大语言模型临床实践应用范围综述

施呈昊[1] 屠馨怡[2] 史佳伟[1] 陈红爽[1] 王琴潞[1] 邹海欧[1]

([1] 中国医学科学院北京协和医学院护理学院 北京 100144
[2] 浙江大学医学院附属第二医院护理部 杭州 310009)

〔摘要〕 **目的/意义** 对大语言模型在临床实践中的应用情况进行范围综述，为其推广提供参考。**方法/过程** 检索 PubMed、Embase、万方和 CNKI 数据库，筛选大语言模型应用于临床实践的相关文献，对纳入文献进行内容提取、汇总和分析。**结果/结论** 大语言模型在提供治疗建议、辅助疾病诊断、健康宣教、分析文本影像资料等方面具有应用价值，然而在答案正确率、个案化等方面的表现不尽如人意。总体来说，大语言模型在临床实践中展现出显著潜力，但仍须采取必要的措施以把控应用风险、确认适用范围。

〔关键词〕 大语言模型；人工智能；临床实践；范围综述；ChatGPT
〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2024.09.003

A Scoping Review of the Application of Large Language Models in Clinical Practice

SHI Chenghao[1], TU Xinyi[2], SHI Jiawei[1], CHEN Hongshuang[1], WANG Qinlu[1], ZOU Haiou[1]

[1] School of Nursing, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100144, China; [2] Department of Nursing, the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310009, China

〔**Abstract**〕 **Purpose/Significance** The scoping review summarizes the application of large language models in clinical practice, and provides references for their promotion. **Method/Process** PubMed, Embase, Wanfang and CNKI databases are searched to screen the literature related to the application of large language models in clinical practice, and the content of the included literature is extracted, summarized and analyzed. **Result/Conclusion** Large language models have application value in providing treatment suggestions, assisting disease diagnosis, health education, analyzing text image data, etc. However, their performance in answer accuracy and individualization is not satisfactory. In general, large language models show significant potential in clinical practice, but necessary measures must be taken to control the application risks and confirm the scope of application.

〔**Keywords**〕 large language models; artificial intelligence (AI); clinical practice; scoping review; ChatGPT

1 引言

大语言模型（large language models，LLM）作为人工智能（artificial intelligence，AI）的重要分支[1-2]，自 20 世纪 80 年代提出以来广受关注。LLM 能够通过在多种语言的大量有限文本数据集上训练，从而具有模仿人类回答的能力[3]。最初语言模型主要基于简单的统计方法，只能处理一些基本的文本任务[4]。2017 年谷歌公司提出 Transformer 模型，通过利用注意力机制提高模型处理语言的能力，进一步推动 LLM 的发展[5]。2022 年 11 月底

〔修回日期〕 2024-08-09
〔作者简介〕 施呈昊，硕士研究生；通信作者：邹海欧，博士，教授。

ChatGPT 正式上线后，其出色的自然语言处理生成能力，使其在贸易、金融、信息检索、教育等各个领域迅速普及[6-9]。在医疗领域，LLM 能够提供 24 小时医疗问题帮助，通过数据集添加和微调，满足不同专科的个性化干预需求，减轻医护人员工作负担，提高医疗服务效率和质量[10-12]。但是 LLM 训练所用数据集有限，且其中部分数据可能存在错误，会影响 LLM 的表现[7, 13]，导致其在医疗领域的应用更为敏感，未来有待进一步研究以解决这一弊端[14]。本研究探索 LLM 在临床实践中的实际应用效果，分析其优劣，为未来开发用于临床环境的 LLM 提供参考建议。

2 资料与方法

2.1 确立研究问题

依据 Arksey H 等[15]提出的范围综述框架，确立研究问题：一是目前 LLM 在临床实践中的应用方向有哪些，表现出哪些作用和优势；二是目前 LLM 在临床实践中应用的局限性有哪些；三是未来为更好地促进 LLM 在临床实践中的应用，应从哪些方面优化改进。

2.2 文献纳入排除标准

纳入标准：LLM 应用于实际医院场景中；研究内容为 LLM 在医院场景下应用的影响；文献类型为原始研究。排除标准：非中英文文献；综述类研究；无法获取全文；LLM 应用于脱离实际临床案例的简单问答。

2.3 文献检索策略

为全面检索 LLM 在临床医疗及护理实践中具体应用的相关文献，计算机检索 PubMed、Embase、万方、CNKI，以“large language models”“health/nursing/clinical”为英文检索词，“大语言模型”“临床/医疗/护理”为中文检索词，采用主题词结合关键词的方式检索，检索时限为建库至 2024 年 4 月 22 日。

2.4 文献筛选与数据提取

两名经过护理科研培训的研究者按照纳入与排除标准独立筛选文献、提取数据，并交叉核对，出现冲突时寻求第三方专家评估总结。资料提取内容包括：第一作者、发表年月、研究地区、应用模型类型、应用方向、具体作用等。

3 结果

3.1 文献检索结果

共检索获得 4 061 篇文献，经查重、阅读文题和摘要初筛，剩余 203 篇文献。根据文献纳入排除标准复筛，最终纳入 32 篇，见图 1。

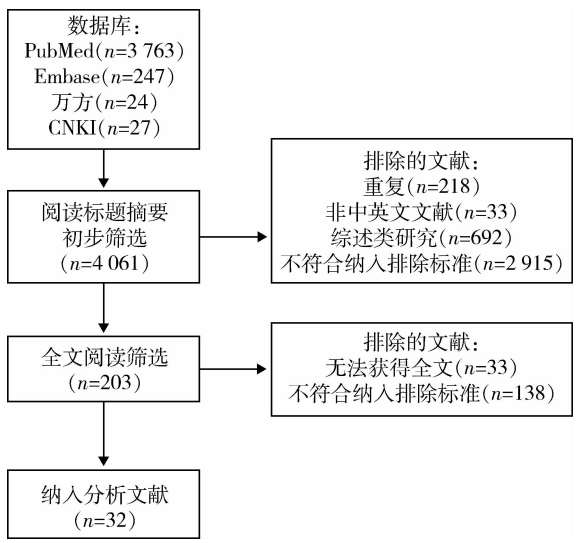

图 1 文献筛选流程

3.2 纳入文献基本信息

纳入的 32 篇文献来自 11 个国家，其中 16 篇来自美国，其他来自德国（$n=3$）、中国（$n=3$）、澳大利亚（$n=3$）等。从应用的 LLM 类型看，29 篇研究测试了 OpenAI 公司开发的 ChatGPT 及其迭代版本在临床实践中的应用，其次分别为 Google Bard（$n=8$）、New Bing（$n=5$）、Claude（$n=2$）等。纳入文献基本特征，见表 1。

表 1 纳入文献基本特征

第一作者	发表年月	国家	应用模型类型	应用方向	具体作用
Chari S[16]	2023.03	美国	BERT、SciBERT	分析文本影像资料；预测疾病风险	分析患者的实验室检查报告、用药清单等，辅助医生评估患者健康状况；预测 2 型糖尿病患者在未来 1 年内患慢性肾病的风险
Doshi R[17]	2023.06	美国	ChatGPT 3.5、ChatGPT 4、Google Bard、New Bing	通俗化医学资料	简化放射报告，增强患者对医学报告的理解，改善医患沟通
Robinson A[18]	2023.06	英国	ChatGPT 4	完成文书工作	自动生成腹腔镜阑尾切除术的手术笔记，减轻医生压力
Nov O[19]	2023.07	美国	ChatGPT	通俗化医学资料；处理行政工作	解答患者疑问，尤其是低风险问题信任度高；慢性病管理；协助处理行政任务
Liu S[20]	2023.07	美国	CLAIR – Long、ChatGPT	健康宣教	解答患者疑问，提高沟通效率，且具有一定个性化；生成电子记录以记录问答过程
Yang H[10]	2023.07	中国	ChatGLM	提供治疗建议	将模型集成到医院信息系统中，为糖尿病患者生成治疗建议
Seth I[21]	2023.08	澳大利亚	ChatGPT	提供治疗建议；辅助疾病诊断	辅助诊断腕管综合征，并提供相关治疗建议
Bernstein I A[22]	2023.08	美国	ChatGPT	提供治疗建议；处理行政工作	回答眼部护理相关问题；改进临床工作流程
Ayoub M[23]	2023.08	美国	ChatGPT	提供治疗建议；辅助疾病诊断	辅助处理心血管急性事件，提供术后护理说明；根据症状提供初步诊断建议
Ayoub N F[24]	2023.08	美国	ChatGPT	健康宣教	提供医学知识，为耳鼻喉科、头颈外科、儿科等科室的患者提供健康宣教，促进非专业人士理解
Krusche M[25]	2023.09	德国	ChatGPT 4	辅助疾病诊断	筛查和识别风湿病
Eppler M B[26]	2023.09	美国	ChatGPT	通俗化医学资料	生成易于患者理解的医学相关材料的摘要，促进医患交流
Kianian R[27]	2023.09	美国	ChatGPT、Google Bard	健康宣教	一定程度提高葡萄膜炎宣教材料可读性和准确性
Pugliese G[28]	2023.09	意大利	ChatGPT 3、ChatGPT 4	辅助疾病诊断	坏死性外耳道炎诊断和数据分析，且诊断可靠、分析快速
Nastasi A J[29]	2023.10	美国	ChatGPT	提供治疗建议	在预防保健、急性疾病管理、缓和医疗领域提供临床相关知识，简单场景下可提供决定性医学建议
Decker H[30]	2023.10	美国	ChatGPT 3.5	完成文书工作	生成易读、准确的手术知情同意书
Nazario – Johnson L[31]	2023.10	美国	ChatGPT、Glass AI	分析文本影像资料	分析神经影像学资料，提高工作效率
Franco D R[32]	2023.11	澳大利亚	ChatGPT 3.5	提供治疗建议	提供精神病学临床病例个性化治疗方案，提高方案制定效率
Bushuven S[33]	2023.11	德国	ChatGPT	提供治疗建议；辅助疾病诊断	为儿科急救提供诊断和处理意见

续表 1

第一作者	发表年月	国家	应用模型类型	应用方向	具体作用
Truhn D[34]	2023. 11	德国	ChatGPT 4	提供治疗建议；分析文本影像资料	分析磁共振成像检查报告，并提供骨科治疗建议
Mira F A[35]	2023. 11	法国	ChatGPT	辅助疾病诊断	辅助阻塞性睡眠呼吸暂停的诊断，在医疗资源有限地区具有可观的潜力
Savage T[36]	2023. 11	美国	BioMed – RoBERTa	预测疾病风险	辅助目前用于深静脉血栓风险预测的最佳实践警报（best practive alerts，BPAs），以更有效地筛选风险患者
Horiuchi D[37]	2023. 11	日本	ChatGPT	辅助疾病诊断；分析文本影像资料	对神经科患者的病历和影像报告生成诊断，面对简单资料时表现出潜力
Song H[38]	2023. 11	中国	Google Bard、Claude、ChatGPT 4、New Bing	健康宣教	用于尿路结石患者的疑问解答和健康宣教
Sun H[39]	2023. 11	中国	ChatGPT	健康宣教	2 型糖尿病管理，能有效识别食物成分，辅助营养咨询和饮食建议
Sarangi P K[40]	2023. 12	印度	ChatGPT	分析文本影像资料；通俗化医学资料	分析并简化放射学报告，使其更易于被医护人员和患者理解
Rouhi A D[41]	2024. 01	美国	ChatGPT 3. 5、Google Bard	通俗化医学资料	改善主动脉狭窄患者宣教材料的可读性
Abi – Rafeh J[42]	2024. 01	美国	Google Bard	提供治疗建议；辅助疾病诊断	整形外科术后并发症的辅助诊断，并提供治疗建议
Zhou Y[43]	2024. 02	澳大利亚	ChatGPT	提供治疗建议；分析文本影像资料	针对一例 53 岁男性股骨颈骨折的临床案例，包括影像学资料分析、辅助治疗决策以及术后管理等方面的应用
Pradhan F[44]	2024. 02	美国	ChatGPT、DocsGPT、Google Bard、New Bing	健康宣教	生成肝硬化患者教育材料
Fisch U[45]	2024. 02	瑞士	Google Bard、New Bing、Claude 2、GTP 3. 5、GTP 4、Llama、PaLM	提供治疗建议；辅助疾病诊断	辅助细菌性脑膜炎的诊断，并提供治疗建议
Huo B[46]	2024. 04	加拿大	ChatGPT 3. 5、ChatGPT 4、New Bing、Google Bard、Perplexity AI	提供治疗建议；辅助疾病诊断；健康宣教	基于临床指南辅助医生诊断胃食管反流，并提供治疗建议；向患者提供疾病相关信息

3. 3 LLM 在临床实践中的应用方向与优势

3.3.1 诊断与治疗建议 在辅助诊断和治疗建议方面[10,21-23,25,28-29,32-35,37,42,43,45]，LLM 表现出较高潜力。LLM 不仅可以辅助诊断疾病，还能协助医生制定治疗方案，从而提高决策的速度和效率。

3.3.2 数据分析与处理 通过自动化文本和影像数据分析[16, 31, 34, 37, 40, 43]，LLM 可显著提升医疗数据处理效率，使医护人员能够迅速获得关键信息，更快地响应临床需求并减轻工作压力。

3.3.3 风险预测与筛查 LLM 可以从大量数据集中识别高风险患者，预测潜在健康问题，并及时提醒医护人员，提高筛查的特异性和敏感性[16, 36]。

3.3.4 文书与行政工作处理 LLM 的应用不仅限于临床决策支持，还包括处理大量文书[18, 30]和行政工作[19, 22]，如自动完成病历记录、出入院报告和其

他文档，显著减少医护人员在这些任务上的时间投入，从而更好地为患者服务。

3.3.5 患者交流与健康宣教 LLM 能将复杂的医学信息转换成易于理解的语言，帮助患者更好地理解健康状况和治疗选项[17, 19, 26, 40-41]，从而改善医患、护患沟通，并提升患者满意度[20, 24, 27, 38, 39, 44]。

3.4 LLM 在临床实践中应用的局限性

3.4.1 缺少个案化 病例十分复杂，包括大量影像学资料和详细历史病历，导致 LLM 在提供个案化医疗建议时存在局限性[21,23,28-29,31,34]。

3.4.2 答案可靠性问题 由于现有模型多基于通用知识库，而未经专业、可靠的医学相关知识库训练[7,13,46]，LLM 应用在临床实践时，答案可靠性不足[10,16-17,19,22-23,25-27,30,32-33,35,37-41,44-46]。

3.4.3 用户接受度与依赖问题 LLM 的回答有时可能缺乏足够的依据[46]，导致用户信任度降低，机器生成的答案缺乏人类情感温度[20]、应用成本高[26]也会影响用户接受度。

3.4.4 隐私与数据安全问题 由于患者健康档案的敏感性，将患者病历资料上传至提供 LLM 服务的服务商云端，存在隐私泄漏的可能，对患者生活造成一定困扰，解决隐私和数据安全问题在 LLM 应用中至关重要[32,46]。

4 讨论

4.1 模型微调与优化

为了优化 LLM 在特定医疗任务中的表现，必须对其进行持续的微调[18,28]。通过专门的医疗数据训练，增强其对医疗术语和病例的理解能力[19-21,29,31]，对提高 LLM 在临床实践中的应用效果尤为重要。目前用于优化 LLM 在特定领域表现的常见方法主要有全参微调、LoRA、GraphRAG、检索增强生成以及提示词工程。其中，GraphRAG 由于相对低廉的成本与较强的可靠性被越来越多的开发者应用。GraphRAG 通过构建知识图谱来增强 LLM 对复杂医疗场景的理解与处理能力，不仅能够处理来自电子病历、实验室报告等的数据，还可以将这些信息整合成知识图谱，从而提升 LLM 的上下文理解能力；这种优化方法的另一个优势在于能够生成带有来源标注的回答，确保每个生成结果可以追溯到其原始数据来源，从而提高用户对回答的信任感，这对于需要严格把控数据准确性和安全性的医疗领域尤为重要[47]。通过微调与优化，LLM 可以进一步改善对患者数据的处理，提供更加个性化、基于证据的医疗建议。

4.2 基于 LLM 的智能决策支持系统开发

将 LLM 与医院信息系统（hospital information system，HIS）相结合，有助于显著提升医疗决策的精准度和效率[10]。开发者应当使 LLM 的调用足够便捷、高效，可以通过侧边栏、浮窗等形式将 LLM 的问答界面集成到 HIS 界面中，使 LLM 能够自动获取当前选中患者的数据，并据此回答医护人员关心的问题，提高回答的准确率和个性化水平；将 LLM 集成入 HIS，还能使其获得将当前病例与历史病例进行匹配的能力，从而提供相似病例的诊断和治疗参考，推荐个性化的治疗方案，包括药物选择和剂量调整等。除了被动问答外，LLM 还应主动监测患者的各项数据，如生命体征、实验室指标等，当这些数据出现异常时，及时提醒医护人员并推荐相应的处理措施，在提高工作效率的同时减轻医护人员的工作负担。

4.3 建立有效监督机制

在 LLM 应用于医疗领域的过程中，建立有效的监督机制是保障使用者权益的关键。受限于当前技术水平，LLM 应用于临床实践存在诸多不确定性，容易产生风险[10,16-17,19,22-23,25-27,30,32-33,35,37-41,44-46]；且由于涉及患者个人信息，隐私问题不容忽视[48]。因此相关行政机构应当根据当前技术水平及研究现状，评估 LLM 的临床适用性和潜在风险[16-17,22-23,25-36,39]，完善相关法律法规以规范患者数据的来源和用途、LLM 在临床实践中的应用范围和方式[49]；医疗机构应当建立完善的 LLM 监控系统及危险词库，实时追踪 LLM 的操作和输出，当 LLM 输出危险词库中的内容时及时干预，以便修正

可能的错误或偏差，确保 AI 的临床应用风险始终处于可控范围。在应用过程中，应使用加密技术保护数据，建立严格的数据访问和处理制度，确保对患者数据的全面保护[50]。此外还需建立方便、快捷的反馈渠道，使 LLM 在应用过程中暴露的具体问题得到及时处理。

5 结语

LLM 在临床实践领域具有显著的应用潜力，在疾病诊断、治疗建议、健康宣教等方面表现良好，可以加速临床决策过程，优化医患、护患交流，并协助处理行政任务。然而，LLM 的使用也面临着准确性不足、特异性低和用户接受度不足等挑战。未来研究应重点关注提高 LLM 在临床环境中的性能和可靠性，同时考虑个案化的解决方案以满足复杂多变的医疗需求。总体而言，LLM 在临床实践中的应用呈现积极前景，但仍待深入探索与发展以充分发挥其潜力。

作者贡献：施呈昊负责研究设计、文献搜集与整理、数据分析与解释、论文撰写与修订；屠馨怡负责文献搜集与整理、数据分析与解释、论文撰写与修订；史佳伟、陈红爽、王琴潞负责数据分析与解释、论文修订；邹海欧负责项目管理、提供指导。
利益声明：所有作者均声明不存在利益冲突。

参考文献

1 KORTELING J, VAN DE BOER - VISSCHEDIJK G C, BLANKENDAAL R, et al. Human - versus artificial intelligence [EB/OL]. [2024 - 04 - 14]. https://www.frontiersin.org/journals/artificial - intelligence/articles/10.3389/frai.2021.622364/full.

2 ARORA A, ARORA A. The promise of large language models in health care [J]. Lancet, 2023, 401 (10377): 641.

3 BROWN T, MANN B, RYDER N, et al. Language models are few - shot learners [EB/OL]. [2024 - 02 - 14]. https://arxiv.org/abs/2005.14165.

4 MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [EB/OL]. [2024 - 02 - 14]. https://arxiv.org/abs/1301.3781v1.

5 DEVLIN J, CHANG M W, LEE K, et al. Bert: pre - training of deep bidirectional transformers for language understanding [EB/OL]. [2024 - 02 - 14]. https://arxiv.org/abs/1810.04805v2.

6 BROWN T B, MANN B, RYDER N, et al. Language models are few - shot learners [EB/OL]. [2024 - 04 - 14]. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a - Paper.pdf.

7 DENG J, LIN Y. The benefits and challenges of ChatGPT: an overview [J]. Frontiers in computing and intelligent systems, 2022, 2 (2): 81 - 83.

8 CHATTERJEE J, DETHLEFS N. This new conversational AI model can be your friend, philosopher, and guide and even your worst enemy [J]. Patterns, 2023, 4 (1): 1 - 3.

9 ZHU Y, YUAN H, WANG S, et al. Large language models for information retrieval: a survey [EB/OL]. [2024 - 04 - 14]. https://ar5iv.labs.arxiv.org/html/2308.07107.

10 YANG H, LI J, LIU S, et al. Exploring the potential of large language models in personalized diabetes treatment strategies [EB/OL]. [2024 - 02 - 14]. https://www.medrxiv.org/content/10.1101/2023.06.30.23292034v1.

11 REDDY S. Evaluating large language models for use in healthcare: a framework for translational value assessment [EB/OL]. [2024 - 04 - 14]. https://www.sciencedirect.com/science/article/pii/S2352914823001508.

12 彭婉琳, 陈德凤, 李蓓, 等. ChatGPT 人工智能语言机器人在护理领域应用现状与展望 [J]. 全科护理, 2023, 21 (35): 4934 - 4937.

13 STOKEL - WALKER C, VAN N R. What ChatGPT and generative AI mean for science [J]. Nature, 2023, 614 (7947): 214 - 216.

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

15 ARKSEY H, O' MALLEY L. Scoping studies: towards a methodological framework [J]. International journal of social research methodology, 2005, 8 (1): 19 - 32.

16 CHARI S, ACHARYA P, GRUEN D M, et al. Informing clinical assessment by contextualizing post - hoc explanations of risk prediction models in type - 2 diabetes [EB/OL]. [2024 - 04 - 14]. https://www.sciencedirect.com/science/article/pii/S093336572300012X.

17 DOSHI R, AMIN K, KHOSLA P, et al. Utilizing large lan-

guage models to simplify radiology reports: a comparative analysis of ChatGPT 3.5, ChatGPT 4.0, Google Bard, and Microsoft Bing [EB/OL]. [2024-02-14] https://www.medrxiv.org/content/10.1101/2023.06.04.23290786v2.

18 ROBINSON A, AGGARWAL S J. When precision meets penmanship: ChatGPT and surgery documentation [J]. Cureus, 2023, 15 (6): e40546.

19 NOV O, SINGH N, MANN D. Putting ChatGPT's medical advice to the (turing) test: survey study [EB/OL]. [2024-04-14]. https://mededu.jmir.org/2023/1/e46939.

20 LIU S, MCCOY A B, WRIGHT A P, et al. Leveraging large language models for generating responses to patient messages [EB/OL]. [2024-02-14]. https://www.medrxiv.org/content/10.1101/2023.07.14.23292669v1.

21 SETH I, XIE Y, RODWELL A, et al. Exploring the role of a large language model on carpal tunnel syndrome management: an observation study of ChatGPT [J]. The journal of hand surgery, 2023, 48 (10): 1025-1033.

22 BERNSTEIN I A, ZHANG Y V, GOVIL D, et al. Comparison of ophthalmologist and large language model chatbot responses to online patient eye care questions [J]. JAMA network open, 2023, 6 (8): e2330320.

23 AYOUB M, BALLOUT A A, ZAYEK R A, et al. Mind + Machine: ChatGPT as a basic clinical decisions support tool [J]. Cureus, 2023, 15 (8): e43690.

24 AYOUB N F, LEE Y J, GRIMM D, et al. Head-to-head comparison of ChatGPT versus google search for medical knowledge acquisition [EB/OL]. [2024-02-14]. https://aao-hnsfjournals.onlinelibrary.wiley.com/doi/10.1002/ohn.465.

25 KRUSCHE M, CALLHOFF J, KNITZA J, et al. Diagnostic accuracy of a large language model in rheumatology: comparison of physician and ChatGPT-4 [J]. Rheumatology international, 2024, 44 (2): 303-306.

26 EPPLER M B, GANJAVI C, KNUDSEN J E, et al. Bridging the gap between urological research and patient understanding: the role of large language models in automated generation of layperson's summaries [J]. Urology practice, 2023, 10 (5): 436-443.

27 KIANIAN R, SUN D, CROWELL E L, et al. The use of large language models to generate education materials about uveitis [J]. Ophthalmology retina, 2024, 8 (2): 195-201.

28 PUGLIESE G, MACCARI A, FELISATI E, et al. Are artificial intelligence large language models a reliable tool for difficult differential diagnosis? An a posteriori analysis of a peculiar case of necrotizing otitis externa [J]. Clinical case reports, 2023, 11 (9): e7933.

29 NASTASI A J, COURTRIGHT K R, HALPERN S D, et al. A vignette-based evaluation of ChatGPT's ability to provide appropriate and equitable medical advice across care contexts [J]. Scientific reports, 2023, 13 (1): 17885.

30 DECKER H, TRANG K, RAMIREZ J, et al. Large language model-based chatbot vs surgeon-generated informed consent documentation for common procedures [J]. JAMA network open, 2023, 6 (10): e2336997.

31 NAZARIO-JOHNSON L, ZAKI H A, TUNG G A. Use of large language models to predict neuroimaging [J]. Journal of the American college of radiology, 2023, 20 (10): 1004-1009.

32 FRANCO D R, AMANULLAH S, MATHEW M, et al. Appraising the performance of ChatGPT in psychiatry using 100 clinical case vignettes [EB/OL]. [2024-04-14]. https://www.sciencedirect.com/science/article/pii/S1876201823003026X.

33 BUSHUVEN S, BENTELE M, BENTELE S, et al. "ChatGPT, Can you help me save my child's life?"-diagnostic accuracy and supportive capabilities to lay rescuers by ChatGPT in prehospital basic life support and paediatric advanced life support cases-an in-silico analysis [J]. Journal of medical systems, 2023, 47 (1): 123.

34 TRUHN D, WEBER C D, BRAUN B J, et al. A pilot study on the efficacy of GPT-4 in providing orthopedic treatment recommendations from MRI reports [J]. Scientific reports, 2023, 13 (1): 20159.

35 MIRA F A, FAVIER V, DOS S S N H, et al. ChatGPT for the management of obstructive sleep apnea: do we have a polar star[EB/OL]. [2024-02-14]. https://link.springer.com/article/10.1007/s00405-023-08270-9.

36 SAVAGE T, WANG J, SHIEH L. A large language model screening tool to target patients for best practice alerts: development and validation [EB/OL]. [2024-04-14]. https://medinform.jmir.org/2023/1/e49886.

37 HORIUCHI D, TATEKAWA H, SHIMONO T, et al. Accuracy of ChatGPT generated diagnosis from patient's medical history and imaging findings in neuroradiology cases [J].

Neuroradiology, 2024, 66 (1): 73 - 79.

38 SONG H, XIA Y, LUO Z, et al. Evaluating the performance of different large language models on health consultation and patient education in urolithiasis [J]. Journal of medical systems, 2023, 47 (1): 125.

39 SUN H, ZHANG K, LAN W, et al. An AI dietitian for type 2 diabetes mellitus management based on large language and image recognition models: preclinical concept validation study [EB/OL]. [2024 - 04 - 14]. https://www.jmir.org/2023/1/e51300/.

40 SARANGI P K, LUMBANI A, SWARUP M S, et al. Assessing ChatGPT's proficiency in simplifying radiological reports for healthcare professionals and patients [J]. Cureus, 2023, 15 (12): e50881.

41 ROUHI A D, GHANEM Y K, YOLCHIEVA L, et al. Can artificial intelligence improve the readability of patient education materials on aortic stenosis? A pilot study [EB/OL]. [2024 - 02 - 19]. https://link.springer.com/article/10.1007/s40119 - 023 - 00347 - 0.

42 ABI - RAFEH J, MROUEH V J, BASSIRI - TEHRANI B, et al. Complications following body contouring: performance validation of bard, a novel AI large language model, in triaging and managing postoperative patient concerns[EB/OL]. [2024 - 02 - 19]. https://link.springer.com/article/10.1007/s00266 - 023 - 03819 - 9.

43 ZHOU Y, MOON C, SZATKOWSKI J, et al. Evaluating ChatGPT responses in the context of a 53 - year - old male with a femoral neck fracture: a qualitative analysis [J]. European journal of orthopaedic surgery & traumatology: orthopedie traumatologie, 2024, 34 (2): 927 - 955.

44 PRADHAN F, FIEDLER A, SAMSON K, et al. Artificial intelligence compared with human - derived patient educational materials on cirrhosis [J]. Hepatol communication, 2024, 8 (3): e0367.

45 FISCH U, KLIEM P, GRZONKA P, et al. Performance of large language models on advocating the management of meningitis: a comparative qualitative study [J]. BMJ health & care informatics, 2024, 31 (1): e100978.

46 HUO B, CALABRESE E, SYLLA P, et al. The performance of artificial intelligence large language model - linked chatbots in surgical decision - making for gastroesophageal reflux disease [EB/OL]. [2024 - 04 - 14]. https://link.springer.com/article/10.1007/s00464 - 024 - 10807 - w.

47 WU J, ZHU J, QI Y. Medical graph RAG: towards safe medical large language model via graph retrieval - augmented generation [EB/OL]. [2024 - 08 - 09]. https://arxiv.org/abs/2408.04187.

48 夏光辉, 曹艳林, 陈炳澍, 等. 大模型人工智能技术在医疗服务领域应用的专家共识 [J]. 中国卫生法制, 2023, 31 (5): 124 - 126.

49 MESKO B, TOPOL E J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare [EB/OL]. [2024 - 04 - 22]. https://link.springer.com/article/10.1007/s00464 - 024 - 10807 - w#citeas.

50 VERSPEURT J. Staying compliant and in control of your data when using ChatGPT and other large language models [EB/OL]. [2023 - 11 - 17]. https://insights.radix.ai/blog/staying - compliant - and - in - control - of - your - data - when - using - chatgpt - and - other - large - language - models.