

专家述评

[专家简介] 李涛, 现任解放军总医院医学创新研究部副研究员。主要研究方向为医疗大模型与具身智能医疗机器人。牵头新一代人工智能国家科技重大专项和后勤科研重点项目等多项国家和军队课题。发表论文 25 篇, 授权和受理专利 13 项, 参编专著 5 部。获军队科技进步二等奖一项。获解放军总医院“军事医学科技创新先进个人”和北京市卫健委“首都医学创新学者”。兼任中国生物医学工程学会临床医学工程分会委员和中国研究型医院学会转化医学分会理事。

[专家简介] 张政波, 生物医学工程专业博士, 解放军总医院医学创新研究部某中心主任, 生物医学工程教研室主任, 解放军医学院博士研究生导师, 博士后合作导师, 北京航空航天大学大学生物与医学工程学院兼职导师, 麻省理工学院计算生理学实验室和哈佛医学院 BID 医学中心 Rey 研究所访问学者, 工信部人工智能医疗器械真实世界数据应用中心负责人。在医疗大数据与医学人工智能、医疗物联网与数字健康技术、急危重症预测预警模型与算法等方面开展了大量探索性研究工作。承担国家自然科学基金青年项目与面上项目、国家科技支撑课题、军委科技委前沿创新项目、首长保健专项、北京市科委医工联合类项目、解放军总医院转化医学专项、医疗大数据项目等 10 余项课题, 发表论文 100 余篇, 其中以第一作者和通信作者发表 SCI/EI/Medline 论文 60 多篇 (含 Lancet Digital Health、eClinical Medicine 及计算机领域知名期刊), 申请专利 50 多项, 研究成果获军队科技进步二等奖一项、一等奖一项, 解放军总医院医疗成果一等奖一项。主译麻省理工学院教材《电子病历二次分析利用-医疗大数据分析导论》, 由人民卫生出版社出版。

大语言模型在临床医学领域的应用、挑战和展望

孙磊¹, 汪安安^{1,2,3}, 宋一敏¹, 董敬^{1,2,3}, 刘晓莉¹, 梁洪¹, 李丽轩¹, 宋欣雨¹, 范勇¹, 贾志龙¹, 李涛^{1,2,3}, 张政波^{1,2,3}

¹解放军总医院医学创新研究部, 北京 100853; ²解放军总医院医疗大数据应用技术国家工程研究中心, 北京 100853; ³解放军总医院医学工程实验室, 北京 100853

摘要: 大语言模型正以前所未有的方式与临床医学领域融合并改变临床诊疗的范式, 其不仅在加速知识传播、优化临床决策、提升患者体验等方面发挥着重要作用, 还为医疗系统的整体智能化升级提供了强大动力。本文对大语言模型在临床医学领域的典型应用进行了梳理, 分析了大语言模型存在的问题与面临的挑战, 并展望了大语言模型的发展趋势。通过讨论大语言模型的应用情况和未来前景, 希望能够激励医务工作者和研究人员共同探索这一领域, 把以大语言模型为代表的人工智能新技术应用于临床实践, 提升医疗机构的工作效能和服务质量。

关键词: 大语言模型; 人工智能; 具身智能; 应用

中图分类号: R71

文献标志码: A

文章编号: 2095-5227(XXXX)XX-XXXX-XX

DOI: 10.12435/j.issn.2095-5227.2024.145

引用本文: 孙磊, 汪安安, 宋一敏, 等. 大语言模型在临床医学领域的应用、挑战和展望 [J]. 解放军医学院学报, 2024, X (X): XX-XX.

Applications, Challenges, and Prospects of Large Language Models in the Field of Clinical Medicine

SUN Lei¹, WANG An'an^{1,2,3}, SONG Yimin¹, DONG Jing^{1,2,3}, LIU Xiaoli¹, LIANG Hong¹, LI Lixuan¹, SONG Xinyu¹, FAN Yong¹, JIA Zhilong¹, LI Tao^{1,2,3}, ZHANG Zhengbo^{1,2,3}

¹ Medical Innovation Research Division, Chinese PLA General Hospital, Beijing 100853, China; ² National Engineering Research Center for Medical Big Data Application Technology, Chinese PLA General Hospital, Beijing 100853, China; ³ Medical Engineering

收稿日期: 2024-07-02

基金项目: 新一代人工智能国家科技重大专项 (2021ZD0140409)

作者简介: 孙磊, 女, 博士。研究方向: 医学人工智能。Email: rubysl@126.com 共同第一作者: 汪安安, 女, 学士, 工程师。研究方向: 医学大数据与医学人工智能。Email: wangananjob@163.com

通信作者: 李涛, 男, 硕士。Email: litao301hospital@163.com

张政波, 男, 博士。Email: zhengbozhang@126.com

Laboratory, Chinese PLA General Hospital, Beijing 100853, China

Corresponding authors: LI Tao. Email: litao301hospital@163.com ZHANG Zhengbo. Email: zhengbozhang@126.com

Abstract: Large language models are integrating with the field of clinical medicine in unprecedented ways and changing the paradigm of medical diagnosis and treatment. Large language models play a significant role not only in accelerating the dissemination of knowledge, optimizing clinical decision-making, and enhancing the patient experiences, but also provide a powerful impetus for the overall intelligent upgrade of the healthcare system. In this article, the typical applications of large language models in the field of clinical medicine are reviewed, the existing problems and challenges of large language models are analyzed, and the development trends of large language models in the field of clinical medicine are prospected. By discussing the application and future prospects of large language models, we aim to inspire medical professionals and researchers to explore this field together, apply new artificial intelligence technologies represented by large language models to clinical practice, and thereby improve the work efficiency and service quality of medical institutions.

Keywords: large language models; artificial intelligence; embodied intelligence; application

Cited as: Sun L, Wang AA, Song YM, et al. Applications, Challenges, and Prospects of Large Language Models in the Field of Clinical Medicine [J]. Acad J Chin PLA Med Sch, 2024, X (X): xx-xx.

大语言模型 (large language models, LLMs) 是一种由包含数百亿及以上参数的深度神经网络构建的语言模型。随着人工智能 (artificial intelligence, AI) 技术的飞速发展, LLMs 已经成为当代科技发展的热点和前沿。LLMs 的研究正在经历一个空前的繁荣时期, 将深刻影响社会、经济 and 科技的发展, 为人类带来便利和挑战。在临床医学领域, LLMs 在提高诊疗效率和医疗服务质量、降低医疗成本、改善患者体验等方面具有巨大潜力。

临床医学领域是数据密集型领域, 如何从海量临床诊疗数据中提取有价值的信息并进行高效利用, 一直以来都是亟待解决的重大问题, LLMs 作为先进的自然语言处理技术, 其强大的文本理解和生成能力为解决这一难题提供了新的契机。目前, 部分 LLMs 在某些特定任务中取得了令人瞩目的成果, 已有部分医院开始尝试将 LLMs 应用于临床实践, 但 LLMs 在临床医学领域的应用总体尚处于探索阶段, LLMs 在医疗系统的嵌入程度有限。一方面, 医疗数据的安全性和隐私保护问题限制了 LLMs 在实际应用中的广泛推广; 另一方面, LLMs 的可解释性和可靠性仍有待提高, 需要符合临床医学领域的严格标准。

1 LLMs 发展简史

语言模型的目标是建模自然语言的概率分布, 其发展可以追溯到 20 世纪 50 年代, 当时的计算机科学家就开始探索如何让计算机理解和生成自然语言。在过去的几十年里, 随着计算能力的提高, 语言模型得到了快速发展。2017 年, Vaswani 等^[1]提出了 Transformer 架构, 其核心技术原理是模型通过自注意力机制为序列中每个词语分配不同的注意力权重, 从而有效捕捉词语之

间的长距离依赖关系和上下文信息, 再通过位置编码定位词语在序列中的位置信息, 保证了模型对顺序的敏感性, 从而实现对自然语言的高效处理。模型训练包括预训练和微调两个阶段。预训练阶段: 模型在大规模无标注文本语料库中进行无监督或自监督训练, 目的是让模型捕捉到语言的通用模式和结构, 使其具备通用且强大的自然语言表示能力; 微调阶段: 在小规模有标注数据集中训练模型, 使其更好地适应特定任务或应用场景的需求。在临床医学领域, 通常会使用与临床诊疗相关的数据集对模型进行微调, 学习专业术语和临床医学领域特有的语言模式, 微调使模型能够将预训练阶段学到的通用知识与特定领域的专业知识结合起来, 显著提升模型在特定任务中的表现。预训练和微调的方法极大地提高了模型的泛化能力。2018 年, OpenAI 发布了基于 Transformer 架构的预训练模型 GPT, 并在此后陆续推出了该系列的多个进阶版本^[2-3], 以 GPT 为代表的大规模预训练语言模型的出现, 使自然语言处理进入了预训练微调范式新时代。2022 年, OpenAI 发布了全新对话式大型语言模型 ChatGPT, 其在连续对话能力、生成内容质量、语言理解能力和逻辑推理能力上均有显著提升, 标志着通用大型语言基础模型的重大突破。2023 年 4 月发布的 GPT-4, 具备了多模态理解能力, 在图像处理和语音交互方面表现出色, 不仅增强了 LLMs 处理图文信息的能力, 也为未来 LLMs 的应用开辟了新的可能性。2024 年 5 月, OpenAI 发布了新一代生成式 AI 模型 GPT-4o, GPT-4o 的多模态处理能力实现了显著提升, 标志着 AI 技术在生成式模型领域的又一个重大进步。

2 LLMs 在临床医学领域的应用

当前,我国人口老龄化问题日益严重,慢性病患者数量也在加速增长,这对医疗服务的质量和及时性提出了更高的要求,以 LLMs 为代表的 AI 技术为满足上述要求提供了新的工具和方法。医师在临床工作中需要处理大量的患者信息,包括病程记录、医学影像报告、实验室检查结果等,LLMs 的应用可以高效处理这些文本和数据,提取关键信息,为医师提供辅助诊断和临床决策支持(图 1)。基于 LLMs 的智能问诊、辅助诊断、个性化治疗、智慧护理以及电子病历自动生成等应用,能够显著提升医疗机构医疗服务的效率和准确性^[4-5]。加速推进 LLMs 在临床医学领域的应用不仅可以带来技术上的革新,更是对传统医疗范式的深刻改变,对现有医疗体系的深度重构,预示着一个更加智能、高效、人性化的医疗新时代的到来。

2.1 智能问诊

LLMs 凭借其深厚的语言理解和生成能力,不仅能够模拟医师与患者的自然对话,进行症状询问和健康咨询,还能基于庞大的临床医学知识库和真实世界数据,提供初步的诊断建议和健康方案,可应用于在线问诊平台、医院导诊系统和医疗健康咨询等。基于 LLMs 的智能问诊系统,能够 24 h 不间断服务,不仅可以为患者提供即时的医疗援助,减轻初级医师的工作压力,还能有效缓解医疗资源分布不均的现状,为医疗健康服务的可及性和效率带来革命性的提升。Pushpanathan 等^[6]评估了 3 种基于 LLMs 的聊天机器人在回答眼科症状相关常见问题时的准确性和全面性,结果显示 ChatGPT-4 在回答眼部症状查询时准确性最高,显著优于 ChatGPT-3.5 和 Google Bard。Hershenhouse 等^[7]评估了 ChatGPT-3.5 回答前列腺癌诊断、治疗和术后随访相关问题的能力,结果显示 ChatGPT-3.5 生成的回答在准确性、完整性和清晰度上得到了泌尿科医师的高度评价。Benirschke 等^[8]研究发现,ChatGPT-4 在帮助

病理学专业人士回答一般病理学问题时,准确性和完整性较高,具有显著节省时间的潜力。Li 等^[9]开发了一种名为 ChatDoctor 的医疗聊天机器人模型,通过使模型具备从线上和离线医疗领域知识库中检索信息的能力,观察到了其对患者响应的准确性有了显著提升。Lahat 等^[10]评估了 OpenAI 的 ChatGPT 聊天机器人在回答各种胃肠病学相关问题时的性能,结果发现 ChatGPT 在回答患者关于胃肠道健康的问题方面具有一定的潜力,但其准确性还有待提高,需要进一步改进和完善。Makrygiannakis 等^[11]比较了 4 种 LLMs 对正畸学领域临床相关问题的回答能力,结果表明 LLMs 具有支持循证正畸学的潜力,但如果在没有谨慎考虑的情况下使用,可能会导致不准确的医疗决策,为了保证将 LLMs 成功融入实践,必须进行更多临床验证并改进模型。这些研究提供了对临床医学领域聊天机器人性能的深入分析,揭示了 LLMs 在临床医学问答领域的巨大应用潜力。然而,这些模型在不同细分领域的表现存在较大差异,因为不同疾病在诊疗过程中所需的诊疗数据类型和诊疗思维链等差异巨大,所以 LLMs 的应用仍需针对特定领域进行优化和完善,以确保医疗信息的准确性和可靠性。

尽管 LLMs 在知识处理方面表现出色,但问诊往往需要结合细致的临床观察,模型的问诊能力与医师的临床经验相比仍有差距,如何提升模型问诊的准确性,是智能问诊系统在临床医学领域应用的关键。此外,智能问诊不应被视为医师的替代品,而应定位为辅助工具,如何促进人机协作,是未来临床实践中的重要课题。

2.2 辅助诊断

辅助诊断是指利用各种技术手段和工具来帮助医师进行疾病识别、分析和决策的过程。LLMs 的核心在于其强大的自然语言理解能力和知识推理能力,通过深度学习海量医学文献、临床指南和病例记录,LLMs 可以具备深度理解医学专业词汇和上下文的能力,能够进一步分析患者的病历信息、临床症状、实验室检查结果和医学影像报告,为医师提供全面而深入的诊断建议。在某些情况下,LLMs 还能够通过分析患者的诊疗史和实时症状,预测疾病的发展趋势和可能的并发症。Obradovich 等^[12]综述了 LLMs 在精神病学中的潜在应用,文中指出 LLMs 可以快速分析患者数据,包括病历、治疗记录和诊断报告等,从而评估精神疾病的严重程度,还可以根据患者的症



图 1 LLMs 在临床医学领域的应用

Fig.1 Applications of LLMs in the Field of Clinical Medicine

状和表现,提供可能的诊断建议。Wu 等^[13]探讨了如何使用 LLMs 增强甲状腺结节超声影像诊断的准确性和一致性,发现 LLMs 显示出了改善甲状腺结节诊断的潜力,与 Bard 和 ChatGPT-3.5 相比,ChatGPT-4 具有更好的一致性和准确性。Krusche 等^[14]评估了 ChatGPT-4 在风湿病诊断中的准确性,并与风湿病专家的诊断进行了比较,结果显示 ChatGPT-4 在风湿病诊断中展现出潜力,可作为辅助工具加速诊断过程。Sandmann 等^[15]对 LLMs 在临床决策支持任务中的表现进行了系统分析,结果表明 GPT-4 在诊断方面优于 GPT-3.5。Savage 等^[16]研究了 LLMs 在医学诊断推理方面的可解释性,GPT-4 在诊断推理提示下能够模仿医师的常见推理过程,这说明 LLMs 有潜力通过提供可解释的推理过程来帮助医师评估其诊断结果的可信度,从而克服“黑箱”局限性。以上研究表明,LLMs 能够高效分析患者数据,提供诊断建议,并有望提高诊断的准确性和一致性,特别是 ChatGPT-4 模型,其在多个研究中表现出优于先前版本的性能,而 GPT-4 模型在可解释性方面取得的进展,有助于克服“黑箱”问题。这些发现预示着 LLMs 未来可能成为临床医学领域的重要辅助工具,但在实际应用中,模型在实时获取患者的多维度诊疗信息方面仍存在困难,其稳定性和安全性需要经过严格验证与持续优化,以确保医疗服务的精准性和安全性。

将 LLMs 应用于辅助诊断,可以提高临床诊断的准确性和效率,具有重要的意义和广阔的应用前景。未来,随着技术的不断进步和诊疗数据的不断积累,LLMs 有望成为医师的重要助手,为患者提供更加优质和高效的医疗服务。

2.3 个性化治疗

在传统的医疗模式中,医师通常会根据疾病的一般特征为患者制定治疗方案,这一治疗策略往往忽视了个体间显著的生理差异与疾病表现多样性。随着精准医学的快速发展,临床医师根据患者的个体差异选择最合适的治疗方案成为可能^[17],个性化治疗方案的制定正逐步成为提升治疗效果和患者体验的关键。LLMs 能够快速、准确地分析每例患者的病情特点,整合患者遗传背景、生活习惯、过往病史等因素,辅助医师为患者制定更加个性化和精准化的治疗方案,实现治疗效果的最大化和不良反应的最小化。此外,LLMs 还可以与基因组学和药物代谢动力学等学科结合,进一步优化个性化治疗方案,确保治疗的安全性

和有效性。在肿瘤治疗中,LLMs 可以分析基因组数据,预测患者对特定化疗药物的响应性,指导个性化用药决策;在精神疾病和慢性疾病管理领域,LLMs 可以根据患者的情绪状态、疾病进展等个性化信息,推荐最适合的心理干预或生活方式调整方案。Perlis 等^[18]研究了使用 GPT-4 turbo 辅助双相情感障碍抑郁发作的治疗方案选择,结果显示,经过增强的 GPT-4 turbo 模型识别最佳治疗方案的准确率为 50.8%,优于基础模型和社区医师,但该增强模型偶尔会选择专家认为较差或禁忌的治疗方案。Yang 等^[19]提出了一种基于 BERT 和 TabNet 的多模态多任务学习模型,用于评估精神疾病患者的康复状况,医师不仅可以依据模型分析结果制定个性化的康复计划,还可以根据患者的实时数据调整治疗方案,从而改善患者的康复结果和生活质量。这两项研究共同展示了 LLMs 在辅助治疗方案选择和个性化康复计划制定方面的应用潜力。GPT-4 turbo 的应用虽然在某些情况下存在局限,但其在大多数情况下能够辅助专家做出合理的治疗选择。基于 BERT 和 TabNet 的多模态多任务学习模型则为精神疾病患者的个性化治疗和康复提供了新的思路和方法。这些研究为未来的临床实践提供了有价值的参考。

在肿瘤学、罕见病治疗以及慢性疾病管理等领域,个性化治疗尤为重要,LLMs 的应用可以显著提高治疗效果和患者生活质量。LLMs 在个性化治疗方案制定方面的应用,标志着医疗保健向更加精准、高效的方向迈进了一大步。随着个性化治疗需求的日益增加,LLMs 有望在未来发挥更加关键的作用,推动精准医疗的发展。

2.4 智慧护理

在智慧护理方面,LLMs 的应用场景日益丰富,从智能健康顾问到个性化康复指导,从精神支持到家庭护理教育,LLMs 正在帮助医护人员更好地理解患者的需求,提供更加贴心和高效的护理服务。借助先进的自然语言处理技术,LLMs 能够模拟专业的医疗对话,为患者提供实时的健康咨询和病情监测;在康复过程中,LLMs 能够基于患者的具体情况,生成个性化饮食建议、运动指导、康复计划、疾病预防知识等,促进患者自我管理,提高生活质量;面对疾病,患者往往承受着巨大的心理压力,LLMs 能够理解并回应患者的情感需求,提供心理安慰和辅导,尤其是在慢性病管理和康复过程中,给予患者必要的精神支持,减轻焦虑和抑郁情绪;除此之外,LLMs 还能

提供详尽的护理知识和技能指导,从而提升家庭成员的护理能力,确保患者在家也能得到专业级的照料。Hobensack 等^[20]综述了 LLMs 在护理实践中的应用,文章指出 LLMs 可以协助护士完成护理工作,也可以帮助护士进行患者教育。Nashwan 和 Abujaber^[21]探讨了 LLMs 在护理计划中的应用潜力,LLMs 可用于分析患者数据、制定护理计划、促进与患者沟通以及提升护理人员专业技能。Sheng 等^[22]综述了 LLMs 在糖尿病护理中的应用潜力和前景,LLMs 可以提供个性化的健康指导、营养建议、血糖监测辅助、药物依从性支持、并发症监测和心理健康支持,为患者和医疗保健专业人员提供全面的数据支持和糖尿病护理管理。护理工作是连接医师治疗方案与患者实际需求之间的桥梁,贯穿于整个医疗流程,是医疗服务中不可或缺的一环,上述文献综述表明,LLMs 在护理领域的应用正逐步拓宽,从协助日常工作、患者教育到护理计划的制定,LLMs 展现出其多功能性和实用性。特别是在糖尿病护理中,LLMs 的潜力尤为突出,能够提供个性化支持和全面的数据管理,有望提升护理质量和患者体验,预示着未来护理工作将更加智能化和人性化。

LLMs 在智慧护理领域的应用,不仅体现了技术进步带来的医疗变革,也彰显了以人为本的医疗理念,将显著提升医疗护理服务效率和个性化程度,从而为患者提供更加优质的护理服务。

2.5 病历自动生成

LLMs 在医疗文本信息的生成、数据挖掘与分析方面展现出了前所未有的能力。LLMs 利用其深度学习能力,可以高效地处理和理解临床医学领域的非结构化文本数据。在信息生成方面,LLMs 能够根据给定的模板,自动撰写病历、临床报告、患者教育材料和临床指南;在数据挖掘方面,LLMs 能够从电子健康记录、医学研究论文等资料中提取关键信息,帮助医疗专业人员快速获取所需数据;在数据分析方面,LLMs 能够通过复杂算法和大规模数据训练,揭示疾病发展的趋势和治疗措施的效果,为临床决策提供支持。这些应用不仅提高了医疗文档处理的效率,还增强了数据分析的深度和广度。Van Veen 等^[23]评估了 8 种不同的 LLMs 在 4 个不同的临床文本摘要任务(包括放射报告、患者问题、病历记录和医师-患者对话)中的表现,结果表明,将 LLMs 纳入临床工作流程中,可以有效减轻医师的工作负担。Saj-

jadul Islam 等^[24]开发了多个自动生成患者主诉的电子健康记录工具,结果显示 BioGPT-Large 模型在自动生成患者主诉方面表现优异。Huang 等^[25]研究了 ChatGPT 从临床笔记中提取结构化数据的能力,结果表明 ChatGPT 在提取肺癌病理报告中关于 TNM 分期和病理组织学类型的信息方面表现出良好的性能,准确率达 89%。Liu 等^[26]评估了使用 GPT-4 总结临床决策支持系统警报评论的能力,发现 AI 生成的摘要具有良好的清晰度、准确性和完整性,并且在完整性方面显著优于人工生成的摘要。Gabriel 等^[27]综述了 LLMs 在围术期医学中的应用,LLMs 可应用于临床文档的处理和分析,如对文档进行分类、总结、提取关键信息,并识别特定患者群体,从而帮助医师更快速、准确地了解患者病情,制定个性化的诊疗方案,提高工作效率。综合上述研究,LLMs 在文本处理和数据分析方面展现出了广泛的应用和变革性的潜力,无论是在临床文本摘要、电子健康记录生成方面,还是在结构化数据提取方面,LLMs 都表现出了高效处理医疗信息的能力,这不仅有助于减轻医护人员的工作负担,还能提升诊疗的准确性和效率,对于优化临床工作流程和提升患者护理质量具有重要意义。

在临床医学实践中,门诊病历和病房病历是临床决策和信息传递的重要依据,其质量对于临床诊疗的全流程管理具有重要意义,但病历书写和录入工作也给医护人员带来了巨大的工作负担。有文献报道,医师与患者互动 1 h,可能需要花费 2 h 整理文档记录,而护士处理文档任务可能会占据高达 60% 的时间,并造成相当大的工作压力,这些任务分散了医护人员对患者诊疗和护理的注意力,并造成医护人员的工作满意度下降^[23]。LLMs 的应用,不仅促进了医疗信息价值的释放,提高了利用效率,还极大地减轻了医护人员的工作负担,降低了医疗机构的运营成本,促进了医疗信息化和智能化的发展。

2.6 基于多模态大模型的机器人辅助诊疗

解放军总医院医学创新研究部研究团队致力于具身智能机器人在临床诊疗中的应用研究,与达闼机器人有限公司合作开发的机器人模型 RobotGPT 先后支撑了解放军总医院临床科室的机器人独立查房、健康宣教、康复训练、药品递送和清洁消毒等示范应用,并开发了用于脊柱关节炎辅助诊断的 LLMs^[28]。辅助诊疗机器人可以快速、准确地与患者进行交流,收集患者的问诊信息,

辅助医师进行问诊和查房,不但能提高医师问诊和查房的效率,减轻工作负担,使医师有更多时间关注复杂病例,还能改善患者的就医体验,提升患者满意度。同时,团队在病房开展面向住院患者的机器人健康宣教临床验证工作,机器人可以根据不同患者的情况进行多次健康宣教,协助医护人员进行患者健康管理,有助于患者提高健康意识,也能够促进疾病的预防、治疗和康复(图2)。未来随着医疗场景世界模型和双足人形机器人技术的不断进步,模型驱动下的辅助诊疗机器人将全方位嵌入医疗机构的各个诊疗环节。

除以上应用外,LLMs在医学研究与药物研发、临床试验、智能监护与预警、康复训练、医疗行政管理、远程医疗、医学伦理学、医学教育、学术写作等方面均有着广泛的应用^[29-36]。

3 存在的问题与挑战

LLMs为临床医学实践带来了诸多创新和便利,但同时也伴随着一系列的问题和挑战。研究者们应审慎地考虑和应对可能出现的问题,从数据隐私保护到模型的准确性和可靠性,从法律合规性到伦理责任界定,每一个环节都需要细致入微的考量,以促进对LLMs在临床医学领域应用

的深入理解和负责任的实践(图3)。

3.1 数据安全

临床诊疗信息的隐私保护是不容忽视的问题^[4]。LLMs的训练过程中,如果使用了可识别的患者数据,且缺乏相应的安全措施,那么模型就有可能无意中记住并泄露这些敏感信息,从而带来潜在风险;即使患者数据经过匿名化处理,也可能遭受对抗性攻击导致泄露,而且存在重新识别的风险^[37-39]。如何在利用LLMs高效处理信息的同时,确保患者数据的安全与隐私,是技术开发者 and 临床医学领域从业者共同面临的考验,需要采取数据脱敏和匿名化处理、使用联邦学习技术、开发评估工具、模型安全审查、提升网络安全等措施来保护患者的隐私和数据的安全,以促进LLMs的持续发展^[37-39]。

LLMs的应用依赖于丰富、高质量的数据,数据利用对于科学研究和社会福祉的重要性不言而喻,过分保护数据可能会限制数据的可访问性,从而影响模型的训练和优化,进而阻碍临床医学领域的研究和创新。通过建立安全受控的数据共享平台、制定和实施严格的数据治理框架、确保合规性和伦理审查、鼓励患者参与和知情同意等措施,研究者可以在尊重个人隐私的同时充分利

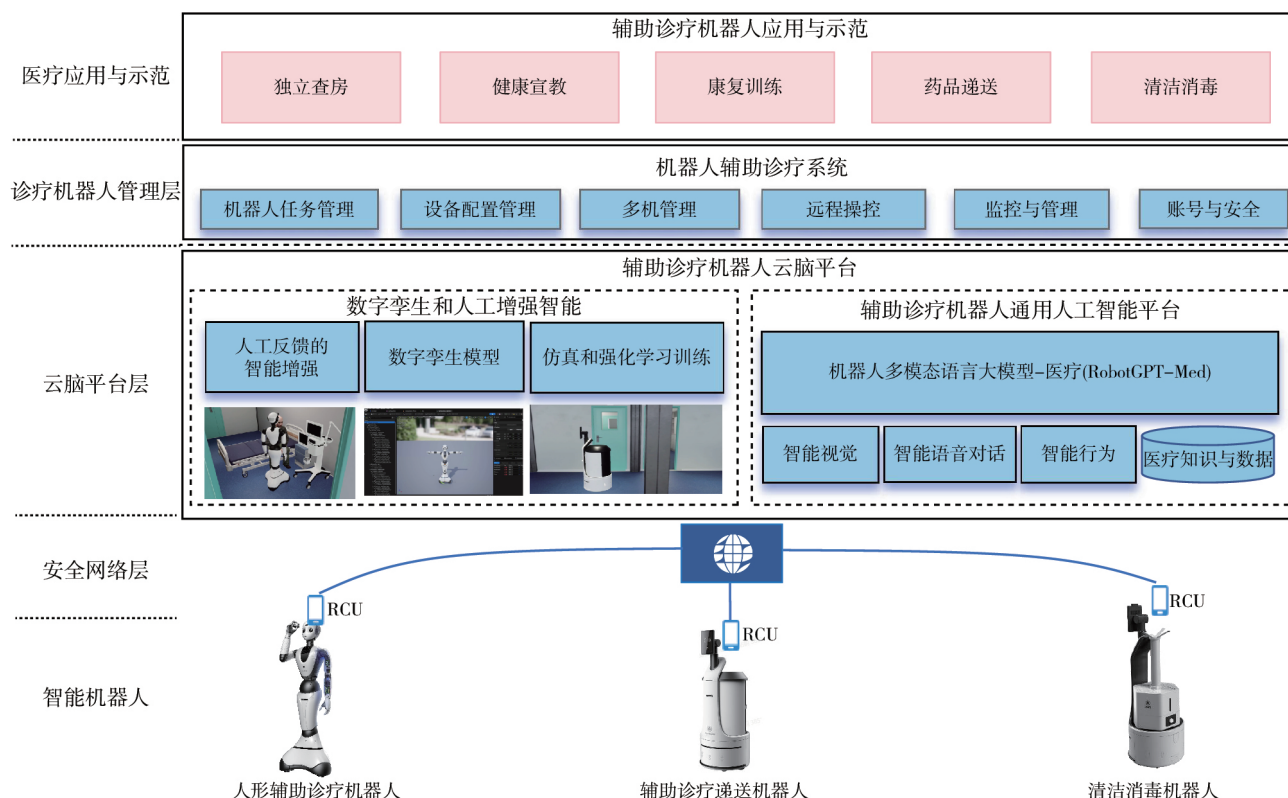


图2 辅助诊疗机器人系统

Fig.2 Auxiliary Diagnosis and Treatment Robot System



图 3 LLMs 在临床医学领域应用存在的问题与挑战

Fig.3 Problems and Challenges of LLMs in the Field of Clinical Medicine

用数据,从而促进医学研究的进步。

医院信息系统 (hospital information system, HIS)、实验室信息系统 (laboratory information system, LIS) 和医学影像存档与通讯系统 (picture archiving and communication system, PACS) 是医院非常重要的信息系统,这些医疗业务系统分别负责医院的医疗业务运行、检查检验和医学影像管理。通过将 LLMs 嵌入到 HIS、LIS 和 PACS 系统中,虽然理论上可以实时提取相关信息和数据,从而更为迅捷地辅助医护人员开展患者诊疗,医院可能会实现更加高效和精准的医疗服务,但对于管理这些医院信息系统的部门而言,因为缺乏测试验证,LLMs 与这些业务系统的对接,可能造成医院业务系统运行的不稳定。一旦造成医院业务系统的混乱,将给医院的日常临床诊疗运行带来极大风险。所以在实际工作中,还没有 LLMs 被嵌入到 HIS、LIS 和 PACS 系统中。医院的信息管理部门最重视的是医院信息系统的安全运行,管理者一方面担心嵌入的 LLMs 会影响各业务系统的正常运行,另一方面也忧虑其会对业务系统内的敏感数据造成影响,因此对 LLMs 实时读取医疗业务系统数据持保留态度。

面对上述挑战,医院管理者需采取前瞻性策略,确保技术进步与医疗实践的和谐共生,但同时也要谨慎行事,在做出决策之前,应全面深入地评估潜在利益和风险,制定相应的风险管理策略,并确保所有的操作都符合当前的医疗标准和法规要求。通过这种方式,医院可以在保护患者数据的同时,利用 AI 提升医疗服务的质量。

3.2 技术问题

尽管 LLMs 在诸多任务中表现出色,但其自身仍存在一系列技术问题。LLMs 的幻觉是指 LLMs 在生成文本时,可能会产生与事实不符、缺乏逻辑的内容,这种现象可能是由于模型在训练过程中接触到了大量的虚假信息、错误数据或在训练过程中过度拟合了数据,导致其在生成文本时出现了偏差^[40-42]。LLMs 的鲁棒性是指模型在面对输入数据中的噪声、异常值或对抗性样本时,能够保持性能稳定,输出可靠结果的能力。鲁棒性越强的模型,实际应用中就越能提供可靠的性能,从而更好地服务于用户需求。泛化性问题是模型在处理未见过的数据时可能出现的性能下降情况,尽管 LLMs 在训练数据上表现良好,但其可能无法很好地泛化到新的或不同的数据集。这种局限性可能会限制模型在多样化和不断变化的现实世界场景中的应用。LLMs 自身存在的技术问题对模型的可靠性和应用范围造成了一定的限制。研究者和开发者正不断探索解决方案,如采用更先进的训练方法提高模型的鲁棒性和泛化能力,开发更精细的评估指标来检测和减少幻觉,通过这些努力,我们可以期待 LLMs 在未来能够以更加成熟和可靠的方式服务于社会。

除上述技术问题外,LLMs 还面临着来自模型架构、多模态大模型预训练、模型参数微调等方面的挑战。LLMs 通常具有数亿甚至千亿级别的参数,这使得模型在训练和部署过程中对计算资源的需求极高,包括 GPU 资源和内存,如何在保证模型性能的同时降低参数规模,成为一大挑战。除此之外,如何平衡模型复杂度与效率,如何提高模型的可解释性,都是 LLMs 架构方面存在的挑战^[43]。在模型预训练方面,随着多模态大模型的广泛开发,对模型预训练提出了新的挑战,获取高质量、多样化的多模态数据,设计合理的数据融合和模态对齐策略^[44],提高多模态模型在多种下游任务中的表现等,都是亟待解决的问题。模型参数微调也是 LLMs 训练的关键环节,实现高效的知识迁移,解决预训练数据与微调数据可能存在的分布不一致的问题,平衡模型的拟合能力与泛化能力,避免灾难性遗忘,是模型参数微调中的重要挑战^[45]。为了解决这些问题,未来的研究需要继续探索更高效的模型架构、创新的预训练方法以及更有效的微调策略,以促进 LLMs 在临床医学领域的广泛应用。

3.3 伦理与道德

LLMs 在为人类带来便利和效益的同时,也带来了一系列伦理问题,如 LLMs 在训练过程中,依赖于大量的历史文本数据,这其中包括了可能存在的文化、性别、地域等偏见和歧视性内容,因此在生成文本、回答问题时可能会带有偏见和主观色彩,研究者和开发者需持续监控模型的输出,通过优化数据处理的调整模型算法等措施,努力消除偏见和不公正的倾向^[39]。另一个重要的伦理问题是,当 LLMs 参与到医疗决策中时是否应该承担相应的责任?如决策输出责任、产生错误信息的责任、隐私泄露的责任等,如果要承担这些责任,具体的标准和方法是什么?这涉及复杂的法律和伦理考量,加强监管、明确责任主体并建立清晰的问责机制,是确保医疗安全和患者权益的关键^[46]。

在道德层面,LLMs 存在被滥用的风险,其可能被用于生成虚假医疗信息或误导性建议,并导致患者采取错误的诊疗措施,从而影响患者的健康。为了防止虚假医疗信息的传播,开发者和使用者需要加强对 LLMs 输出内容的审核,确保其真实性和可靠性,同时采取模型蒸馏、对抗训练和检索增强生成等措施来降低 LLMs 的滥用风险^[47]。面对 LLMs 使用过程中的伦理和道德问题,我们需要制定全面、完善的规范和标准,加强对 LLMs 的监管,指导 LLMs 的研究和应用,确保其健康、可持续的发展。

4 展望

在 AI 浪潮中,LLMs 正逐步成为临床医学领域创新的关键驱动力,尽管目前 LLMs 在临床医学领域的应用仍处于起步阶段,但其未来的发展潜力不可限量。展望 LLMs 的技术发展,其将从单模态迭代到多模态,应用形式将从 AI Agent 实体化发展到具身智能,这将极大地拓展 LLMs 的应用前景。

4.1 多模态大模型

未来,LLMs 在临床医学领域的发展将聚焦于多模态大模型。多模态大模型能够同时理解和处理多种不同形式的输入数据,如文本、图像、声音等,通过整合来自不同模态的信息,提供更丰富和准确的分析结果,从而实现更高效的应用。与单一模态的模型相比,多模态大模型具有多方面的优势:(1)多模态大模型能够将不同类型的数据进行有效融合,挖掘出更深层次的信息,为医疗决策提供更全面的依据;(2)通过交叉验证不同

模态的信息,可以提高诊断和治疗方案的准确性,减少漏诊和误诊的可能;(3)多模态大模型能够推荐更加个性化的治疗方案,为精准医疗提供了新的可能性^[48];(4)多模态大模型能够自动处理和整合大量来自不同模态的数据,进一步提高医疗工作的效率。2024 年,Google 推出了 Med-Gemini 多模态模型,它基于 Gemini 模型强大的多模态和长上下文推理能力,能够高效地处理和整合来自文本、图像、视频等多种模态的数据,在多模态医学对话、医学研究和教育等领域展现出卓越的性能和潜力^[49]。Med-Gemini 模型在医学文本摘要和转诊信生成等任务中的表现超越了人类专家^[49],未来还会有更多医疗领域多模态大模型的出现,进一步推动医疗行业向更智能、更高效的方向发展,全面优化医疗服务流程。

LLMs 在临床医学领域的发展方向将是多元化和深层次的。技术创新将赋予 LLMs 更强的数据处理能力和更广的应用范围;应用拓展将使 LLMs 更深入地融入临床诊疗流程,提高医疗系统的响应速度和个性化服务水平。随着研究的深入和技术的成熟,我们期待 LLMs 能够为临床医学领域带来更多的创新和突破,为人类健康事业贡献力量。

4.2 AI Agent

AI 智能体 (AI Agent) 作为能够感知环境、独立决策并执行任务的智能体,其核心在于融合了先进的机器学习与 AI 技术,赋予了系统高度的自主性和适应性。从企业级应用到个人生活助手, AI Agent 正逐渐渗透到社会的每一个角落。

在临床医学领域, AI Agent 的引入为全新的工作模式提供了可能。我们可以设想如下工作场景: Doctor Agent 作为医疗决策支持的核心,集成了丰富的医学知识库和临床经验,能够辅助医师进行病情分析、诊断建议、治疗方案设计等工作。它不仅能快速查阅全球最新的医学研究成果,还能基于个体患者的病史和实时健康数据,提供个性化的治疗建议,极大地提高了诊断的准确性和效率,减轻了医师的工作负担。相比之下, Nurse Agent 则更多聚焦于患者的日常护理和健康监护,扮演着智能护理助手的角色。它通过持续监测患者的生理指标、情绪状态和生活习惯,及时提醒患者服药、安排复诊、调整生活习惯等,有效补充了传统护理工作的不足。在这一智能协作体系中, Doctor Agent 与 Nurse Agent 的交互尤为关键,两者通过共享患者信息和治疗进

度,实现了无缝对接,确保了护理和治疗方案的一致性和连贯性。Doctor Agent 在制定治疗策略时,可参考 Nurse Agent 提供的患者日常反馈和身体状况变化,使方案更加贴合患者的实际需求;而 Nurse Agent 在执行护理任务时,可即时响应 Doctor Agent 的指令调整,确保执行效率和安全性。这种协同作业模式,不仅优化了资源分配,更是在人机协作的探索中,为患者带来了更加全面、细致且个性化的医疗照护体验。

4.3 具身智能机器人

LLMs 具有理解和生成自然语言的能力,这使得它们在处理复杂任务和提供个性化服务方面具有巨大潜力。然而,这些模型往往缺乏与外部世界的直接互动能力,因此需要具身智能机器人作为其物理实体。具身智能机器人作为 LLMs 的载体,可以将模型的智能决策转化为实际行动,通过与环境进行交互来执行各种任务,并将相关信息反馈给 LLMs,从而实现模型的优化和升级。

“具身”的本质是赋予一个物理实体(通常是机器人)在现实世界中执行任务并具有与环境进行互动的能力,该物理实体需要具备多模态感知、智能决策和智能控制等能力,而“智能”的本质是对世界知识的抽象,智能的加入改变了机器人传统的控制方式,即从传统的基于固定算法或模型的控制转向基于数据驱动学习和自适应算法的控制。未来,LLMs 和 AI Agent 驱动的具身智能机器人有可能嵌入医院业务系统,可以实时获取患者的病历信息、检查结果和治疗记录,为 LLMs 提供更加全面的数据支持。同时,LLMs 可以基于患者医疗数据生成诊断建议、治疗方案和护理计划,再通过具身智能机器人将这些决策转化为实际行动。这些具身智能机器人将在临床医学领域发挥越来越重要的作用,它们将与 HIS、LIS、PACS 等系统集成,共同构成医院的核心业务系统,让智慧医疗的愿景逐步成为现实。

从 LLMs - AI Agent - 具身智能这一 AI 技术演进路径来看,具身智能被视为 AI 发展的重要里程碑,它融合了感知、行动与认知能力,使具身智能机器人有望成为通用机器人领域的关键组成部分,为 AI 的自主进化提供强有力的支撑,推动 AI 向更高级的形态跃迁。辅助诊疗机器人作为一种典型的具身智能机器人,基于本体的视觉、语音、行走和抓取等能力,将在部署于云端或边端的机器人云脑平台和多模态医疗大模型的共同驱动下,辅助医护人员开展临床诊疗工作。展望未

来,辅助诊疗机器人在临床医学领域的应用将进一步深化,成为推动医疗技术进步和服务模式创新的重要力量。

5 结论

综上所述,LLMs 在临床医学领域展现出巨大的潜力和价值,为医疗服务的现代化转型注入了强劲动力。通过智能化的信息处理与分析,LLMs 能够显著优化服务流程,实现医疗资源的高效配置与利用。LLMs 还能够提供精准的辅助诊断、个性化的健康咨询以及持续的病情监测,确保医疗服务的专业性和个性化,满足不同患者的特定需求。更为重要的是,通过智能交互,LLMs 可以增强医患沟通,提升患者满意度和信任度,对提升患者体验产生深远影响。此外,LLMs 能够自动化处理大量重复性工作,使医护人员能够更加专注于临床诊疗和患者关怀,从而全面提升医疗服务的整体效能。

随着技术的不断发展,LLMs、AI Agent 和具身智能机器人等技术的融合将成为 AI 技术在临床医学领域垂直应用的重要趋势。LLMs 与医疗需求的深度融合会深刻改变医疗机构开展临床诊疗的范式,这将是一个不断迭代,持续优化与进化的过程。为实现 LLMs 在临床医学领域的健康可持续发展,提升我国在全球 AI 领域的核心竞争力,需要国家机关、医疗机构、科研机构和企业共同努力,建立完善的标准和监管体系,加强伦理审查,确保技术的公正性和透明度,同时不断优化算法,提高模型的实用性和可靠性^[50-51]。未来,我们期待看到更多创新应用,使 LLMs 不仅成为提升医疗服务效率的工具,还能搭建起增强医患沟通、推动医学研究进步的桥梁,携手共创一个更加智能、精准且富含人文关怀的医疗未来。

作者贡献 孙磊:文章的撰写和修改;汪安安、董敬:文献查询,文章修改;宋一敏:配图制作;刘晓莉、梁洪、李丽轩、宋欣雨、范勇、贾志龙:文献的查询和整理;李涛:论文的设计、指导和修改;张政波:论文指导。

利益冲突 所有作者声明无利益冲突。

参考文献

- 1 Vaswani A, Shazeer NM, Parmar N, et al. Attention is all you need [J/OL]. <https://doi.org/10.48550/arXiv.1706.03762>.
- 2 Brown TB, Mann B, Ryder N, et al. Language Models are Few-Shot Learners [J/OL]. <https://doi.org/10.48550/arXiv.2005.14165>.
- 3 Ouyang L, Wu J, Jiang X, et al. Training language models to

- follow instructions with human feedback [J/OL]. <https://doi.org/10.48550/arXiv.2203.02155>.
- 4 Clusmann J, Kolbinger FR, Muti HS, et al. The future landscape of large language models in medicine [J]. *Commun Med*, 2023, 3 (1): 141.
 - 5 Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al. Large language models in medicine [J]. *Nat Med*, 2023, 29 (8): 1930-1940.
 - 6 Pushpanathan K, Lim ZW, Er Yew SM, et al. Popular large language model chatbots' accuracy, comprehensiveness, and self-awareness in answering ocular symptom queries [J]. *iScience*, 2023, 26 (11): 108163.
 - 7 Hershenhouse JS, Mokhtar D, Eppler MB, et al. Accuracy, readability, and understandability of large language models for prostate cancer information to the public [J/OL]. <https://doi.org/10.1038/s41391-024-00826-y>.
 - 8 Benirschke RC, Wodskow J, Prasai K, et al. Assessment of a large language model's utility in helping pathology professionals answer general knowledge pathology questions [J]. *Am J Clin Pathol*, 2024, 161 (1): 42-48.
 - 9 Li YX, Li ZH, Zhang K, et al. ChatDoctor: a medical chat model fine-tuned on a large language model meta-AI (LLaMA) using medical domain knowledge [J]. *Cureus*, 2023, 15 (6): e40895.
 - 10 Lahat A, Shachar E, Avidan B, et al. Evaluating the utility of a large language model in answering common patients' gastrointestinal health-related questions: are we there yet? [J]. *Diagnostics*, 2023, 13 (11): 1950.
 - 11 Makrygiannakis MA, Giannakopoulos K, Kaklamanos EG. Evidence-based potential of generative artificial intelligence large language models in orthodontics: a comparative study of ChatGPT, google bard, and microsoft Bing [J/OL]. <https://doi.org/10.1093/ejo/cjae017>.
 - 12 Obradovich N, Khalsa SS, Khan WU, et al. Opportunities and risks of large language models in psychiatry [J]. *NPP*, 2024, 2: 8.
 - 13 Wu SH, Tong WJ, Li MD, et al. Collaborative enhancement of consistency and accuracy in US diagnosis of thyroid nodules using large language models [J]. *Radiology*, 2024, 310 (3): e232255.
 - 14 Krusche M, Callhoff J, Knitz J, et al. Diagnostic accuracy of a large language model in rheumatology: comparison of physician and ChatGPT-4 [J]. *Rheumatol Int*, 2024, 44 (2): 303-306.
 - 15 Sandmann S, Riepenhausen S, Plagwitz L, et al. Systematic analysis of ChatGPT, Google search and Llama 2 for clinical decision support tasks [J]. *Nat Commun*, 2024, 15 (1): 2050.
 - 16 Savage T, Nayak A, Gallo R, et al. Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine [J]. *NPJ Digit Med*, 2024, 7 (1): 20.
 - 17 Stumpp NE, Sauer-Zavala S. Evidence-based strategies for treatment personalization: a review [J]. *Cogn Behav Pract*, 2022, 29 (4): 902-913.
 - 18 Perlis RH, Goldberg JF, Ostacher MJ, et al. Clinical decision support for bipolar depression using large language models [J]. *Neuropsychopharmacology*, 2024, 49 (9): 1412-1416.
 - 19 Yang HY, Zhu D, He SY, et al. Enhancing psychiatric rehabilitation outcomes through a multimodal multitask learning model based on BERT and TabNet: an approach for personalized treatment and improved decision-making [J]. *Psychiatry Res*, 2024, 336: 115896.
 - 20 Hobensack M, Von Gerich H, Vyas P, et al. A rapid review on current and potential uses of large language models in nursing [J]. *Int J Nurs Stud*, 2024, 154: 104753.
 - 21 Nashwan AJ, Abujaber AA. Harnessing large language models in nursing care planning: opportunities, challenges, and ethical considerations [J]. *Cureus*, 2023, 15 (6): e40542.
 - 22 Sheng B, Guan ZY, Lim LL, et al. Large language models for diabetes care: potentials and prospects [J]. *Sci Bull*, 2024, 69 (5): 583-588.
 - 23 Van Veen D, Van Uden C, Blankemeier L, et al. Adapted large language models can outperform medical experts in clinical text summarization [J]. *Nat Med*, 2024, 30 (4): 1134-1142.
 - 24 Sajjadul Islam KM, Nipu AS, Madiraju P, et al. Autocompletion of chief complaints in the electronic health records using large language models [C] //2023 IEEE International Conference on Big Data (BigData). IEEE, 2023: 4912-4921.
 - 25 Huang JW, Yang DM, Rong RC, et al. A critical assessment of using ChatGPT for extracting structured data from clinical notes [J]. *NPJ Digit Med*, 2024, 7 (1): 106.
 - 26 Liu SR, McCoy AB, Wright AP, et al. Why do users override alerts? Utilizing large language model to summarize comments and optimize clinical decision support [J]. *J Am Med Inform Assoc*, 2024, 31 (6): 1388-1396.
 - 27 Gabriel RA, Mariano ER, McAuley J, et al. How large language models can augment perioperative medicine: a daring discourse [J]. *Reg Anesth Pain Med*, 2023, 48 (11): 575-577.
 - 28 Wang AN, Wu YN, Ji XJ, et al. Assessing and optimizing large language models on spondyloarthritis multi-choice question answering: protocol for enhancement and assessment [J]. *JMIR Res Protoc*, 2024, 13: e57001.
 - 29 Pool J, Indulska M, Sadiq S. Large language models and generative AI in telehealth: a responsible use lens [J]. *J Am Med Inform Assoc*, 2024, 31 (9): 2125-2136.
 - 30 Kjell ONE, Kjell K, Schwartz HA. Beyond rating scales: With targeted evaluation, large language models are poised for psychological assessment [J]. *Psychiatry Res*, 2024, 333: 115667.
 - 31 Landman R, Healey SP, Loprinzo V, et al. Using large language models for safety-related table summarization in clinical study reports [J]. *JAMIA Open*, 2024, 7 (2): ooae043.
 - 32 Datta S, Lee K, Paek H, et al. AutoCriteria: a generalizable clinical trial eligibility criteria extraction system powered by large language models [J]. *J Am Med Inform Assoc*, 2024, 31 (2): 375-385.
 - 33 Rengers TA, Thiels CA, Salehinejad H. Academic surgery in the era of large language models: a review [J]. *JAMA Surg*, 2024, 159 (4): 445-450.
 - 34 Jiang LY, Liu XC, Nejatian NP, et al. Health system-scale language models are all-purpose prediction engines [J]. *Nature*, 2023, 619 (7969): 357-362.
 - 35 Boonstra MJ, Weissenbacher D, Moore JH, et al. Artificial intelligence: revolutionizing cardiology with large language models [J]. *Eur Heart J*, 2024, 45 (5): 332-345.
 - 36 Balas M, Wadden JJ, Hébert PC, et al. Exploring the potential utility of AI large language models for medical ethics: an expert panel evaluation of GPT-4 [J]. *J Med Ethics*, 2024, 50 (2): 90-96.
 - 37 Ong JCL, Chang SYH, William W, et al. Ethical and regulatory challenges of large language models in medicine [J]. *Lancet Digit Health*, 2024, 6 (6): e428-e432.
 - 38 Karabacak M, Margetis K. Embracing large language models for medical applications: opportunities and challenges [J].

- Cureus, 2023, 15 (5) : e39305.
- 39 Deng CY, Duan YQ, Jin X, et al. Deconstructing The Ethics of Large Language Models from Long-standing Issues to New-emerging Dilemmas [J/OL] . <https://doi.org/10.48550/arXiv.2406.05392>.
- 40 Gilbert S, Harvey H, Melvin T, et al. Large language model AI chatbots require approval as medical devices [J] . *Nat Med*, 2023, 29 (10) : 2396-2398.
- 41 Azamfirei R, Kudchadkar SR, Fackler J. Large language models and the perils of their hallucinations [J] . *Crit Care*, 2023, 27 (1) : 120.
- 42 Schwartz IS, Link KE, Daneshjou R, et al. Black box warning: large language models and the future of infectious diseases consultation [J] . *Clin Infect Dis*, 2024, 78 (4) : 860-866.
- 43 Yin JQ, Bose A, Cong GJ, et al. Comparative Study of Large Language Model Architectures on Frontier [J/OL] . <https://doi.org/10.48550/arXiv.2402.00691>.
- 44 Harrison RM, Dereventsov A, Bibin A, et al. Zero-Shot Recommendations with Pre-Trained Large Language Models for Multimodal Nudging [J/OL] . <https://doi.org/10.48550/arXiv.2309.01026>.
- 45 Lawton N, Kumar A, Thattai G, et al. Neural Architecture Search for Parameter-Efficient Fine-tuning of Large Pre-trained Language Models [J/OL] . <https://doi.org/10.48550/arXiv.2305.16597>.
- 46 Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on large language models (LLMs) [J] . *NPJ Digit Med*, 2024, 7 (1) : 183.
- 47 Ferdous MM, Abdelguerf M, Ioup E, et al. Towards Trustworthy AI: A Review of Ethical and Robust Large Language Models [J/OL] . <https://doi.org/10.48550/arXiv.2407.13934>.
- 48 Truhn D, Eckardt JN, Ferber D, et al. Large language models and multimodal foundation models for precision oncology [J] . *NPJ Precis Oncol*, 2024, 8 (1) : 72.
- 49 Saab K, Tu T, Weng WH, et al. Capabilities of Gemini Models in Medicine [J/OL] . <https://doi.org/10.48550/arXiv.2404.18416>.
- 50 Anderson B. How to bridge innovation and regulation for responsible AI in healthcare [J] . *Nat Med*, 2024, 30 (5) : 1231.
- 51 Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare [J] . *NPJ Digit Med*, 2023, 6 (1) : 120.

(责任编辑: 孙菲)