型带来偏差^[75-76]。LLMs生成的文本映射了其训练数据,假设文本中存在某些种族、性别、语言和文化相关的偏

见^[77-79],模型可能会延续这一偏见,如果医生、患者和社会因为与可信度相关的因素而失去信任,那么即使经过良好开发和验证的人工智能设备在实际应用中可能仍然会失败。最近一项针对四个商用LLMs在医疗场景中的表现评估显示,所有模型在回答涉及种族医学的问题时均存在传播有害种族观念的现象,且回答内容呈现明显的不一致性^[80]。这一发现提示我们LLMs可能成为种族主义观念的传播载体,从而对医疗实践构成潜在威胁。中医,作为一个有着深厚文化背景和独特思维方式的医学体系,构建中医LLMs有助于中医文化传播,但在开发应用LLMs时,同样需要关注以上提到的风险问题。

对LLMs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。国家应尽快制定相关的法律法规,明确LLMs的知识产权归属和责任分配,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LLMs在中医药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。

LLMs训练需要大量的个人健康信息和诊疗记录,并且在模型提供疾病咨询、中医诊疗建议等应用时,存在个人敏感信息泄露的隐患。因此,在数据采集、模型训练、部署、输出和应用等多个重要环节,必须保证信息的处理分析符合医疗数据保护的标准规定,以确保模型输出的内容不会侵犯用户隐私。与此同时,监管机构应建立健全的框架,保护私人信息,防止未经授权访问或滥用医疗信息系统。针对不同应用场景的AI模型,需要采取灵活且全面的监管策略,确保其在临床实践中的合规性和责任担当^[81]。

7 结语

随着人工智能技术的不断进步,LLMs在中医药领域的应用将更加深入和广泛,为中医药的创新研究和传承提供强有力的支撑。尽管目前存在一些局限性和挑战,但快速发展的LLMs有望融入中医临床工作并提高工作效率。然而,由于它在内容生成中可能产生虚假信息,对LLMs进入辅助诊断和治疗或许仍应采取谨慎态度。未来,我们必须继续批判性地评估这些技术的影响,并采取相关行动和措施以减轻它们可能对社会造成的负面影响。

[利益冲突] 本文不存在任何利益冲突。

参考文献

- 1 BERT: Pre-training of deep bidirectional transformers for language understanding[C/OL]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019: 4171–4186[2025-09-17]. <https://aclanthology.org/N19-1423/>.
- 2 Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. *Adv Neural Inf Process Syst*, 2020,33:1877–1901.
- 3 Le Scao T, Fan A, Akiki C, et al. BLOOM: A 176B-parameter open-access multilingual language model [preprint]. *arXiv*: 2211.05100. 2023. <https://doi.org/10.48550/arXiv.2211.05100>
- 4 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.
- 5 Han Q, Tian S, Zhang J. A PubMedBERT-based classifier with data augmentation strategy for detecting medication mentions in tweets [DB/OL]. *arXiv*:2112.02998. 2021. <https://doi.org/10.48550/arXiv.2112.02998>
- 6 Luo R Q, Sun L A, Xia Y C, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining[J]. *Brief Bioinform*, 2022, 23(6):bbac409.
- 7 Venigalla A, Frankle J, Carbin M. Biomedlm: a domain-specific large language model for biomedical text[J]. *MosaicML*. 2022, 23(3):2.
- 8 Thirunavukarasu A J, Ting D S J, Elangovan K, et al. Large language models in medicine[J]. *Nat Med*, 2023, 29(8):1930–1940.
- 9 Suzuki R, Arita T. An evolutionary model of personality traits related to cooperative behavior using a large language model[J]. *Sci Rep*, 2024, 14(1):5989.
- 10 Tuckute G, Sathe A, Srikant S, et al. Driving and suppressing the human language network using large language models[J]. *Nat Hum Behav*, 2024, 8(3):544–561.
- 11 Hu M C, Lan K C, Fang W C, et al. Automated tongue diagnosis on the smartphone and its applications[J]. *Comput Methods Programs Biomed*, 2019, 174:51–64.
- 12 孟月, 崔殿坤, 刘立安, 等. 从人脸生物特征进行中医体质辨识的思路和方法探讨[J]. *中华中医药杂志*, 2020, 35(9):4.
- 13 Li J, Yuan P, Hu X J, et al. Discussion on identification of TCM constitution based on facial biological features[J]. *China Journal of Traditional Chinese Medicine and Pharmacy*, 2020, 35(9):4.
- 14 Li J, Yuan P, Hu X J, et al. A tongue features fusion approach to predicting prediabetes and diabetes with machine learning[J]. *J Biomed Inform*, 2021, 115:103693.
- 15 Chen J H, Sun Y Y, Li J, et al. In-depth metaproteomics analysis of tongue coating for gastric cancer: a multicenter diagnostic research study[J]. *Microbiome*, 2024, 12(1):6.
- 16 Cui J, Song L C. Wrist pulse diagnosis of stable coronary heart disease based on acoustics waveforms[J]. *Biomed Res Int*, 2022, 214:106550.
- 17 Quanyu E. Pulse signal analysis based on deep learning network[J]. *Biomed Res Int*, 2022, 2022:6256126.
- 18 Dong X, Zhao C X, Song X P, et al. PresRecST: a novel herbal prescription recommendation algorithm for real-world patients with integration of syndrome differentiation and treatment planning[J]. *J Am Med Inform Assoc*, 2024, 31(6):1268–1279.
- 19 宋逸杰, 马素亚, 戴亚盛, 等. 人工智能辅助中医辨证的关键问题与技术挑战[J]. *中国工程科学*, 2024, 26(2):234–244.
- 20 Song Y J, Ma S Y, Dai Y H, et al. AI-assisted TCM syndrome differentiation: key issues and technical challenges[J]. *Strategic Study of CAE*, 2024, 26(2):234–244.
- 21 杨涛, 王欣宇, 朱垚, 等. 大语言模型驱动的中医智能诊疗研究思路与方法[J]. *南京中医药大学学报*, 2023, 39(10):967–971.
- 22 Yang T, Wang X Y, Zhu Y, et al. Research ideas and methods of intelligent diagnosis and treatment of traditional Chinese medicine driven by large language model[J]. *Journal of Nanjing University of Traditional Chinese Medicine*, 2023, 39(10):967–971.
- 23 Moor M, Banerjee O, Abad Z S H, et al. Foundation models for generalist medical artificial intelligence[J]. *Nature*, 2023, 616(7956): 259–265.
- 24 Truhn D, Reis-Filho J S, Kather J N. Large language models should be used as scientific reasoning engines, not knowledge databases[J]. *Nat Med*, 2023, 29(12):2983–2984.
- 25 陈曦. 中国工程院院士张伯礼:人工智能与中医药结合是必然趋势[N]. *科技日报*, 2024-05-17(5).
- 26 Liu C, Sun K J, Zhou Q Q, et al. CPML-ChatGLM: parameter-efficient fine-tuning ChatGLM with Chinese patent medicine instructions[J]. *Sci Rep*, 2024, 14(1):6403.
- 27 Yang G X, Liu X H, Shi J Y, et al. TCM-GPT: efficient pre-training of large language models for domain adaptation in Traditional Chinese Medicine[J]. *Comput Methods Programs Biomed Update*, 2024, 6: 100158.
- 28 Chen J, Wang X, Ji K, et al. HuatuoGPT-II, one-stage training for medical adaption of llms[DB/OL]. *arXiv preprint arXiv*: 2311.09774, 2023.<https://doi.org/10.48550/arXiv.2311.09774>
- 29 Yang S H, Zhao H J, Zhu S B, et al. Zhongjing: enhancing the Chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, 38(17):19368–19376.
- 30 Zhou Z Z, Yang T, Hu K F. Traditional Chinese medicine epidemic prevention and treatment question-answering model based on LLMs [C]//2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Istanbul, Turkiye. IEEE, 2023:4755–4760.
- 31 Hua R, Dong X, Wi Y, et al. Lingdan: enhancing encoding of traditional Chinese medicine knowledge for clinical reasoning tasks with large language models[J]. *J Am Med Inform Assoc*, 2024, 31(9): 2019–2029.
- 32 张君冬, 杨松桦, 刘江峰, 等. AIGC赋能中医古籍活化: Huang-Di 大

- 模型的构建[J]. 图书馆论坛, 2024, 44(10):103-112.
- Zhang J D, Yang S H, Liu J F, et al. AIGC empowering the revitalization of ancient books on traditional Chinese medicine: building the Huang-Di Large language model[J]. Library Forum, 2024, 44(10):103-112.
- 30 杨乐乐,王哲,姚克宇,等. 大语言模型在中医药领域的应用展望思考[J]. 中华中医药学刊, 2025, 43(2):16-24.
 - 31 Liu C, Sun K J, Zhou Q Q, et al. CPMI-ChatGLM: parameter-efficient fine-tuning ChatGLM with Chinese patent medicine instructions[J]. Sci Rep, 2024, 14(1):6403.
 - 32 潘晨浩,周明. ChatGPT在中医药领域中的应用前景[J]. 医学信息, 2023, 36(18):15-20.
 - Pan C H, Zhou M. The application prospect of ChatGPT in the field of traditional Chinese medicine[J]. Journal of Medical Information, 2023, 36(18):15-20.
 - 33 吕关成. 智慧中医诊疗新模式对患者就医体验的影响[J]. 中医药管理杂志, 2023, 31(12):227-229.
 - 34 Li Y M, Peng X Q, Li J F, et al. Relation extraction using large language models: a case study on acupuncture point locations[J]. J Am Med Inform Assoc, 2024, 31(11):2622-2631.
 - 35 张童,王一凡,王若佳,等. 面向中医药领域专业能力的生成式大语言模型对比研究[J/OL]. 中华中医药学刊, 1-18 [2025-04-05].
 - Zhang T, Wang Y F, Wang R J, et al. Comparative study of generative large language models for professional abilities in the field of traditional Chinese medicine[J/OL]. Chinese Archives of Traditional Chinese Medicine, 1-18 [2025-04-05].
 - 36 张默,金雪,李宇明,等. 基于多模态大模型探索智能化医疗在中医舌诊中的临床研究[J]. 中药药理与临床, 2025, 41(3):72-76.
 - Zhang M, Jin X, Li Y M, Application of intelligent medical care on tongue diagnosis in traditional Chinese medicine based on multimodal large models[J]. Pharmacology and Clinics of Chinese Materia Medica, 2025, 41(3):72-76.
 - 37 林树元,黄慧雯,刘畅,等. 大模型视角下中医四诊合参的认知机制与多模态研究方法[J]. 中华中医药杂志, 2025, 40(1):97-102.
 - Lin S Y, Huang H W, Liu C, et al. Cognitive mechanisms and multimodal research methodologies of traditional Chinese medicine diagnostics under the large language model perspective[J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2025, 40(1): 97-102.
 - 38 Betzler B K, Chen H C, Cheng C Y, et al. Large language models and their impact in ophthalmology[J]. Lancet Digit Health, 2023, 5(12): e917-e924.
 - 39 Wan P X, Huang Z G, Tang W J, et al. Outpatient reception *via* collaboration between nurses and a large language model: a randomized controlled trial[J]. Nat Med, 2024, 30(10):2878-2885.
 - 40 刘奕清,李迎,杨洪军,等. 基于大语言模型的中医智能预问诊系统的构建与验证[J]. 中医杂志, 2025, 66(9):895-900.
 - Liu Y Q, Li Y, Yang H J, et al. The construction and validation of an intelligent pre-diagnosis system for traditional Chinese medicine based on large language models[J]. Journal of Traditional Chinese Medicine, 2025, 66(9):895-900.
 - 41 Buchanan B B, Allen G B, Pumphrey C M, et al. Engaging medical students in the foundational curriculum using third-party resources[J]. Med Teach, 45(3), 336-336.
 - 42 Wothe J K, Wanberg L J, Hohle R D, et al. Academic and wellness outcomes associated with use of anki spaced repetition software in medical school[J]. J Med Educ Curric, Dev, 2023, 10: 23821205231173289.
 - 43 Benítez T M, Xu Y, Boudreau J D, et al. Harnessing the potential of large language models in medical education: promise and pitfalls[J]. J Am Med Inform Assoc, 2024, 31(3):776-783.
 - 44 Seth P, Hueppchen N, Miller S D, et al. Data science as a core competency in undergraduate medical education in the age of artificial intelligence in health care[J]. JMIR Med Educ, 2023, 9:e46344.
 - 45 颜士刚,胡修磊,李文光. 生成式大语言模型能有效实现对话式教学吗[J]. 现代远程教育研究, 2025, 37(2):52-61, 72.
 - Yan Shigang, Hu Xiulei, Li Wenguang. Can Generative Large Language Models Effectively Implement Dialogic Teaching[J]. Modern Distance Education Research, 2025, 37(2):52-61, 72.
 - 46 王芳,张洪春. 师承教育对中医文化守正创新的重要价值及其实践路径[J]. 中医杂志, 2023, 64(15):1513-1518.
 - Wang F, Zhang H C. The important value and practical path of inheritance education in the integrity and innovation of traditional Chinese medicine culture[J]. Journal of Traditional Chinese Medicine, 2023, 64(15):1513-1518.
 - 47 Christensen, Steffan Witttrup McPhee. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021[J]. Lancet, 2024, 403(10440):2133-2161.
 - 48 汤少梁,赵楠,龙秋予,等. 基于ChatGLM的中医妇科知识图谱自动化构建与临床决策支持研究[J]. 中华中医药学刊, [2025-05-23].
 - Tang S L, Zhao N, Long S Q, et al. Research on the automated construction and clinical decision support of traditional Chinese medicine Gynecology knowledge graphs based on ChatGLM[J]. Chinese Archives of Traditional Chinese Medicine, [2025-05-23]
 - 49 Burns S T, Amobi N, Chen J V, et al. Readability of patient discharge instructions[J]. J Gen Intern Med, 2022, 37(7):1797-1798.
 - 50 DeSai C, Janowiak K, Secheli B, et al. Empowering patients: simplifying discharge instructions[J]. BMJ Open Quality, 2021, 10(3): e001419.
 - 51 Huang T, Safranek C, Socrates V, et al. Patient-representing population's perceptions of GPT-generated versus standard emergency department discharge instructions: randomized blind survey assessment [J]. J Med Internet Res, 2024, 26:e60336.
 - 52 Dunn C, Hunter J, Steffes W, et al. Artificial intelligence-derived dermatology case reports are indistinguishable from those written by

- humans: a single-blinded observer study[J]. *J Am Acad Dermatol*, 2023, 89(2):388–390.
- 53 Qureshi R, Shaughnessy D, Gill K A R, et al. Are ChatGPT and large language models “the answer” to bringing us closer to systematic review automation?[J]. *Systematic Reviews*, 2023, 12(1):72.
 - 54 Mahuli S A, Rai A, Mahuli A V, et al. Application ChatGPT in conducting systematic reviews and meta-analyses[J]. *Br Dent J*, 2023, 235(2):90–92.
 - 55 Khraisha Q, Put S, Kappenberg J, et al. Can large language models replace humans in systematic reviews? Evaluating GPT-4’s efficacy in screening and extracting data from peer-reviewed and grey literature in multiple languages[J]. *Res Synth Methods*, 2024, 15(4):616–626.
 - 56 Gwon Y N, Kim J H, Chung H S, et al. The use of generative AI for scientific literature searches for systematic reviews: ChatGPT and microsoft Bing AI performance evaluation[J]. *JMIR Med Inform*, 2024, 12:e51187.
 - 57 Sanmarchi F, Bucci A, Nuzzolese A G, et al. A step-by-step researcher’s guide to the use of an AI-based transformer in epidemiology: an exploratory analysis of ChatGPT using the STROBE checklist for observational studies[J]. *Z Gesundh Wiss*, 2023:1–36.
 - 58 Rengers T A, Thiels C A, Salehinejad H. Academic surgery in the era of large language models: a review[J]. *JAMA Surg*, 2024, 159(4):445–450.
 - 59 Kahn M G, Callahan T J, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data[J]. *EGEMS*, 2016, 4(1):1244.
 - 60 杨涛, 朱学芳. 中医辨证智能化研究现状及发展趋势[J]. *南京中医药大学学报*, 2021, 37(4):597–601.
 - Yang T, Zhu X F, Discussion on the status and development trend of research on intellectualization of Chinese medicine syndrome differentiation[J]. *J Nanjing Univ Tradit Chin Med*, 2021, 37(4):597–601.
 - 61 Gao C A, Howard F M, Markov N S, et al. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers[J]. *BioRxiv*, 2023, 388."https://doi.org/10.1056/NEJMs2214184:1233–1239.
 - 62 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.
 - 63 Spitale G, Biller-Andorno N, Germani F. AI model GPT-3 (dis)informs us better than humans[J]. *Sci Adv*, 2023, 9(26):eadh1850.
 - 64 Martinho A, Kroesen M, Chorus C. A healthy debate: exploring the views of medical doctors on the ethics of artificial intelligence[J]. *Artif Intell Med*, 2021, 121:102190.
 - 65 ChatGPT, Egli A., GPT-4, and other large language models: the next revolution for clinical microbiology?[J]. *Clin Infect Dis*, 2023, 77(9):1322–1328.
 - 66 Rai A. Explainable AI: from black box to glass box[J]. *J Acad Mark Sci*, 2020, 48:137–141.
 - 67 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.
 - 68 Pan S R, Luo L H, Wang Y F, et al. Unifying large language models and knowledge graphs: a roadmap[J]. *IEEE Trans Automat Contr*, 2024, 36(7):3580–3599.
 - 69 Xu J, Kim S, Song M, et al. Building a PubMed knowledge graph[J]. *Sci Data*, 2020, 7(1):205.
 - 70 焦逸, 何庆勇, 曾凤, 等. 人工智能实现中医四诊的发展现状、问题及解决路径[J/OL]. *中华中医药学刊*, 1–7 [2025–03–19].
 - Jiao Y, He Q Y, Zeng F, et al. The development status, problems and solutions of artificial intelligence to realize the four diagnoses of traditional Chinese medicine. [J/OL]. *Chinese Archives of Traditional Chinese Medicine*, 1–7 [2025–03–19].
 - 71 Laacke S, Gauckler C. Why personalized large language models fail to do what ethics is all about[J]. *Am J Bioeth*, 2023, 23(10):60–63.
 - 72 Porsdam Mann S, Earp B D, Møller N, et al. AUTOGEN: a personalized large language model for academic enhancement—ethics and proof of principle[J]. *Am J Bioeth*, 2023, 23(10):28–41.
 - 73 Porsdam Mann S, Earp B D, Møller N, et al. AUTOGEN and the ethics of co-creation with personalized LLMs—reply to the commentaries[J]. *Am J Bioeth*, 2024, 24(3):W6–W14.
 - 74 Glenn Cohen I. What should ChatGPT mean for bioethics?[J]. *Am J Bioeth*, 2023, 23(10):8–16.
 - 75 Khan S. Artificial intelligence and machine learning in clinical medicine[J]. *J Acad Mark Sci*, 2023, 388(25):2398.
 - 76 Embi P J. Algorithmic vigilance—advancing methods to analyze and monitor artificial intelligence-driven health care for effectiveness and equity[J]. *JAMA Netw Open*, 2021, 4(4):e214622.
 - 77 Lucy L, Bamman D. Gender and representation bias in GPT-3 generated stories[C]//*Proceedings of the Third Workshop on Narrative Understanding*. Virtual. Stroudsburg, PA, USA: ACL, 2021:48–55.
 - 78 Guo W, Caliskan A. Detecting emergent intersectional biases: contextualized word embeddings contain a distribution of human-like biases[C]//*Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. Virtual Event USA. ACM, 2021:122–133.
 - 79 Abid A, Farooqi M, Zou J. Large language models associate Muslims with violence[J]. *Nat Mach Intell*, 2021, 3(6):461–463.
 - 80 Omiye J A, Lester J C, Spichak S, et al. Large language models propagate race-based medicine[J]. *NPJ Digit Med*, 2023, 6(1):195.
 - 81 Meskó B, Topol E J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare[J]. *NPJ Digit Med*, 2023, 6(1):120.

Large Language Models in Chinese Medicine: Potential Limitations

LI Hua¹, WANG Yuman¹, LIU Xiaoyu¹, SUN Runxue², WANG Shaopo², YANG Qian^{2,3}, DU Yanru^{2,4}

(1. College of Traditional Chinese Medicine, Hebei University of Chinese Medicine, Shijiazhuang 050020, China; 2. Hebei Hospital of Traditional Chinese Medicine, Shijiazhuang 050011, China; 3. Key Laboratory of Hebei Province for Turbid-Toxic Syndrome, Shijiazhuang 050031, China; 4. Hebei Key Laboratory of Gastroenterology of Integrated Traditional Chinese and Western Medicine, Shijiazhuang 050011, China)

Abstract: The emergence of generative pre-training transformer (GPT) and large language models (LLMs) has brought transformative applications to the medical field. In traditional Chinese medicine, LLMs offer unique opportunities to address inefficiencies in clinical workflows and improve the patient experience. However, the prospects have challenges, including data quality, security and privacy, disinformation, ethics, and other issues. This paper systematically elaborates on the application of Large Language Models (LLMs) in Traditional Chinese Medicine (TCM), highlighting their value in improving TCM services, enhancing teaching effectiveness, and optimizing healthcare management processes, while conducting an in-depth analysis of potential technical and ethical challenges during real-world implementation. Preventive measures are necessary to ensure the safe and unbiased use of large language models in TCM clinical practice. We encourage clinicians and researchers to address current challenges and optimize large language models while reducing associated risks. The deployment and implementation of large language models in TCM clinical practice will significantly contribute to the dissemination and development of TCM culture.

Keywords: Large language models, Traditional Chinese medicine, Patient experience, Scheduled workflow, Challenges

(责任编辑: 刘玥辰)