

医疗基础大模型的应用现状、挑战及展望

李小鹏^{1,2} 喻慧心¹ 常青¹ 赵玉虹^{1,2}

(¹ 中国医科大学附属盛京医院 沈阳 110004 ² 中国医科大学健康管理学院 沈阳 110122)

[摘要] **目的/意义** 系统评估医疗基础大模型 (medical foundation model, MFM) 的发展脉络、应用现状及核心挑战, 以评估其临床转化潜力与局限。**方法/过程** 通过系统文献综述, 分析 MFM 从通用模型到医学特化的技术演进, 进而分析其在数据、模型、伦理等方面的关键挑战。**结果/结论** MFM 通过多模态融合、检索增强等技术, 能有效提升诊疗效率与决策可信度。然而, 数据孤岛、算法偏见及生成不确定性限制其规模化应用。未来应推动其向可解释、可验证的医学智能体演进, 并构建技术创新、数据治理与伦理监管协同的生态, 促进 MFM 安全可持续发展。

[关键词] 医疗基础大模型; 多模态融合; 隐私与伦理; 智慧医疗

[中图分类号] R-058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2026.01.003

Application Status, Challenges and Prospects of Medical Foundation Models

LI Xiaopeng^{1,2}, YU Huixin¹, CHANG Qing¹, ZHAO Yuhong^{1,2}

¹Shengjing Hospital, China Medical University, Shenyang 110004, China; ²School of Health Management, China Medical University, Shenyang 110122, China

[Abstract] **Purpose/Significance** To systematically review the developmental trajectory, current applications and core challenges of medical foundation models (MFM), and to assess their potential and limitations in clinical transformation. **Method/Process** Through a systematic literature review, the technical evolution of MFMs from general-purpose foundation models to domain-specific medical adaptations is analyzed. Its key challenges in terms of data, models and ethics are analyzed. **Result/Conclusion** MFM can effectively enhance the efficiency of diagnosis and treatment as well as the credibility of decision-making through technologies such as multimodal fusion and retrieval augmentation. However, data silos, algorithmic biases and generation uncertainties limit its large-scale application. In the future, it is necessary to promote its evolution towards interpretable and verifiable medical agents, and build a coordinated ecosystem that integrates technological innovation, data governance and ethical supervision to promote the safe and sustainable development of MFM.

[Keywords] medical foundation model (MFM); multimodal fusion; privacy and ethics; smart healthcare

1 引言

近年来, 我国高度重视人工智能 (artificial in-

telligence, AI) 在医疗健康领域的创新应用与产业化发展。2022 年《“十四五”全民健康信息化规划》^[1] 强调以数字化、智能化促进医疗健康行业转型; 2024 年《卫生健康行业人工智能应用场景参考

[修回日期] 2025-11-16

[作者简介] 李小鹏, 硕士研究生, 发表论文 2 篇; 通信作者: 赵玉虹, 博士, 教授。

[基金项目] 国家重点研发计划项目 (项目编号: 2023YFC3604600); 辽宁省科技计划联合计划 (项目编号: 2024JH2/102600322)。

指引》^[2]旨在推动“人工智能+”在医疗领域的创新应用。一系列递进式、多层次政策框架,为医疗基础大模型(medical foundation model, MFM)^[3]在医疗健康领域规范化应用提供了制度保障。然而,随着 MFM 的深入发展,其对模型可信性、可解释性以及偏差控制等方面的要求日益突显,相关研究不再局限于技术层面,而是更多涉及医疗规范、责任边界与安全体系构建等综合性议题。

基于上述背景,本研究系统梳理大语言模型在医学领域的应用现状,结合最新研究进展,总结其发展路径,深入分析其面临的挑战,并在此基础上展望未来发展方向,旨在为医学研究者提供结构清晰、可追溯的参考,助力行业规范化发展,从而为构建智慧医疗生态、推动我国医疗健康体系向数字化、智能化演进提供持续动能,助力行业高质量发展。

2 医疗基础大模型在医学领域的发展及应用

在政策引导与技术革新驱动下, MFM 作为 AI 技术重要突破,正展现其在医疗健康领域的变革性潜力。MFM 指专为医疗场景设计的基础大模型,基于大规模医疗多模态数据的自监督预训练,实现跨模态理解、生成与推理^[4-5]。与通用基础大模型(foundation model, FM)相比, MFM 的训练数据高度领域特化,需要高质量医疗数据集,而非广义互联网数据; MFM 强调临床可解释性、可验证推理与隐私合规,而非通用任务泛化; MFM 评估标准聚焦医学基准,而非通用指标。

MFM 的发展历程可追溯至自然语言处理(natural language processing, NLP)和计算机视觉领域范式转变。自 2017 年 Transformer 被提出^[6],其多头分层自注意力机制深刻重塑 NLP 领域发展轨迹。2018 年后, BERT^[7]、PaLM^[8]、LLaMA^[9-10]、GPT-4^[11]等文本模型, ViT、Swin Transformers^[12-13]等视觉模型,推动“预训练+微调”学习范式广泛应用。2020 年 GPT-3 模型发布,其卓越少样本学习能力奠定 FM 的基础。然而,当时 MFM 发展面临显著技术挑战:一方面,“幻觉”问题导致输出信息可靠性不足;另一方面,缺乏针对医疗领域的专业

化微调策略,难以适应临床场景特殊需求。此外,领域高质量标注数据严重匮乏^[13],单纯依赖模型参数规模扩张面临边际效益递减。

2022 年 ChatGPT 通过基于人类反馈的强化学习(reinforcement learning from human feedback, RLHF)^[14]与监督微调(supervised fine-tuning, SFT)^[15]显著提升输出质量。将该方法论迁移至医学场景,为 MFM 发展提供了助力,利用指令微调、链式思维提示^[16]和知识蒸馏^[17]等技术,可增强模型在医学任务中的推理能力与领域适配性。同时, MedMCQA^[18]、PubMedQA^[19]、MMLU-pro^[20]等生物医学专用评测数据集逐步完善,为模型性能评估提供标准化依据^[21]。

截至 2023 年,相关技术持续迭代与突破,如 GPT-4 系列优化算力调度与层级并行机制,使模型具有更低延迟与相对良好推理效率^[11]; Gemini 2.0^[22]采用 Flash Thinking 架构,基于层次化多路并行解码,缩短多模态即时推理任务响应时间并提升准确性。针对医疗健康领域,其应用深度与广度拓展,应用效能显著提升,如 DeepSeek R1 在中英文眼科复杂临床推理任务中的准确率超越此前所有公开模型^[23]。同时,多模态整合与多任务学习^[24]使 MFM 能够协同处理文本、影像与音频等多源数据^[25],为智慧医疗发展提供创新技术路径。MFM 在医学领域的应用成果显著:已在电子病历语义理解与结构化处理^[26]、医疗问答^[27]、疾病辅助诊断^[28]等核心场景形成成熟解决方案,提升医疗服务效率和质量;同时,在医学教育^[29]、药物研发^[30]、医疗虚拟数字人^[31]等新兴领域展现出显著的应用潜力。

本研究采用系统文献综述法,以“大语言模型”“医学”“medical foundation model”等为关键词,检索中国知网、PubMed 和 Web of Science 等文献数据库,发文时间限定为 2023—2025 年。全面研读文献并结合专家意见,为清晰呈现 MFM 研究方向与应用路径,依据任务目标与技术取向,将纳入研究划分为 5 类:会话型临床模型、多模态临床诊断与报告模型、检索增强的可验证临床决策支持模型、域特化与参数高效医学语言模型、知识增强的可验证临床推理模型,见表 1。近年来已有研究在

真实临床工作中验证 MFM 的可行性与效能。例如, Wang B 等^[32]证实 ChatGLM2 - 6B 应用于临床数据提取可使转录时间减少 80.7%; Wan P 等^[33]通过随机对照试验发现 MFM 能显著提升患者满意度和沟通效率; Garcia P 等^[34]证明基于 GPT 的患者信息回复草稿可有效减轻临床医生工作负担。随时间推移, MFM 研究呈现“功能化 - 证据化 - 工程化”并行演进趋势。具体而言, 会话型与多模态方向在交互性与跨模态表征上推动临床辅助流程自动化;

检索增强与知识增强的推理方法是抑制生成性“幻觉”、提升可追溯性的关键路径; 域特化与参数高效思路为资源受限环境的临床部署提供务实方案。上述发展路径不仅清晰勾勒 MFM 研究的演进脉络, 更逐步构建起从单一任务到全流程智能化的技术生态。在该生态中, 医学域自适应与临床评估相互促进, 各应用场景间的协同优化形成正反馈循环, 从而持续加速 MFM 在医疗健康领域的深度整合与创新。

表 1 2023—2025 年代表性 MFM 概况

模型	开发时间(年)	开发机构	类型	模型特点
PsyLLM ^[35]	2023	悉尼大学等	会话型临床模型	面向心理支持的多场景会话模型, 含高危筛查
BianQue ^[36]	2023	华南理工大学等	会话型临床模型	链式多轮问答微调, 支持连续且个性化的患者建议生成
ChatDoctor ^[37]	2023	得克萨斯大学等	会话型临床模型	基于 LLaMA 进行医患对话微调, 实现持续交互与自适应诊疗
ClinicalGPT ^[38]	2023	北京邮电大学等	检索增强的可验证临床决策支持模型	利用真实电子病历微调, 采用多源数据融合推理机制
PMC - LLaMA ^[39]	2023	上海科技大学等	检索增强的可验证临床决策支持模型	使用领域语料库预训练, 根据数据集使用指令进行微调
LLaVA - Med ^[40]	2023	微软	多模态临床诊断与报告模型	基于视觉 - 语言少样本协同的临床图文助手
MedPaLM2 ^[41]	2023	谷歌	知识增强的可验证临床推理模型	采用知识增强架构, 已在医学执照考试等基准测试中接近专家水平
BioMedGPT ^[42]	2024	利哈伊大学等	域特化与参数高效医学语言模型	开源、轻量的视觉 - 语言医学通用模型
BioMedLM - 2.7B ^[43]	2024	斯坦福大学等	域特化与参数高效医学语言模型	小规模生物医学语言模型, 在多项选择与问答中达到较高水平
Me - LLaMA ^[44]	2024	耶鲁大学等	域特化与参数高效医学语言模型	面向医学领域的轻量化基座, 采用“持续预训练 + 高效微调”策略
DeepDR - LLM ^[45]	2024	上海交通大学等	多模态临床诊断与报告模型	针对糖尿病视网膜病变的影像 - 文本多模态诊断模型
Polaris ^[46]	2024	Hippocratic AI	会话型临床模型	安全优先的大语言模型群组架构, 降低会话“幻觉”
扁鹊 2.0 ^[36]	2024	华南理工大学	会话型临床模型	结合混合指令与真实医患问诊数据, 增强问诊能力
Medical mT5 ^[47]	2024	巴斯克大学等	域特化与参数高效医学语言模型	多语种文本 - 文本医学模型
Zhongjing ^[48]	2024	郑州大学	会话型临床模型	基于大规模真实医患对话的强化学习训练模型
IHILLM - RAG ^[49]	2025	清华大学	检索增强的可验证临床决策支持模型	私有化检索增强生成方案, 强调安全性与个性化交互
AlphaMed ^[50]	2025	帝国理工大学等	知识增强的可验证临床推理模型	通过规则强化学习训练的可推理医疗问答模型
MedReason - 8B ^[51]	2025	加州大学等	知识增强的可验证临床推理模型	基于高质量知识图谱生成来验证推理链的模型
HealthGPT ^[52]	2025	浙江大学等	多模态临床诊断与报告模型	采用低秩适配的视觉 - 文本统一生成模型
FineMedLM - o1 ^[53]	2025	复旦大学等	会话型临床模型	结合高质量合成数据的监督微调与直接偏好优化, 增强长对话与推理能力
MedGemma ^[54]	2025	谷歌	多模态临床诊断与报告模型	聚焦图像与文本深度融合, 为诊断与治疗提供更有力的支持

3 医疗基础大模型在医学领域应用面临的挑战

3.1 数据复杂性

医疗健康数据涵盖电子健康记录、医学影像、基因组等多模态数据，其异构性与高维性对 MFM 处理能力提出严峻挑战。当前基于特征级融合的多模态处理方法难以建立有效的跨模态关联，影响联合表征质量与决策可靠性。医疗健康数据常见结构化程度不足、标准不一、数据噪声等问题，以及手写病历光学字符识别错误、影像格式不统一、化验指标记录异常等，严重削弱数据质量，降低模型可用的有效信息比例。在分布上，机构间标准与流程差异形成数据孤岛，阻碍协同共享；少数族裔与罕见病数据匮乏，直接引入模型偏差风险。面对这些问题，应开发深度融合的多模态模型并行处理文本、影像等数据，结合差分隐私和统一数据标准，打破数据孤岛，促进 MFM 在医疗领域的应用。

3.2 可解释性、知识整合与生成可靠性

MFM 在医疗健康场景应用面临技术 - 临床耦合的多重挑战。首先，MFM “黑箱” 特性使其推理过程难以提供证据级解释，尤其在处理多模态上下文时。其次，医学知识整合局限，现有检索增强生成等技术多停留于知识后置拼接，难以实现原生整合，导致知识利用效率低下。MFM 存在生成可靠性缺陷，最突出的是医学内容生成时的“幻觉”现象。Busch F 等^[55]认为，FM 常有 3 重可靠性缺陷，即模型输出不准确性、不完整性及根本性错误结论；该现象在 Zhang B 等^[56]构建的 CliniFact 数据集中得到量化验证。最后，高性能大模型通常有上千亿参数，训练与推理需大量算力，高昂成本构成临床应用的重要障碍。未来研究方向包括：一是将可解释性工具与证据链回溯结合，输出时呈现支持结论的原始文献或病例证据，便于临床复核；二是采用知识蒸馏与参数高效微调进行模型压缩，构建“蒸馏 - 场景化评测 - 在线监控”闭环，模型压缩后，在与目标临床场景一致的数据上评估一致性，

部署后持续监测模型漂移与“幻觉”指标。

3.3 数据隐私、伦理与偏见

医疗数据具有高敏感性，要求严格遵守隐私法规。直接使用未脱敏数据训练 MFM 易导致泄露，因此数据去标识化、差分隐私、联邦学习等技术在医疗领域广泛应用^[57]。然而，尚无明确法规界定 MFM 在临床决策错误时的责任归属。此外，训练数据中的系统性偏见易被放大，导致诊疗不公，Omar M 等^[58]研究发现，MFM 在急诊决策等场景中可能产生基于性别或种族的偏见。对此，应引入公平性评估指标和跨群体检测，建立 MFM 全生命周期伦理与监管框架，通过知识图谱和去偏见微调降低决策偏见^[59]；明确责任归属，通过持续审计完善监管；通过跨学科协作优化 MFM 临床适配并在基层部署，利用模型压缩和边缘计算，多语种扩展训练语料，支持全球医疗公平。

3.4 医学数据集建设与更新

高质量多模态医学数据集是 MFM 优化与评估的基石，然而当前数据生态面临多重挑战。一是现有公开数据集时效性不足，难覆盖新发疾病与罕见病，且缺乏多中心、多模态临床数据整合。二是美国《健康保险携带和责任法案》、欧盟《通用数据保护条例》等法律法规保障数据安全的同时，增加跨机构、跨国数据共享难度。三是数据标注依赖临床医务人员，成本高昂，尤其是影像与文本联合标注。为此，应探索基于半监督与弱监督学习的智能数据标注系统，以提升效率；推动建立开源协作机制，鼓励医疗机构与科研团队联合发布多模态公共数据集，为 MFM 持续迭代优化提供高质量、多样化数据支持，有效促进技术研发向临床应用转化。

4 未来发展方向

4.1 从会话模型到医学智能体演进

大语言模型在医学领域正从单向工具向可交互、可协同、可解释的智能体形态演进。基于 MFM 的医学智能体可嵌入临床 workflow，支持多步骤向

诊、患者随访及康复指导等复杂场景。相比早期会话模型，医学智能体支持更精确的交互语义、状态维护与任务管理策略。未来研究应优先构建场景化安全评估指标体系，优化人机协同介入点与权限策略，并通过受控临床试验验证智能体在真实工作流程中的安全性和效能。此外，应系统评估智能体在多角色协作场景中的责任划分与伦理风险，明确人类最终决策权的保留与上报机制。

4.2 多模态诊断的可解释推理与证据闭环

多模态诊断模型旨在将影像、检验与电子病历文本等多源信息整合为可审计诊断结论。MFM 解释研究正由注意力可视化转向因果推理，通过知识图谱与神经网络交互，建立从临床症状到诊断结论的显式推理路径，并在报告中按模态标注支持证据与置信区间，便于临床回溯。应强化模态间证据一致性与因果可验证性，构建模块化解释接口、证据标签化与版本化机制，形成内外证据协同且可追溯的闭环。

4.3 检索增强的临床决策支持研究

检索增强临床决策支持模型将预训练模型与外部知识库、指南与电子健康记录结合，使输出结果具备可追溯证据，减少“幻觉”并提升临床可验证性。关键在于建立语义索引与版本化机制，确保证据来源清晰、时效可控；同时通过结构化片段校验模态一致性。应设计检索信任度量与证据优先级策略，在证据冲突时输出不确定性标记。此类模型能在复杂诊疗场景下提供可溯源诊疗建议，支持临床人员快速核查与比对证据，成为辅助决策的证据生成器，而非“黑箱”答案提供者。

4.4 领域特化与参数高效微调的临床部署策略

医疗场景下，领域知识深度与隐私约束使扩大参数规模并非最优路径。更可行的是以领域特化和参数高效微调实现可部署化：在通用基座上进行持续医学预训练，结合轻量化方法（如 LoRA、Adapter）完成任务适配，实现性能与成本平衡。结合知识蒸馏、量化与差分隐私，实现院内离线更新与在

线服务。应构建“基座域适配 - 参数高效微调 - 场景化回测 - 蒸馏与压缩 - 在线监控”闭环更新流程，并制定可量化验收标准，支持隐私敏感与资源受限场景快速部署。

4.5 知识增强的可验证推理与可信智能体框架

知识增强临床推理模型嵌入了医学知识图谱、本体体系与临床指南，使生成结果具备知识溯源与逻辑可检验性。与检索增强不同，其可实现结构化知识与神经推理深度融合，强化结论可解释性与可追溯性。系统实现层面，为支撑复杂推理机制在真实医疗流程中可执行、可监管，应构建以医学智能体和模型上下文协议（model context protocol, MCP）为核心的协同框架。智能体负责任务分解、外部工具调用与多轮信息整合；MCP 定义模型、知识库与外部系统交互标准，实现上下文状态与证据链结构化传递。该体系为临床 AI 提供从知识驱动推理到工程化部署的桥梁。结合人在环中机制、证据版本化与在线“幻觉”监测，实现模型推理全流程可审计与安全闭环，构建可信、可复现、可监管的 MFM 生态。

5 结语

本研究系统综述 MFM 发展脉络、关键技术与临床应用路径，梳理其在多模态融合、检索增强生成、参数高效微调及知识可验证推理等方面的最新进展。MFM 正推动医学人工智能从工具型向智能体范式演进，为智慧医疗精准化与普惠化奠定基础。本研究仍存在一定局限：一是受限于数据更新周期与模型开源透明度，部分最新成果可能未被纳入；二是伦理、隐私与治理等议题的交叉研究视角有待进一步深化。未来研究应以构建面向临床应用的多维度 MFM 评价体系为核心，将安全性、临床效用、可解释性与证据可追溯性等量化指标纳入统一框架；同时推进 MFM 与结构化知识图谱及因果推理模块深度融合，将生成性输出锚定为可证实证据链，以抑制“幻觉”并实现可审计推理路径。

作者贡献：李小鹏负责研究设计、文献调研、论文

撰写; 喻慧心、常青负责研究设计、论文修订; 赵玉虹负责提出选题、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 “十四五”全民健康信息化规划 [EB/OL]. [2025 - 08 - 13]. <https://www.nhc.gov.cn/guihuaxxs/c100133/202211/fae867435fea479c828cb50047526a69.shtml>.
- 2 卫生健康行业人工智能应用场景参考指引 [EB/OL]. [2025 - 08 - 13]. <https://www.nhc.gov.cn/guihuaxxs/c100133/202411/3dee425b8dc34f739d63483c4e5c334c.shtml>.
- 3 MOOR M, BANERJEE O, ABAD Z S H, et al. Foundation models for generalist medical artificial intelligence [J]. *Nature*, 2023, 616 (7956): 259 - 265.
- 4 LIÉVIN V, HOTHER C E, MOTZFELDT A G, et al. Can large language models reason about medical questions [J]. *Patterns*, 2024, 5 (3): 100943.
- 5 KHAN W, LEEM S, SEE K B, et al. A comprehensive survey of foundation models in medicine [J]. *IEEE reviews in biomedical engineering*, 2025 (5): 3531360.
- 6 VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]. Long Beach: *Advances in Neural Information Processing Systems*, 2017.
- 7 DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]. Minneapolis: *The 2019 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies*, 2019.
- 8 ANIL R, DAI A M, FIRAT O, et al. PaLM 2 technical report [EB/OL]. [2025 - 08 - 13]. <https://doi.org/10.48550/arXiv.2305.10403>.
- 9 TOUVRON H, LAVRIL T, IZACARD G, et al. LLaMa: open and efficient foundation language models [EB/OL]. [2025 - 08 - 13]. <https://doi.org/10.48550/arXiv.2302.13971>.
- 10 TOUVRON H, MARTIN L, STONE K, et al. LLaMa 2: open foundation and fine-tuned chat models [EB/OL]. [2025 - 08 - 13]. <https://doi.org/10.48550/arXiv.2307.09288>.
- 11 ACHIAM J, ADLER S, AGARWAL S, et al. GPT-4 technical report [EB/OL]. [2025 - 08 - 13]. <https://doi.org/10.48550/arXiv.2303.08774>.
- 12 LIU Z, LIN Y, CAO Y, et al. Swin Transformer: hierarchical vision transformer using shifted windows [C]. Montreal: *The IEEE/CVF International Conference on Computer Vision*, 2021.
- 13 ZHANG K, MENG X, YAN X, et al. Revolutionizing health care: the transformative impact of large language models in medicine [J]. *Journal of medical internet research*, 2025, 27 (1): e59069.
- 14 KIRK R, MEDIRATTA I, NALMPANTIS C, et al. Understanding the effects of RLHF on LLM generalisation and diversity [EB/OL]. [2025 - 08 - 13]. <https://doi.org/10.48550/arXiv.2310.06452>
- 15 LIU H, LI C, WU Q, et al. Visual instruction tuning [C]. New Orleans: *The 37th Conference on Neural Information Processing Systems (NeurIPS 2023)*, 2023.
- 16 WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models [C]. New Orleans: *The 36th Conference on Neural Information Processing Systems (NeurIPS)*, 2022.
- 17 GOU J, YU B, MAYBANK S J, et al. Knowledge distillation: a survey [J]. *International journal of computer vision*, 2021, 129 (6): 1789 - 1819.
- 18 PAL A, UMAPATHI L K, SANKARASUBBU M. Medmcqa: a large-scale multi-subject multi-choice dataset for medical domain question answering [EB/OL]. [2025 - 08 - 13]. <https://proceedings.mlr.press/v174/pal22a/pal22a.pdf>.
- 19 JIN Q, DHINGRA B, LIU Z, et al. PubMedQA: a dataset for biomedical research question answering [C]. Hong Kong: *The 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP - IJCNLP)*, 2019.
- 20 WANG Y, MA X, ZHANG G, et al. Mmlu-pro: a more robust and challenging multi-task language understanding benchmark [C]. Vancouver: *The 38th Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2024.
- 21 SINGHAL K, TU T, GOTTWEIS J, et al. Towards expert-level medical question answering with large language models [EB/OL]. [2025 - 08 - 13]. <https://doi.org/10.48550/arXiv.2305.09617>.
- 22 TEAM G, ANIL R, BORGEAUD S, et al. Gemini: a family of highly capable multimodal models [EB/OL]. [2025 - 08 - 13]. <https://doi.org/10.48550/arXiv.2312.11805>.
- 23 XU P, WU Y, JIN K, et al. Deepseek-r1 outperforms Gemini 2.0 pro, OpenAI o1, and o3-mini in bilingual complex ophthalmology reasoning [J]. *Advances in oph-*

- thalmology practice and research, 2025, 5 (3): 189 – 195
- 24 CHEN Z, WANG W, CAO Y, et al. Expanding performance boundaries of open – source multimodal models with model, data, and test – time scaling [EB/OL]. [2025 – 08 – 13]. <https://doi.org/10.48550/arXiv.2412.05271>.
- 25 曹建峰, 徐艳玲. 医疗领域多模态 AI 模型的机遇、挑战与治理应对 [J]. 中国医学伦理学, 2024, 37 (9): 1023 – 1029.
- 26 李佳林, 郜斌宇, 陈卉. 基于大语言模型的电子病历文本结构化 [J]. 中国医疗设备, 2025, 40 (5): 42 – 46, 52.
- 27 王天笑, 于佳婕, 钱琨. 基于医疗大模型的患者咨询问答安全及伦理管控 [J]. 中国卫生信息管理杂志, 2025, 22 (2): 202 – 207.
- 28 方俊泽, 高怀婷, 邢素霞, 等. 人工智能胸部疾病辅助诊断系统在基层医疗机构全科医生诊断中的实效性评估 [J]. 中国全科医学, 2025, 28 (31): 3948 – 3953.
- 29 王亚楠, 蒋知轩, 何昊洋, 等. 大语言模型发展历程及其在临床医学教育中的应用 [J]. 华西医学, 2025, 40 (5): 777 – 782.
- 30 孙连伟, 连雪全, 杨佳林, 等. 图神经网络在药物 – 靶标预测领域的研究进展 [J]. 医学信息, 2025, 38 (8): 167 – 171.
- 31 王燕婷, 蒋雪炳, 陈方, 等. 虚拟数字人联合移动式智能采血系统改善患者就医体验 [J]. 中国卫生质量管理, 2025, 32 (4): 23 – 26, 36.
- 32 WANG B, LAI J, CAO H, et al. Enhancing the interoperability and transparency of real – world data extraction in clinical research; evaluating the feasibility and impact of a ChatGLM implementation in Chinese hospital settings [J]. European heart journal – digital health, 2024, 5 (6): 712 – 724.
- 33 WAN P, HUANG Z, TANG W, et al. Outpatient reception via collaboration between nurses and a large language model: a randomized controlled trial [J]. Nature medicine, 2024, 30 (10): 2878 – 2885.
- 34 GARCIA P, MA S P, SHAH S, et al. Artificial intelligence – generated draft replies to patient inbox messages [J]. JAMA network open, 2024, 7 (3): e24320.
- 35 LAI T, SHI Y, DU Z, et al. Supporting the demand on mental health services with AI – based conversational large language models (LLMs) [J]. BioMedInformatics, 2023, 4 (1): 8 – 33.
- 36 CHEN Y, WANG Z, XING X, et al. BianQue: balancing the questioning and suggestion ability of health LLMs with multi – turn health conversations polished by ChatGPT [EB/OL]. [2025 – 08 – 13]. <https://doi.org/10.48550/arXiv.2310.15896>.
- 37 LI Y, LI Z, 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.
- 38 WANG G, YANG G, DU Z, et al. ClinicalGPT: large language models finetuned with diverse medical data and comprehensive evaluation [EB/OL]. [2025 – 08 – 13]. <https://doi.org/10.48550/arXiv.2306.09968>.
- 39 WU C, LIN W, ZHANG X, et al. PMC – LLaMA: toward building open – source language models for medicine [J]. Journal of the American medical informatics association, 2024, 31 (9): 1833 – 1843.
- 40 LI C, WONG C, ZHANG S, et al. LLaVA – Med: training a large language – and – vision assistant for biomedicine in one day [EB/OL]. [2025 – 08 – 13]. <https://neurips.cc/virtual/2023/poster/73643>.
- 41 SINGHAL K, AZIZI S, TU T, et al. Large language models encode clinical knowledge [J]. Nature, 2023, 620 (7972): 172 – 180.
- 42 LUO Y, ZHANG J, FAN S, et al. BioMedGPT: an open multimodal large language model for biomedicine [J]. IEEE journal of biomedical and health informatics, 2024 (11): 1 – 14.
- 43 BOLTON E, VENIGALLA A, YASUNAGA M, et al. BioMedLM: a 2.7 B parameter language model trained on biomedical text [EB/OL]. [2025 – 08 – 13]. <https://doi.org/10.48550/arXiv.2403.18421>.
- 44 XIE Q, CHEN Q, CHEN A, et al. Me – LLaMA: foundation large language models for medical applications [J]. NPJ digital medicine, 2025, 8 (1): 141.
- 45 LI J, GUAN Z, WANG J, et al. Integrated image – based deep learning and language models for primary diabetes care [J]. Nature medicine, 2024, 30 (10): 2886 – 2896.
- 46 MUKHERJEE S, GAMBLE P, AUSIN M S, et al. Polaris: a safety – focused LLM constellation architecture for health-care [EB/OL]. [2025 – 08 – 13]. <https://doi.org/10.48550/arXiv.2403.13313>.
- 47 GARCIA – FERRERO I, AGERRI R, CABRIO E, et al. Medical mT5: an open – source multilingual text – to – text LLM for the medical domain [C]. Torino: The 2024 Joint International Conference on Computational Linguistics, Language Resources

- and Evaluation (LREC – COLING 2024), 2024.
- 48 YANG S, ZHAO H, ZHU S, et al. Zhongjing: enhancing the Chinese medical capabilities of large language model through expert feedback and real – world multi – turn dialogue [C]. Vancouver: The AAAI Conference on Artificial Intelligence, 2024.
- 49 ZHOU Z, YANG Y, REN T. IHILLM – RAG: a safe and private medical large language model based on intelligent hardware interaction and retrieval – augmented generation (RAG) [C]. Xiamen: Fourth International Computational Imaging Conference (CITA 2024), 2024.
- 50 LIU C, WANG H, PAN J, et al. Beyond distillation: pushing the limits of medical LLM reasoning with minimalist rule – based RL [EB/OL]. [2025 – 08 – 13]. <https://doi.org/10.48550/arXiv.2505.17952>.
- 51 WU J, DENG W, LI X, et al. MedReason: eliciting factual medical reasoning steps in LLMs via knowledge graphs [EB/OL]. [2025 – 08 – 13]. <https://doi.org/10.48550/arXiv.2504.00993>.
- 52 LIN T, ZHANG W, LI S, et al. HealthGPT: a medical large vision – language model for unifying comprehension and generation via heterogeneous knowledge adaptation [EB/OL]. [2025 – 08 – 13]. <https://doi.org/10.48550/arXiv.2502.09838>.
- 53 YU H, CHENG T, CHENG Y, et al. FineMedLM – o1: enhancing the medical reasoning ability of LLM from supervised fine – tuning to test – time training [EB/OL]. [2025 – 08 – 13]. <https://doi.org/10.48550/arXiv.2501.09213>.
- 54 Google. MedGemma [EB/OL]. [2025 – 08 – 13]. <https://deepmind.google/models/gemma/medgemma/>.
- 55 BUSCH F, HOFFMANN L, RUEGER C, et al. Current applications and challenges in large language models for patient care: a systematic review [J]. *Communications medicine*, 2025, 5 (1): 26.
- 56 ZHANG B, BORNET A, YAZDANI A, et al. A dataset for evaluating clinical research claims in large language models [J]. *Scientific data*, 2025, 12 (1): 86.
- 57 BRAUNECK A, SCHMALHORST L, KAZEMI MAJDABADI M M, et al. Federated machine learning, privacy – enhancing technologies, and data protection laws in medical research: scoping review [J]. *Journal of medical internet research*, 2023, 25 (3): e41588.
- 58 OMAR M, SOFFER S, AGBAREIA R, et al. Sociodemographic biases in medical decision making by large language models [J]. *Nature medicine*, 2025, 31 (6): 1 – 9.
- 59 JI Y, MA W, SIVARAJKUMAR S, et al. Mitigating the risk of health inequity exacerbated by large language models [J]. *NPJ digital medicine*, 2025, 8 (1): 246.

(上接第 8 页)

- 35 SUN Q F, XIA F, LONG Y T, et al. Research on digital technology empowering the modernization of China’s medical governance [J]. *Bulletin of the Chinese academy of sciences*, 2022, 37 (12): 1483 – 1492.
- 36 WELLNER G, MYKHAILOV D. Caring in an algorithmic world: ethical perspectives for designers and developers in building AI algorithms to fight fake news [J]. *Science and engineering ethics*, 2023, 29 (4): 38.
- 37 Nature Machine Intelligence. Striving for health equity with machine learning [J]. *Nature machine intelligence*, 2021, 3 (8): 653.
- 38 ZHANG J, ZHANG Z M. Ethics and governance of trustworthy medical artificial intelligence [J]. *BMC medical informatics and decision making*, 2023, 23 (1): 1 – 15.
- 39 YOU M, XIAO Y, YAO H, et al. Evaluation and regulation of medical artificial intelligence applications in China [J]. *Chinese medical sciences journal*, 2025, 40 (1): 3 – 8.
- 40 MACHADO H, SILVA S, NEIVA L. Publics’ views on ethical challenges of artificial intelligence: a scoping review [J]. *AI and ethics*, 2025, 5 (1): 139 – 167.
- 41 AVINASH A, HARSH A, NIHAARIKA A. Fairness score and process standardization: framework for fairness certification in artificial intelligence systems [J]. *AI and ethics*, 2023, 3 (4): 1143 – 1162.
- 42 许卫卫, 高明, 吉萍. 跨机构, 多学科合作科研项目伦理审查问题和对策 [J]. *医学与哲学*, 2025, 46 (10): 67 – 70.
- 43 WEINER E B, DANKWA – MULLAN I, NELSON W A, et al. Ethical challenges and evolving strategies in the integration of artificial intelligence into clinical practice [J]. *PLOS digital health*, 2025, 4 (4): e0001598.