

大语言模型在中医领域使用的技术及研究应用

曾玉秀, 赵琼, 席崇程*, 冯全生*
(成都中医药大学, 成都 611137)

[摘要] 大语言模型(LLM)和中医药结合,推动了中医药信息化发展,也为新时代中医药的传承创新发展提供了新方向。在LLM研究背景下,概述基于Transformer架构的LLM发展进程,梳理中医药LLM的研究进展。总结构建中医药LLM的主要过程及中医药LLM开发过程中研究者使用的技术。根据相关的研究文献,总结开发的中医药LLM主要应用场景及中医药研究者使用LLM进行的研究探索。同时,分析当前的中医药LLM开发所面临的挑战,在高质量数据的构建、中医药LLM评估的方法、模型的可解释性、中医药LLM多模态融合、中医处方推荐模型开发等方面亟待进一步提升。展望中医药LLM的未来发展,期望能为LLM与中医药的进一步结合提供参考,助力中医药现代化进程发展。

[关键词] 大语言模型; 中医药; 人工智能; 垂直领域大模型

[中图分类号] R856.2;R241;TP18 **[文献标识码]** A **[文章编号]** 1005-9903(2026)14-0050-10

[doi] 10.13422/j.cnki.syfjx.20251011

[网络出版地址] <https://link.cnki.net/urlid/11.3495.r.20250827.1053.001>

[网络出版日期] 2025-08-27 14:16:49 **[增强出版附件]** 内容详见 <http://www.syfjxzz.com> 或 <http://cnki.net>



Technologies and Research Applications of Large Language Models in Traditional Chinese Medicine

ZENG Yuxiu, ZHAO Qiong, XI Chongcheng*, FENG Quansheng*
(Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China)

[Abstract] The integration of large language model (LLM) and traditional Chinese medicine (TCM) promotes the informatization of TCM, and also provides a new direction for the inheritance and innovation of TCM in the new era. Based on the research background of LLM, the development process of LLM based on the Transformer architecture is summarized, and the research progress of LLM in TCM is reviewed. The main process of constructing TCM LLMs and the key techniques used by researchers during model development are summarized. Based on the related literature, the main application scenarios of TCM LLMs and the research explorations conducted by TCM researchers using LLMs are outlined. Meanwhile, the current challenges faced in the development of TCM LLMs are analyzed. Further improvements are urgently needed in the construction of high-quality data, the evaluation methodology of TCM LLMs, the interpretability of the model, multimodal fusion of TCM LLMs, and the development of TCM prescription recommendation models. Looking forward to the future development of LLM in TCM, it is expected to provide a reference for the deeper integration of LLMs and TCM, and facilitate the modernization of TCM.

[Keywords] large language model; traditional Chinese medicine; artificial intelligence; domain-specific large language model

四川古称巴蜀,中医药在四川地区,历经长时间的沉淀,形成了兼容并蓄,学术特色纷呈的川派中医^[1]。自古以来巴蜀地区名医辈出,四川也有“中医之乡”和“中药之库”之称;其复杂的地形地貌,使其拥有丰富的中药资源,生长的道地药材品种多,产量大^[1]。川派中医是中医药的重要组成部分

分,在疾病的治疗上有其特色,发挥着重要作用。

“互联网+”时代的到来,新技术与医疗行业的结合更加紧密,给中医药的发展带来了机遇。《“十四五”中医药信息发展规划》指出要以大数据、人工智能等新一代信息技术创新引领中医药信息化的高质量发展^[2]。当前,人工智能技术广

[收稿日期] 2025-06-24

[基金项目] 四川省哲学社会科学基金项目(SCJJ24ND210)

[第一作者] 曾玉秀,在读硕士,从事中医温病学与中医药信息研究,E-mail:zengyuxiu@cduetcm.edu.cn

[通信作者] *席崇程,博士,讲师,从事中医经典理论研究,E-mail:xichongcheng@bucm.edu.cn;

*冯全生,博士生导师,教授,从事中医温病与感染病学研究,E-mail:fengqs118@163.com

泛应用于辅助中医的诊断和治疗,中医药的数据挖掘,健康管理,辅助筛选中药的潜在作用靶点,解释中药的作用机制,研制新药等方面^[3-4]。近年来,生成式人工智能(AIGC)的出现,极大地推动了人工智能技术从以分析为主向以生成为主的转变。

自OpenAI的聊天生成型预训练变换模型(ChatGPT)发布以来,生成式大语言模型(GLLM)因其强大的自然语言处理和生成能力受到广泛关注,涌现出了OpenAI o4-mini推理模型(O4-mini)、Meta AI大语言模型(LLaMA)、开放科学开放获取多语言模型(BLOOM)、TII猎鹰大模型(Falcon)、LMSYS维库纳大模型(Vicuna)等通用大语言模型(LLM)。LLM具有的良好自然语言处理和涌现能力,能提高中医药LLM在不同任务中的处理和生成能力,为中医药的临床、科研及发展提供强有力的支撑^[5]。本文旨在总结当前中医药LLM的研究进展及开发过程中使用的关键技术,总结LLM在中医药领域的研究应用,为未来LLM和中医药结合提供参考。

1 LLM的研究进展

1.1 LLM的发展 2017年提出的Transformer架构,使用自注意力机制,在语义特征的提取上超越了之前流行的卷积神经网络等深度学习模型,能提取自然语言文本中更加丰富的语义信息^[6]。随后出现了以Transformer架构为基础的基于转换器的双向编码表征(BERT)等预训练语言模型(PLM)。PLM在大规模语料上学习,经过微调适应下游任务。随着技术的不断发展,模型参数量不断增大,LLM应运而生。大多数LLM以Transformer架构为核心,包含数十亿甚至更多参数,在海量文本数据中进行训练,表现出了强大的自然语言处理和生成能力^[7-8]。基于Transformer的LLM在发展过程中,出现了仅编码器(Encoder-only)架构、仅解码器(Decoder-only)架构和同时包含编码器和解码器(Encoder-Decoder)的架构^[9]。

随着参数量的不断变大,研究表明当LLM的参数量超过一定的阈值,模型的性能将得到较大的提升,并将这种能力称为涌现能力(Emergent Ability)^[10]。LLM因参数量的增大而涌现的能力包括①上下文学习(ICL),LLM可以根据自然语言指令或者给出的范例学习特定的任务,无须再进行额外的训练^[7,11];②指令遵循(Instruction Following),通过指令微调的形式,提高模型的泛化能力^[7];③逐步推理(Multi-Step Reasoning),LLM使用思维链(CoT)策略,利用提示能完成多步推理^[12]。

近年来,LLM如雨后春笋般涌现。虽然LLM在通用的多个领域中展现了强大的功能,但模型的有些输出,会背离用户的意图,或者产生与事实不符合的内容^[13-14]。而且在专业领域中,由于LLM缺乏足够深度的专业知识,模型可能无法提供正确的答案^[15-16]。目前LLM可以通过①检索增强生成(RAG),即在LLM模型回答问题之前从外部知识库中检索相关信息^[16];②监督微调(SFT),来自人类反馈的强化学习(RLHF)等方式提高模型在专业领域的性能。此外,为了解决LLM在专业领域的不足,出现了为特定领域而构建的

垂直领域LLM,如金融领域的彭博生成式预训练变换器(BLOOMbergGPT),法律领域的哈佛法律大模型(Harvey),医学领域的医生生成式语言模型(DoctorGLM)、中文医学生成式预训练变换器(ChiMed-GPT)、阿尔帕医疗大模型(AlpaCare)、太乙医学大模型(Taiyi)、百川医疗模型1号(Baichuan-M1)等。

1.2 中医药LLM情况 LLM和中医药结合处于探索阶段,但相关的研究正在快速地发展。本文通过检索中国知网(CNKI)、万方数据库(Wangfang Data)、维普数据库(VIP)、谷歌学术网站、Web of Science数据库收集关于中医药和LLM自建库到2025年7月的相关资料。使用“中医药”和“大语言模型”作为检索关键词,并纳入这些文章中涉及中医药LLM的相关参考文献进行总结分析。中医药LLM相关的模型情况见图1。

2 构建中医药LLM的过程及关键技术

2.1 构建中医药LLM的过程 为了让LLM适用于中医药领域或提高模型的性能,研究者们构建中医药语料库和指令,选用经过训练且性能优异的LLM(称为基座模型)继续训练,微调优化具有中医药知识的LLM,过程见增强出版附加材料。LLM训练主要的步骤为预训练、微调、奖励建模(RM)、强化学习(RL)^[17]。

2.1.1 预训练阶段 目前开发的中医药LLM主要使用的数据包括收集的文本数据如中医古籍、现代中医教科书、医学文献和硕博学位论文、临床指南、专家共识等;自主构建或引入开源的中医药知识图谱;基于真实世界的医患对话数据和临床数据;研究者们构建并已经公开的中医药领域数据,如阿里天池的中医文献数据集、中医辨证数据集(TCM-SD)等。在通用LLM中使用上述数据进行训练,使模型拥有专业的中医药知识。通常选用的基座模型有LLaMA、智谱清言(ChatGLM)、通义千问(Qwen)、百川LLM(Baichuan)、BLOOM等。随着参数量的增大,通常模型的效果会更好,在条件允许下,可以选择参数量更大的模型作为基座模型。

2.1.2 微调阶段 微调是模型进行预训练后,在特定的数据上进行精细调整,使其能更好地适应下游任务^[18]。通常使用小规模、有标注的数据进行训练,以提升模型在特定任务上的性能^[19]。在当前的中医药LLM中,主要使用的微调技术包括高效参数微调(PEFT)、提示微调(Prompt Tuning)、强化学习微调等^[18]。LLM参数量巨大,PEFT只微调模型中少量的参数,但能达到全参微调的效果,主要的方法包括适配器微调(Adapter)、低秩适配(LoRA)、前缀调优(Prefix-Tuning)、提示微调(P-Tuning)等^[18]。提示微调通过构建提示模板,引导模型生成符合预期的输出,在单样本学习(one-shot)、少样本学习(few-shot)等方法后,出现了上下文学习和思维链的方法^[18]。指令微调也是比较受欢迎的微调方法,和提示微调相似,却不同于提示微调。指令微调通过人工创建或使用LLM根据给定的提示得到指令数据集,通常结合PEFT对模型进行微调^[20]。强化学习微调方法主要为基于人类反馈的强化学习(RLHF)和基于人工智能反馈的强化学习(RLAIF)等。

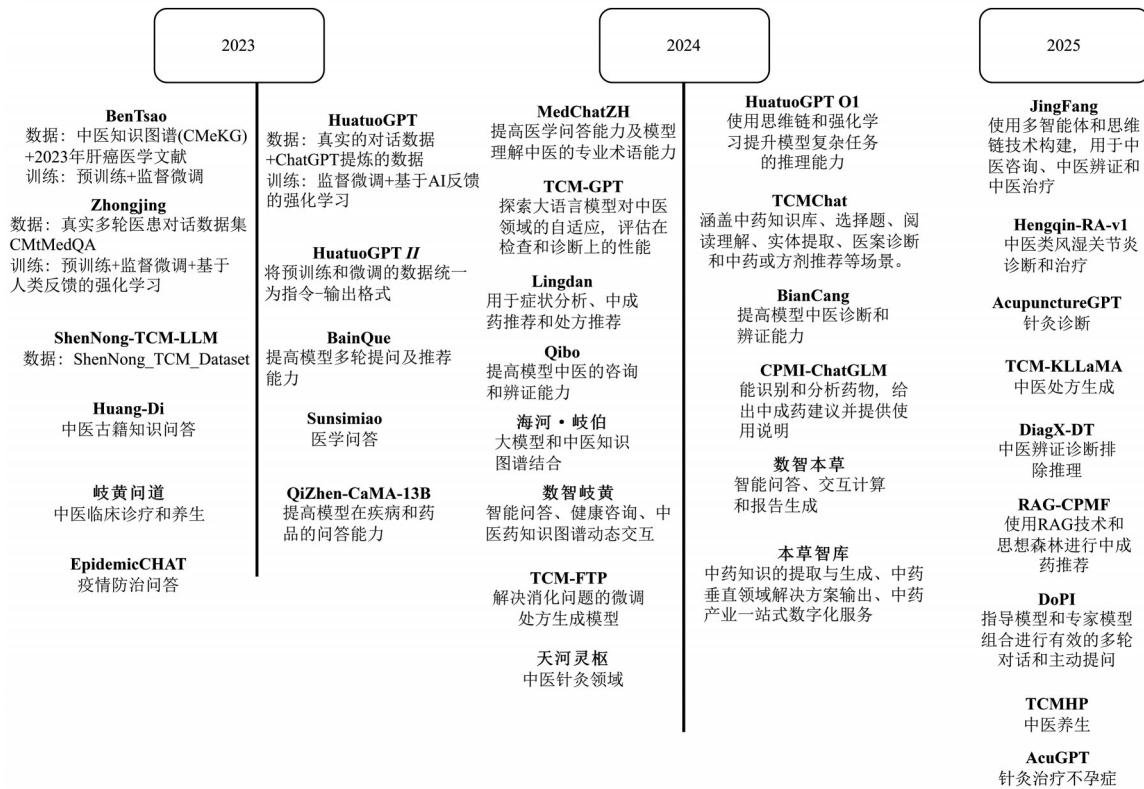


图 1 中医药 LLM 情况

Fig. 1 Case of LLM of traditional Chinese medicine

在预训练模型的基础上,使用有标注的监督微调数据和各种监督微调技术,形成监督微调模型,使其能进行中医的辨证论治、处方推荐等。通过奖励建模和强化学习,使监督微调模型能输出更符合人类意图的结果,形成中医药 LLM。

2.1.3 模型评估阶段 模型的评估有助于了解其在特定任务上的性能,发现模型的不足之处,进而优化模型^[21]。当前主要从模型的性能、鲁棒性、对齐等角度进行评估^[21]。在中医药领域一般使用评估指标或者专家评估等方法,评估 LLM 在中医领域各个场景的能力,优化中医药 LLM。在中医药 LLM 的评估时,需要从多个维度评估模型的性能;在不同的场景中,选择相适应的指标进行评估,以确保评价的准确性^[5]。

2.2 使用的技术

2.2.1 混合专家 (MoE) 和多智能体系统 (Multi-Agent System) MoE 是 Transformer 架构的一种变体,见增强出版附加材料。研究表明在大规模数据的情况下,使用 MoE 和指令调优,模型的效果会更好;而且,在一定范围内,随着集成专家数量的增加,模型得益于能处理不同任务的丰富子网络,在处理复杂任务时表现出色^[22]。在中医药领域,特别是在临床上,面临的情况更加复杂,使用 MoE 可以集成多个专家 (Experts),集合专家的长处,适应中医药不同场景下的应用,让模型能处理复杂不确定的问题,输出的内容更准确^[5]。MoE 通过门控网络 (Gating Network) 激活相应的专家进行训练,使用较少的计算开销,提高了模型的训练和推理的效率^[23-24]。如基于 MoE 框架开发的明医大模型 (MING-MoE)

在多个医疗任务中表现出色^[25]。在中医药领域,使用 MoE 可能因为负载的不平衡,导致某些专家频繁地用到,而一些专家未被充分利用;而且 MoE 模型决策的复杂性使其可解释性难度加大,影响用户对模型的信任度;LLM 融合 MoE 使得模型训练时并行策略变得更加复杂^[26]。与之相似的多智能体系统由多个独立的 Agent 组成,在处理复杂任务时,将其分解成更小的子任务,不同的子任务委派给不同的专业 Agents,智能体之间通过协作实现给定目标^[27]。如 JingFang LLM 中集成了中医全科的 Agent 及中医不同专科的 Agents,模型根据患者的描述,激活相应的专家。

2.2.2 提示工程 提示工程能为模型提供有效的提示,引导模型生成符合要求的回答^[28]。精心设计的 Prompt 能引导模型,使其在中医辨证或知识问答等情景下能生成相关的、准确的并且质量较高的回答^[29]。此外,为了提高模型在复杂任务中的推理能力,提出了思维链 (CoT) 技术。使用 CoT 引导模型进行推理可以显著提高 LLM 执行复杂任务的推理能力^[30]。在中医药领域,思维链技术适合用于模拟名中医诊治疾病的过程,中医古籍知识的推理及推测古籍中由于年代久远而缺失的未知字^[30-31]。虽然提示工程能提高模型的推理能力,但人工构建提示词 (Prompt) 成本较高,而且依赖于设计者的经验,不同的提示,即使是同一个问题,模型生成的回答也不同。提示工程组件图见增强出版附加材料。

2.2.3 检索增强生成技术 在垂直领域,LLM 由于预训练的数据集未包含特定领域的专业知识,可能会产生幻觉现象。在中医药领域,检索增强生成 (RAG) 技术将信息检索

和生成模型相结合,使其在有限语料库的基础上,能根据用户输入的信息从外部知识库(知识库中包含海量的中医药文本及临床数据等)中检索信息,并将这些信息作为提示放入LLM中作为上下文输入,从而指导模型输出更高质量的答案,提高LLM在中医辨证、处方推荐及建议上的准确性^[32-33]。然而RAG很大程度依赖于外部知识库,同时需要更多的计算资源,这可能会影响在实际应用中的实时性^[34]。在检索增强生成方面除RAG技术外,目前常用的还有图检索增强生成(GraphRAG)和知识增强生成(KAG)技术。GraphRAG在传统RAG的基础上引入知识图谱,支持图检索;KAG支持知识和文本互索引,通过逻辑方式引导模型检索和推理^[35]。检索增强生成原理图见增强出版附加材料。

2.2.4 RLHF RLHF是一种微调LLM的技术,使模型的输出更加符合人类意图。其核心在于通过训练奖励模型,与用户交互得到输出结果的人工打分,使用强化学习算法优化监督微调模型的参数,生成符合人类意图的文本^[18]。在中医药领域,使用RLHF技术可以将名老中医的经验和用户的反馈融入LLM,使LLM的输出更贴合医疗实践,提高模型在中医知识问答、处方推荐等方面的准确性^[5]。但RLHF面临着一些挑战。基于人工的打分具有主观性,而且提供高质量的反馈仍然有些困难^[36]。通常情况下,在训练奖励模型时,提供的数据数量越大效果越好,但达到这种需求所需的人工成本较高^[37]。见增强出版附加材料。

2.2.5 RLAIIF 基于RLHF方法微调,通常需要大量的人类反馈标签,花费的成本高,耗费的时间长,因此研究者寻求不依赖于人来监督AI的行为——RLAIIF^[18]。RLAIIF使用通用LLM对结果进行评分,得到评分偏好分布,进行强化学习。研究表明RLAIIF能与RLHF所能达到的性能相媲美,RLAIIF可以作为替代RLHF的一种方案^[38]。而且RLAIIF所需的注释周期短、使用成本更低。

2.2.6 知识蒸馏(KD) KD是一种模型压缩范式,能将知识从复杂模型(教师模型)转移到更简单的对应模型(学生模型),可以将性能较好的闭源LLM转移到开源的LLM中^[39-40]。当前,KD技术广泛应用于医疗领域的临床决策支持,医学总结、知识问答、药物发现、药物推荐(如LEADER)等^[41-42]。在中医药领域,使用KD技术训练学生模型,可以解决数据的隐私问题,可以获得良好的泛化能力,使小模型也能达到近似或优于复杂模型的效果^[43]。中医药领域有庞杂的数据,按传统的LLM方法进行训练,需要较大的计算资源支持,而KD可以将教师模型的中医药知识和中医师的临床经验转移到学生模型,有助于在有限计算资源下部署性能较好的LLM,加速模型的训练及提供更快、更准确的知识回答,精准中医的辨证论治等。但是KD也存在着不足,如通过知识蒸馏后,学生模型的输出可能让人无法理解,而且目前知识蒸馏需要多少数据才能达到较好的效果仍是一个问题。知识蒸馏图见增强出版附加材料。

3 LLM在中医领域的应用

3.1 开发的中医药LLM的应用 用于医学知识的问答,并提供个性化的建议。通用LLM具有出色的问答能力。在此

基础上,通过训练大规模中医数据及医患对话数据,实现中医知识问答,并给出个性化的建议。在数千年中,留下了不少宝贵的中医古籍,为了让古籍在新时代也能焕发出生命力,研究者开发了以《中华医典》为主要数据的Huang-Di LLM^[44]和以历代《伤寒论》为主要数据的中医药古籍LLM,用于古籍知识的问答^[45]。

用于辨证论治。在开发的中医药LLM中,通过设置高质量的prompt,引入强化学习技术或融合中医专家反馈等方法,让模型学习诊疗的思维和辨证规律,进一步提高辨证诊断的性能和准确性^[46]。当前许多研究者正在尝试着构建更加适合于中医辨证诊断的LLM,以应用于临床,如扁仓(BianCang)、诊断X-数字孪生(DiagX-DT)等。与此同时,也出现了更细化的中医药LLM开发,如用于中医类风湿关节炎诊断和诊疗的大模型(第一版)(Hengqin-RA-v1)和用于针灸诊断的LLM(AcupunctureGPT)等。除了基于文本数据进行开发外,也出现了融合患者舌诊图片进行中医药LLM开发的研究^[47]。

目前开发的中医药LLM能根据患者的叙述,分析症状,推荐中成药或中药处方,如图神经网络(GNN)-LLM^[48]、检索增强生成-多模态融合跨模态预训练(RAG-CPMF)^[49]和Lingdan^[46]等。此外“数智本草”等LLM可以用来研究中药复方成分对应的靶点,揭示中药复方的药理机制;可以对中药化学成分进行预测和分析,有望能更好地优化方剂和开发新药。在中药处方推荐上,中医药大语言模型-处方推荐(TCM-LLM-PR)^[50]和中医药知识图谱增强LLaMA模型(TCM-KLLaMA)^[51]都使用精确度、召回率、F1分数优化模型;TCM-KLLaMA相较于其他推荐模型,除了关注症状-药物外,通过构建包含症状、舌像、脉象等信息的知识图谱增强LLM的能力,通过症状互知融合机制(SMKI)增强处理同义症状;出现了基于RAG技术优化的MoE LLM进行处方推荐^[52]。中成药推荐模型中,中成药指令对话通用语言模型(CPMI-ChatGLM)使用双语互译质量评估指标(BLEU)分数、面向摘要评估的召回率评测指标(ROUGE)、基于BERT的自动评分指标(BARTScores)评估模型;而RAG-CPMF根据临床指南构建标准数据集进行评估,随后评估模型在临床应用时的安全性、可用性、平滑度^[49,53]。

3.2 使用LLM进行中医药领域的探索

3.2.1 命名实体的识别与分析及在知识图谱方面的应用

在中医药领域,出现了使用LLM进行医案命名实体识别和分析的研究,如设计提示词,调用LLM,进行医案实体的抽取^[54];利用ChatGPT智能进行宋朝医疗处方识别和病例数据分析^[55]。研究者们基于LLM进行知识图谱的构建和增强知识问答方面的探索。TRAJANOSKA等^[56]发现LLM可以增强知识图谱的构建。研究者们尝试使用LLM增强中医知识图谱的构建,如中医药风湿病诊断-知识蒸馏模型(TCMRD-KD)使用ChatGPT 4.0辅助构建中医古籍中的风湿病知识图谱^[57];李玥等^[58]用LLM抽取国医大师治疗鼻炎的医案构建鼻炎知识图谱;也有使用LLM基于CNKI的文献构建中医妇科的知识图谱的研究^[59]。何宇浩等^[60]的研究中

比较了不同 LLM 按照“病-症-方-药”自动抽取方剂中的实体和关系创建知识图谱的能力,他们发现效果较好的模型构建的知识图谱更好。

在垂直领域,为了能让模型具有适应特定领域的的能力,研究者们提倡将外部知识引入 LLM 中,对知识生成式预训练大模型(KnowledGPT)^[61]框架、图上思维(ToG)^[62]技术和最小映射模型(MinMap)^[63]等进行探索。在中医药领域也探索了知识图谱和 LLM 结合的方法,如疫病通用语言模型(YibingGLM)^[64]将知识图谱和 LLM 结合;董兆安等^[65]用 LLM 对《千金要方》进行知识抽取,构建中医药知识图谱,并使用 GraphRAG 技术将知识图谱的中医药知识融入,开放中医药(大模型/知识系统)(OpenTCM)使用了同样的技术在更大规模的数据上实践,增强 LLM 对中医药知识的理解,以减轻 LLM 出现的幻觉现象,提高知识检索和诊断问答的能力^[66]。ChatTCM^[67]融合思维链和知识图谱提高模型对中医药知识理解和问答的能力。

3.2.2 构建中医药知识问答系统及中药方面的研究 在知识问答方面,研究者们使用 LLM 结合检索增强生成技术构建智能化的中医药知识问答系统^[68]。在中药上,使用 LLM 增强中药药物间的相互作用预测及进行药物-靶点相互作用预测等^[69-70]。基于中药处方推荐模型(PresRecRF)整合分子知识到中药-症状知识图谱中,用 LLM 增强学习中医理论,根据患者情况推荐中药处方及给出各药物的剂量^[71]。研究使用 LLM 预测药物毒性和进行方剂分类^[72-73]。开发的神农用药指导系统(ShennongMGS)用于药物指导和不良反应的预测^[74]。

3.2.3 评估 LLM 的专业能力 在西医和通用的 LLM 中出现了专业的评估基准,而中医药领域的 LLM 评估基准尚未涵盖,因而研究者们致力于中医药 LLM 评估基准的构建及评估中医药 LLM 能力的研究。受使用标准考试以进行客观化评价 LLM 的影响,出现的中医药大模型综合评测基准(TCMBench)^[75]和中医药问答数据集(TCMD)^[76]等数据集使用中医执业医师考试题目作为基准,评估中医药 LLM 的能力。研究者们发现模型在基准上的评价不尽人意且临床的情况复杂,单从标准考试的题目评价还不够。因而探索构建中医药三维评测基准(TCM-3CEval)数据集^[77],从核心知识的掌握、经典文献的理解,临床决策 3 个维度评估中医药 LLM,研究表明当前的模型与实际的中医临床之间仍存在差距^[77]。辨证是中医重要的环节,之前构建的用于辨证的 TCM-SD 缺少辨证推理过程,新构建的中医辨证思维评测基准数据集(TCMEval-SDT)^[78]包含辨证推理过程,用于评估中医临床诊断能力。当前 TCMEval-SDT 包含的疾病类型尚不完善。中医药大模型多任务评测基准框架(MTCMB)^[79]数据集除了关注模型的中医知识、疾病的辨证推理,还关注模型语言理解能力、安全性、处方推荐能力。中医药多模态问答评测基准(TCM-Ladder)除了高质量的问答数据及诊断数据外,还包含了中药药材图片、舌像图片、音频、视频,用于评估中医多模态模型的能力^[80]。当前用于评估的数据集越来越丰富,但在中医辨证推理、中药处方推荐及多模态的数

据上仍有待加强。

研究者们通过构建中医测评数据集,设置 prompt 引导模型进行推理,测评不同的中医药 LLM 能力。如张童等^[81]选用 9 个包含通用和医学领域的 LLM,采用自动评估和专家打分的方法,评估中医药 LLM 的中医药专业知识和临床能力。Li 等^[82]评估了不同的通用 LLM 之间在提取穴位位置关系方面的性能。

3.2.4 使用 LLM 进行中医处方数据的挖掘 数据挖掘是使用各种机器学习或者深度学习算法从大量的数据中提取有价值信息的过程^[83-84]。在中医药领域主要使用数据挖掘技术探析中医药辨治疾病用药规律。常用的算法包括关联规则、聚类分析、复杂网络等,主要的软件包括中医传承辅助平台、古今医案云平台、SPSS、R 语言集成开发环境(RStudio)等^[85]。

鉴于 LLM 拥有卓越的自然语言处理、理解和生成能力,中医领域数据量大,处理有一定的难度,研究者开始探索使用 LLM 进行中医处方数据挖掘。在一项研究中,使用中医治疗骨质疏松症的数据为例,构建自然语言理解能力更好的 LLM,通过人机交互提问的方式,让 LLM 对中医的处方数据自动化挖掘,统计方剂中药物的频率、进行聚类分析和关联规则分析,研究表明自动分析的结果和用 SPSS 统计的结果高度相似^[86]。

3.2.5 探索对抗中医药 LLM 的幻觉 所谓幻觉(hallucination)是 LLM 生成的违背事实或者无意义的文本,基于 LLM 生成的文本由于遵循最大似然原则,容易产生幻觉^[87]。研究者们发现产生幻觉的原因还包括训练时使用的海量数据中存在质量参差不齐的情况,或者训练数据的重复或缺失,数据标注转向由 LLM 辅助生成所带来的问题,模型结构设计,缺乏领域专业知识等^[87-88]。针对 LLM 出现的幻觉现象,研究者们尝试通过构建高质量数据集,使用结构化知识,优化模型结构的设计,或者设计自我检查结构,或者在预训练和微调时引入幻觉缓解策略,引入外部知识、思维链、RLHF、直接偏好优化(DPO)、多智能体等方法加以改进^[88]。

3.2.6 探索中西医相同概念的异同 自西医传入中国到现在,两者在临床上的结合变得更加紧密,共同发挥着较好的疗效。在信息技术不断发展的时代,使用 LLM 技术深入分析和对比这两种医学体系的差异,是一个值得探索的领域。当前,有研究者选定特定的中西医相同的概念,构建中医和西医的文本数据集,微调 LLaMA 模型构建中西医模型,对比分析相同概念之间的异同之处;或者使用多智能体和思维链技术将中医的术语映射到西医中,促进中西医之间的跨学科交流^[89]。

4 挑战与展望

4.1 训练模型的数据 在中医药领域存在着许多非结构化和半结构化数据,而且训练使用的数据庞杂,数据质量参差不齐,此外中医古籍与当代中医书籍又不尽相同,这些都会给训练带来一定的挑战。研究发现使用高质量的语料库进行训练,无须通过增加参数量或网络规模就可以提升性能,降低 LLM 的环境成本^[90]。当前已经构建了涵盖疾病、症状、

治疗方法和药物等信息的中文医疗问答数据集 Huatuo-26m^[91]。虽然中医药领域出现了神农中医药数据(Shennong-TCM)、用于辨证的TCM-SD、中文医学问答数据集(cMedQA2)等数据集,但在中医古籍、中医辨证推理、中药/针灸处方推荐推理方面高质量数据集的构建仍有待进一步加强。

4.2 模型的评估 当前中医药LLM的评价指标不一,用于评估中医药LLM性能的数据集,大多是基于中医药知识测试的角度进行的,与实际临床有差距。目前在不同的场景任务中,通过选择合适的指标进行评估,但这些评估指标在不同的模型中使用又有差异,导致不同中医药LLM之间性能的比较及推动模型间的改进存在一定的阻碍。需要规范或探索更适合中医药领域LLM不同场景的评估方法,或给出相关评估的指导建议用以评价和对比中医药LLM的能力。

4.3 模型的可解释性 中医药LLM的可解释性不足。由于LLM内部机制不清楚,而且其参数量庞大及训练数据量大等,都给模型的可解释性带来挑战,因而人们不能完全信任其生成的知识,导致模型的使用率下降^[92]。对此,研究者提出融合知识图谱、思维链或者是引入更多的中医专业知识用以解释并证明了这些方法具有一定的可行性。未来随着技术不断提升,可以开发可解释性工具对中医药LLM预测的结果进行解释,或者可视化语言模型内部的工作原理等方法能提升模型的可解释性^[92]。

4.4 中医药多模态LLM有待进一步开发 中医药多模态LLM有待进一步的开发。当前大部分中医药LLM基于文本数据进行训练。在医学LLM发展中,出现了麒麟医疗视觉语言大模型(Qilin-Med-VL)融合医学影像及文本的多模态LLM,结合零样本对齐增强视觉语言医疗大模型3代(ZALM3)策略,使用眼科、皮肤科和中医科问答过程的图片和文本进行训练的多模态LLM,以及华佗视觉医疗多模态大模型(HuatuoGPT-Vision)、灵枢统一医疗多模态基础大模型(Lingshu)等模型。中医药领域也出现了基于舌诊图片进行多模态融合LLM的探索。在实际临床中,中医讲究望闻问切,四诊合参进行辨证论治,然而当前的中医药LLM及其结合舌诊图片进行的探索,大部分四诊获得的资料以文本的形式或者知识图谱的形式呈现,小部分舌诊以图片的形式呈现。因而未来,除了进一步发展可以将患者原始舌诊图像进行分析的LLM外,在舌像上,还可以采集舌像的动态图,进一步分析患者舌诊情况;问诊和闻诊中,除了现在的文字形式外,可以开发能识别患者声音如咳嗽声、呼吸声等并分析的模型;在脉诊中,可以开发能识别脉搏跳动的波形及其对应蕴含的含意的模型;采用多智能体协同方式进行辨证开方的中医药LLM用以辅助医生。

当前中医药多模态LLM跨模态间效果不理想,其发展存在一定挑战,比如数据稀疏,缺乏高质量和标准化的多模态数据;中医包含着许多的隐性知识,LLM理解有限;中医药数据纷繁,时间跨度大,多模态数据对齐困难,而且不同模态的数据格式、分辨率等不同,需要更先进的技术进行归一

化和标准化^[93-94]。

4.5 四川地域中医药LLM有待开发 在四川地区,医疗和LLM的结合也变得更加紧密,如“华西医”大模型使用重症监护医学信息库第三版(MIMIC-III)、PubMed生物医学研究问答数据集(PubMedQA)及华西医院的病历等数据进行构建,辅助病例生成等^[95];中医药全产业链LLM“本草智库”汇集中药材的基因信息,中药成分和靶点等数据,为中药全产业链提供支持等^[96]。有研究指出当前的中医药LLM在中医流派上的能力相较于常见的知识而言,仍显不足。四川地域得益于得天独厚的地理环境和历经千年的变化,学术流派林立,呈现出多元化学术特色,LLM与川派中医的结合是一个值得研究的方向。

4.6 中医处方推荐 中医药LLM中特别是中医处方推荐有待进一步开发。目前处方推荐的中医药LLM以训练的数据为基础,在训练数据中出现的频率高的药物更容易被推荐,大部分推荐模型根据训练数据得到症状-中药之间的对应关系,据此推荐药物。然而在实际临床中,大部分情况下,处方并不只是症状-中药这样对应。在TCM-KLLaMA中使用知识图谱融合症状、舌像、脉象进行处方推荐,相较于只关注症状有一定的提升。但中医处方是在中医理论指导下,根据患者病情,辨证之后开具的,当前的推荐模型,缺少中医理论方面的知识,虽然有中药的性味归经等知识,但缺乏相应的中医临床指导用药知识;推荐的处方大多数没有剂量,此外输出的结果只有处方,没有相应的辨证及开方的推理过程等。

5 总结

随着人工智能技术的不断发展,中医药LLM被广泛应用于临床、科研中,LLM为中医药的信息化发展提供了强有力的支撑。在未来,需要构建更高质量的中医古籍及中医辨证及处方推理数据,开发融合多模态数据,构建集望闻问切于一体的中医药LLM,另外在中医古籍的挖掘和问答上仍有待进一步提升,以便更大可能地发挥中医古籍的作用。在模型的评估上,需要规范或探索更适合中医领域不同场景的评估方法。在四川地区,可以深入将LLM和不同流派的中医药结合用于辅助临床、深入挖掘和传承川派中医的学术特色。在中医处方推荐LLM上,除了加入中医药方面的理论知识外,还应有相应的辨证和处方推理过程。

[利益冲突] 本文不存在任何利益冲突。

[参考文献]

- [1] 和中浚,江花,王丽. 川派中医学学术特色研究[J]. 南京中医药大学学报:社会科学版,2022,23(6):371-379.
HE Z J, JIANG H, WANG L, Research on the academic characteristics of Sichuan Chinese medical school[J]. J Nanjing Univ Tradit Chin Med Soc Sci Ed, 2022, 23(6): 371-379.
- [2] 中华人民共和国国务院. “十四五”中医药信息化发展规划[EB/OL]. (2022-11-25)[2025-05-26]. https://www.gov.cn/zhengce/zhengceku/2022-12/06/content_5730292.htm.
The State Council of the People's Republic of China, The "14th

- Five-Year Plan" for the development of Chinese medicine Informatization[EB/OL]. (2022-11-25)[2025-05-26]. https://www.gov.cn/zhengce/zhengceku/2022-12/06/content_5730292.htm.
- [3] 杨菁颖,谢玉芳,张磊昌,等. 中医人工智能技术的现状与未来[J]. 江西中医药,2024,55(10):69-72.
YANG J Y, XIE Y F, ZHANG L C, et al. The current status and future of artificial intelligence technology in Chinese medicine[J]. Jiangxi J Tradit Chin Med, 2024, 55(10): 69-72.
- [4] 代佳晴,姜宇轩,胡婧楠,等. 人工智能在中医药中的发展现状与伦理挑战[J]. 中国医学伦理学,2025,38(2):173-178.
DAI J Q, JIANG Y X, HU J N, et al. Development status and ethical challenges of artificial intelligence in traditional Chinese medicine[J]. Chin Med Ethics, 2025, 38(2): 173-178.
- [5] 萧文科,宋驰,陈士林,等. 中医药大语言模型的关键技术与构建策略[J]. 中草药,2024,55(17):5747-5756.
XIAO W K, SONG C, CHEN S L, et al. Key technologies and construction strategies of large language models for traditional Chinese medicine[J]. Chin Tradit Herb Drugs, 2024, 55(17): 5747-5756.
- [6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Advances in neural information processing systems, NY: Curran Associates, Inc., 2017, 30: 5998-6008.
- [7] ZHAO W X, ZHOU K, LI J, et al. A survey of large language models[J]. arXiv, 2023, doi: 10. 48550/arXiv. 2303. 18223.
- [8] CHEN Y, CHEN D, SONG S, et al. Hrde: Retrieval-augmented large language models for Chinese health rumor detection and explainability [J]. arXiv, 2024, doi: 10. 48550/arXiv. 2407. 00668.
- [9] 陈浩沅,陈罕之,韩凯峰,等. 垂直领域大模型的定制化:理论基础与关键技术[J]. 数据采集与处理,2024,39(3):524-546.
CHEN H L, CHEN H Z, HAN K F, et al. Domain-specific foundation-model customization: Theoretical foundation and key technology[J]. J Data Acquis Process, 2024, 39(3): 524-546.
- [10] WEI J, TAY Y, BOMMASANI R, et al. Emergent abilities of large language models [J]. arXiv, 2022, doi: 10. 48550/arXiv. 2206. 07682.
- [11] MAO H, LIU G, MA Y, et al. A survey to recent progress towards understanding in-context learning[J]. arXiv, 2024, doi: 10. 48550/arXiv. 2402. 02212.
- [12] DIAO S, WANG P, LIN Y, et al. Active prompting with chain-of-thought for large language models [J]. arXiv, 2023, doi: 10. 48550/arXiv. 2302. 12246.
- [13] ZHANG Y, LI Y, CUI L, et al. Siren's song in the AI ocean: A survey on hallucination in large language models[J]. arXiv, 2023, doi: 10. 48550/arXiv. 2309. 01219.
- [14] ADLAKHA V, BEHNAMGHADER P, LU X H, et al. Evaluating correctness and faithfulness of instruction-following models for question answering[J]. Trans Assoc Comput Linguist, 2024, 12: 681-699.
- [15] KANDPAL N, DENG H, ROBERTS A, et al. Large language models struggle to learn long-tail knowledge [C]//Proceedings of the international conference on machine learning (ICML 2023). Honolulu, HI, USA: PMLR, 2023: 15615-15627.
- [16] GAO Y, XIONG Y, GAO X, et al. Retrieval-augmented generation for large language models: A survey[J]. arXiv, 2023, doi: 10. 48550/arXiv. 2312. 10997.
- [17] 吴国栋,秦辉,胡全兴,等. 大语言模型及其个性化推荐研究[J]. 智能系统学报,2024,19(6):1351-1365.
WU G D, QIN H, HU Q X, et al. Research on large language models and personalized recommendation[J]. CAAI Trans Intell Syst, 2024, 19(6): 1351-1365.
- [18] 张钦彤,王昱超,王鹤羲,等. 大语言模型微调技术的研究综述[J]. 计算机工程与应用,2024,60(17):17-33.
ZHANG Q T, WANG Y C, WANG H X, et al. Comprehensive review of large language model fine-tuning[J]. Comput Eng Appl, 2024, 60(17): 17-33.
- [19] WU X K, CHEN M, LI W, et al. LLM fine-tuning: Concepts, opportunities, and challenges[J]. Big Data Cogn Comput, 2025, 9(4): 87.
- [20] YANG F, LIU J, ZHANG X, et al. MAIN: Mutual alignment is necessary for instruction tuning[J]. arXiv, 2025, doi: 10. 48550/arXiv. 2504. 12913.
- [21] 宋佳磊,左兴权,张修建,等. 大语言模型评估方法综述[J]. 宇航计测技术,2025,45(2):1-30.
SONG J L, ZUO X Q, ZHANG X J, et al. A review of large language model evaluation methods[J]. J Astronaut Metrol Meas, 2025, 45(2): 1-30.
- [22] SHEN S, HOU L, ZHOU Y, et al. Mixture-of-experts meets instruction tuning: A winning combination for large language models[J]. arXiv, 2023, doi: 10. 48550/arXiv. 2305. 14705.
- [23] CAI W, JIANG J, WANG F, et al. A survey on mixture of experts [J]. arXiv, 2024, doi: 10. 48550/arXiv. 2407. 06204.
- [24] DU N, HUANG Y, DAI A M, et al. Glam: Efficient scaling of language models with mixture-of-experts [C]//Proceedings of the International conference on machine learning (ICML 2022). Baltimore, Maryland, USA: PMLR, 2022: 5547-5569.
- [25] LIAO Y, JIANG S, WANG Y, et al. MING-MOE: Enhancing medical multi-task learning in large language models with sparse mixture of low-rank adapter experts [J]. arXiv, 2024, doi: 10. 48550/arXiv. 2404. 09027.
- [26] CAI W, JIANG J, WANG F, et al. A survey on mixture of experts in large language models[J]. IEEE Trans Knowl Data Eng, 2025, 37(7): 3896-3915.
- [27] LI X, WANG S, ZENG S, et al. A survey on LLM-based multi-agent systems: Workflow, infrastructure, and challenges [J]. Vicinagearth, 2024, 1(1): 9.
- [28] BANSAL P. Prompt engineering importance and applicability with generative AI[J]. J Comput Commun, 2024, 12(10): 14-23.
- [29] 陈道彬,张子诺,付裕彬,等. 基于Chinese-CLIP模型和Prompt提示机制的图文检索方法[J]. 现代信息科技,2025,9(6):130-134.
CHEN D B, ZHANG Z N, FU Y B, et al. An image-text retrieval method based on Chinese-CLIP model and prompt mechanism

- [J]. *Mod Inf Technol*, 2025, 9(6): 130-134.
- [30] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. *Adv Neural Inf Process Syst*, 2022, 35: 24824-24837.
- [31] 李欣桐, 马素芬, 张丰聪, 等. 中医药领域大语言模型的研究进展与应用前景[J]. *南京中医药大学学报*, 2024, 40(12): 1393-1403. LI X T, MA S F, ZHANG F C, et al. Research progress and application prospect of large language model in the traditional Chinese medicine[J]. *J Nanjing Univ Tradit Chin Med*, 2024, 40(12): 1393-1403.
- [32] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[J]. *Adv Neural Inf Process Syst*, 2020, 33: 9459-9474.
- [33] RAM O, LEVINE Y, DALMEDIGOS I, et al. In-context retrieval-augmented language models[J]. *Trans Assoc Comput Linguist*, 2023, 11: 1316-1331.
- [34] 刘彦宏, 崔永瑞. 基于 Word2Vec 模型与 RAG 框架的医疗检索增强生成算法[J]. *Artif Intell Robot Res*, 2024, 13: 479. LIU Y H, CUI Y R. Enhanced generation algorithm for medical retrieval based on Word2Vec model and RAG framework[J]. *Artif Intell Robot Res*, 2024, 13: 479.
- [35] ZHANG D, XU J, ZHOU J, et al. KAG-Thinker: Teaching large language models to think with human-like reasoning process[J]. *arXiv*, 2025, doi: 10. 48550/arXiv. 2506. 17728.
- [36] CASPER S, DAVIES X, SHI C, et al. Open problems and fundamental limitations of reinforcement learning from human feedback[J]. *arXiv*, 2023, doi: 10. 48550/arXiv. 2307. 15217.
- [37] KAUFMANN T, WENG P, BENGIS V, et al. A survey of reinforcement learning from human feedback[J]. *arXiv*, 2023, doi: 10. 48550/arXiv. 2312. 14925.
- [38] LEE H, PHATALE S, MANSOOR H, et al. Rlaif: Scaling reinforcement learning from human feedback with AI feedback [C]//*Proceedings of the 41st International Conference on Machine Learning*. Vienna, Austria: PMLR, 2024, 235: 26874-26901.
- [39] HAN S Y, WANG M Y, ZHANG J L, et al. A review of large language models: Fundamental architectures, key technological evolutions, interdisciplinary technologies integration, optimization and compression techniques, applications, and challenges[J]. *Electron*, 2024, doi: 10. 3390/sym13245040.
- [40] XU X, LI M, TAO C, et al. A survey on knowledge distillation of large language models [J]. *arXiv*, 2024, doi: 10. 48550/arxiv. 2402. 13116.
- [41] NIU S, MA J, BAI L, et al. Multimodal clinical reasoning through knowledge-augmented rationale generation[J]. *arXiv*, 2024, doi: 10. 48550/arxiv. 2411. 07611.
- [42] LIU Q, WU X, ZHAO X, et al. Large language model distilling medication recommendation model [J]. *arXiv*, 2024, doi: 10. 48550/arxiv. 2402. 02803.
- [43] FANG L, YU X, CAI J, et al. Knowledge distillation and dataset distillation of large language models: Emerging trends, challenges, and future directions[J]. *arXiv*, 2025, doi: 10. 48550/arxiv. 2504. 14772.
- [44] 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]. *Libr Trib*, 2024, 44(10): 103-112.
- [45] 柴景贤, 郎许锋, 李红岩, 等. 基于 Lora 微调的轻量化中医古籍大语言模型研究[J]. *世界科学技术—中医药现代化*, 2025, 27(3): 823-831. CHAI J X, LANG X F, LI H Y, et al. Research on lightweight large language models for ancient traditional Chinese medicine texts based on Lora fine-tuning[J]. *Mod Tradit Chin Med Mater Med-World Sci Technol*, 2025, 27(3): 823-831.
- [46] HUA R, DONG X, WEI 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.
- [47] FU C, FU Z, YAN S, et al. In-context multitask learning for few-shot fine-tuning of large language models in traditional Chinese medicine tongue diagnosis [C]//*2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Hyderabad, India: IEEE, 2025.
- [48] QI J, WANG X, YANG T. Traditional Chinese medicine prescription recommendation model based on large language models and graph neural networks [C]//*2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Istanbul, Turkey: IEEE, 2023.
- [49] QIN S, WANG Y, CUI T, et al. Intelligent Chinese patent medicine (CPM) recommendation framework: Integrating large language models, retrieval-augmented generation, and the largest CPM dataset[J]. *Pharmacol Res*, 2025, 219: 107883.
- [50] HAOYU T, KUO Y, XIN D, et al. TCMLLM-PR: Evaluation of large language models for prescription recommendation in traditional Chinese medicine[J]. *Digit Chin Med*, 2024, 7(4): 343-355.
- [51] ZHUANG Y, YU L, JIANG N, et al. TCM-KLLaMA: Intelligent generation model for traditional Chinese medicine prescriptions based on knowledge graph and large language model[J]. *Comput Biol Med*, 2025, 189: 109887.
- [52] CHAN Y C, HUANG P Y, CHEN Z L, et al. Pilot study of retrieval-augmented generation model in recommending traditional Chinese medicine formulations [C]//*International Conference on Biomedical and Health Informatics*, 2024: *Proceedings of ICBHI 2024*. Tainan, Taiwan: Springer, 2025: 331-341.
- [53] LIU C, SUN K, ZHOU Q, et al. CPML-ChatGLM: Parameter-efficient fine-tuning ChatGLM with Chinese patent medicine instructions[J]. *Sci Rep*, 2024, 14(1): 6403.
- [54] 李盼飞, 杨小康, 白逸晨, 等. 基于大语言模型的中医医案命名实体抽取研究[J]. *中国中医药图书情报杂志*, 2024, 48(2): 108-113.

- LI P F, YANG X K, BAI Y C, et al. Study on named entity extraction in TCM medical records based on large language models[J]. *Chin J Libr Inf Sci Tradit Chin Med*, 2024, 48(2): 108-113.
- [55] LI M F, ZHENG X H. Identification of ancient Chinese medical prescriptions and case data analysis under artificial intelligence GPT algorithm: A case study of Song dynasty medical literature [J]. *IEEE Access*, 2023, 11: 131453-131464.
- [56] TRAJANOSKA M, STOJANOV R, TRAJANOV D. Enhancing knowledge graph construction using large language models[J]. *arXiv*, 2023, doi:10.48550/arXiv.2305.04676.
- [57] LI H, XIA C, HOU Y, et al. TCMRD - KG: Innovative design and development of rheumatology knowledge graph in ancient Chinese literature assisted by large language models[J]. *Front Pharmacol*, 2025, 16: 1535596.
- [58] 李玥, 洪海蓝, 李文林, 等. 大语言模型构建鼻炎医案知识图谱的应用研究[J]. *计算机工程与应用*, 2025, 61(4): 167-175.
- LI Y, HONG H L, LI W L, et al. Study on application of large language model in constructing knowledge graph of medical cases of rhinitis[J]. *Comput Eng Appl*, 2025, 61(4): 167-175.
- [59] 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*, 2026, 44(1): 1-6, 259-263.
- TANG S L, ZHAO N, LONG Q Y, et al. Research on the automated construction and clinical decision support of traditional Chinese medicine gynecology knowledge graphs based on ChatGLM[J]. *Chin Arch Tradit Chin Med*, 2026, 44(1): 1-6, 259-263.
- [60] 何宇浩, 李明, 罗晓兰, 等. 基于GPTs的中医知识图谱实体和关系抽取研究[J]. *上海中医药杂志*, 2024, 58(8): 1-6.
- HE Y H, LI M, LUO X L, et al. Research on entity and relation extraction from traditional Chinese medicine knowledge graphs based on GPTs[J]. *Shanghai J Tradit Chin Med*, 2024, 58(8): 1-6.
- [61] WANG X, YANG Q, QIU Y, et al. Knowledgept: Enhancing large language models with retrieval and storage access on knowledge bases[J]. *arXiv*, 2023, doi:10.48550/arxiv.2308.11761.
- [62] SUN J, XU C, TANG L, et al. Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph[J/OL]. *arXiv*, 2023, doi:10.48550/arxiv.2307.07697.
- [63] WEN Y, WANG Z, SUN J. Mindmap: Knowledge graph prompting sparks graph of thoughts in large language models[J]. *arXiv*, 2023, doi:10.48550/arXiv.2308.09729.
- [64] 周宗楨. 中医疫病文献的知识图谱构建与应用研究[D]. 南京: 南京中医药大学, 2024.
- ZHOU Z Z. Construction and application research of knowledge graph in traditional Chinese medicine literature on epidemic disease[D]. Nanjing: Nanjing University of Chinese Medicine, 2024.
- [65] 董兆安, 秦可豪, 周子力, 等. 基于GraphRAG的中医药知识问答系统[J]. *曲阜师范大学学报: 自然科学版*, 2025, doi:37.1154.N.20250421.1311.003.
- DONG Z A, QIN K H, ZHOU Z L, et al. A traditional Chinese medicine knowledge question-answering system based on GraphRAG[J]. *J Qufu Normal Univ: Natu Sci Ed*, 2025, doi:37.1154.N.20250421.1311.003.
- [66] HE J, GUO Y, LAM L K, et al. OpenTCM: A GraphRAG-empowered LLM-based system for traditional Chinese medicine knowledge retrieval and diagnosis [J]. *arXiv*, 2025, doi:10.48550/arXiv.2504.20118.
- [67] 苑中旭, 李理, 何凡, 等. 融合思维链与知识图谱的中医问答模型[J]. *计算机工程与应用*, 2025, 61(4): 158-166.
- YUAN Z X, LI L, HE F, et al. Traditional Chinese medicine question answering model based on chain-of-thought and knowledge graph[J]. *Comput Eng Appl*, 2025, 61(4): 158-166.
- [68] 海佳丽, 汪润, 袁良志, 等. 基于检索增强的中医药标准知识问答系统构建探索与实践[J]. *数据分析与知识发现*, 2025, 9(7): 165-174.
- HAI J L, WANG R, YUAN L Z, et al. Constructing a retrieval-augmented question-answering system for traditional Chinese medicine standards[J]. *Data Anal Knowl Discov*, 2025, 9(7): 165-174.
- [69] ABUNASSER R J, ALI M Z, JARARWEH Y, et al. Large language models in drug discovery: A comprehensive analysis of drug-target interaction prediction [C]//2024 2nd International Conference on Foundation and Large Language Models (FLLM). IEEE, 2024.
- [70] YUAN S, ZHOU Z, JIN X, et al. Enhancing herbal medicine-drug interaction prediction using large language models[J]. *IEEE J Biomed Health Inf*, 2025, doi:10.1109/jbhi.2025.3558667.
- [71] YANG K, DONG X, ZHANG S, et al. PresRecRF: Herbal prescription recommendation via the representation fusion of large TCM semantics and molecular knowledge[J]. *Phytomed*, 2024, 135: 156116.
- [72] YANG H, XIU J, YAN W, et al. Large language models as tools for molecular toxicity prediction: AI insights into cardiotoxicity [J]. *J Chem Inf Model*, 2025, 65(5): 2268-2282.
- [73] WANG Z, LI K, PENG S, et al. A weighted voting approach for traditional Chinese medicine formula classification using large language models: Algorithm development and validation study [J]. *JMIR Med Inf*, 2025, 13: e69286.
- [74] DOU Y, HUANG Y, ZHAO X, et al. ShennongMGS: An LLM-based Chinese medication guidance system [J]. *ACM Trans Manag Inf Syst*, 2025, 16(2): 1-15.
- [75] YUE W, WANG X, ZHU W, et al. Tembench: A comprehensive benchmark for evaluating large language models in traditional Chinese medicine [J]. *arXiv*, 2024, doi:10.48550/arXiv.2406.01126.
- [76] YU P, SONG K, HE F, et al. TCMD: A traditional Chinese medicine QA dataset for evaluating large language models[J]. *arXiv*, 2024, doi:10.48550/arXiv.2406.04941.
- [77] HUANG T, LU L, CHEN J, et al. TCM-3CEval: A triaxial benchmark for assessing responses from large language models in traditional Chinese medicine [J]. *arXiv*, 2025, doi:10.48550/

- arXiv. 2503. 07041.
- [78] WANG Z, HAO M, PENG S, et al. TCMEval-SDT: A benchmark dataset for syndrome differentiation thought of traditional Chinese medicine[J]. *Sci Data*, 2025, 12(1):437.
- [79] KONG S, YANG X, WEI Y, et al. MTCMB: A multi-task benchmark framework for evaluating LLMs on knowledge, reasoning, and safety in traditional Chinese medicine[J]. arXiv, 2025, doi: 10. 48550/arXiv. 2506. 01252.
- [80] XIE J, YU Y, ZHANG Z, et al. TCM-Ladder: A benchmark for multimodal question answering on traditional Chinese medicine [J]. arXiv, 2025, doi: 10. 48550/arxiv. 2505. 24063.
- [81] 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]. *中华中医药学刊*, 2025, 43(10): 19-27, 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]. *Chin Arch Tradit Chin Med*, 2025, 43(10): 19-27.
- [82] LI Y, PENG X, LI J, et al. Relation extraction using large language models: A case study on acupuncture point locations[J]. *J Am Med Inf Assoc*, 2024, 31(11): 2622-2631.
- [83] 韩家伟, 裴健, 范明, 等. 数据挖掘: 概念与技术[M]. 北京: 机械工业出版社, 2012.
HAN J W, PEI J, FAN M, et al. *Data mining: Concepts and techniques*[M]. Beijing: China Machine Press, 2012.
- [84] 李明, 罗晓兰, 朱邦贤. 中医古籍方剂数据挖掘与知识问答系统构建[J]. *图书馆论坛*, 2025, 45(4): 49-59.
LI M, LUO X L, ZHU B X. Data mining and knowledge Q&A system for traditional Chinese medicine formulas in ancient books [J]. *Libr Trib*, 2025, 45(4): 49-59.
- [85] 潘晔, 娄静, 潘玉颖, 等. 中医药数据挖掘现状分析与创新探索 [J]. *中国中医药信息杂志*, 2022, 29(5): 5-9.
PAN Y, LOU J, PAN Y Y, et al. Status analysis and innovation exploration of TCM data mining technology[J]. *Chin J Inf Tradit Chin Med*, 2022, 29(5): 5-9.
- [86] MA M, TIAN G, WANG K, et al. A study of automated data mining of medication rules in traditional Chinese medicine prescriptions based on large language model[C]//2024 IEEE 7th International Conference on Computer and Communication Engineering Technology (CCET). IEEE, 2024.
- [87] JI Z, LEE N, FRIESKE R, et al. Survey of hallucination in natural language generation[J]. *ACM Comput Surv*, 2023, 55(12): 1-38.
- [88] 刘泽垣, 王鹏江, 宋晓斌, 等. 大语言模型的幻觉问题研究综述 [J]. *软件学报*, 2025, 36(3): 1152-1185.
LIU Z Y, WANG P J, SONG X B, et al. Survey on hallucinations in large language models[J]. *J Softw*, 2025, 36(3): 1152-1185.
- [89] TANG J, WU N, GAO F, et al. From metaphor to mechanism: How LLMs decode traditional Chinese medicine symbolic language for modern clinical relevance[J]. arXiv, 2025, doi: 10. 48550/arXiv. 2503. 02760.
- [90] GUNASEKAR S, ZHANG Y, ANEJA J, et al. Textbooks are all you need[J]. arXiv, 2023, doi: 10. 48550/arXiv. 2306. 11644.
- [91] LI J, WANG X, WU X, et al. Huatuo-26m, a large-scale Chinese medical QA dataset [J]. arXiv, 2023, doi: 10. 48550/arXiv. 2305. 01526.
- [92] ZHAO H, CHEN H, YANG F, et al. Explainability for large language models: A survey[J]. *ACM Trans Intell Syst Technol*, 2024, 15(2): 1-38.
- [93] ALSAAD R, ABD-ALRAZAQ A, BOUGHORBEL S, et al. Multimodal large language models in health care: Applications, challenges, and future outlook[J]. *J Med Internet Res*, 2024, 26: e59505.
- [94] 廉坤, 王学琴, 谭朵廷, 等. 基于人工智能探讨中医诊断的现代化发展[J]. *中国实验方剂学杂志*, 2026, 32(9): 336-346.
LIAN K, WANG X Q, TAN D T, et al. Explore the modern development of traditional Chinese medicine diagnosis based on artificial intelligence[J]. *Chin J Exp Tradit Med Form*, 2026, 32(9): 336-346.
- [95] 石锐, 郑兵, 姚巡, 等. “华西黄医”大模型构建与应用[J]. *中国胸心血管外科临床杂志*, 2025, 32(5): 587-593.
SHI R, ZHENG B, YAO X, et al. Construction and application of the "Huaxi Hongyi" large medical model[J]. *Chin J Clin Thorac Cardiovasc Surg*, 2025, 32(5): 587-593.
- [96] 罗韬, 肖莹佩. 四川“医疗+AI”渐行渐近亟待突破数据瓶颈[EB/OL]. (2025-5-19)[2025-05-26]. <https://epaper.scdaily.cn/shtml/scrb/20250519/326554.shtml>.
LUO T, XIAO Y P. Sichuan's "medical+AI" is getting closer and closer, urgently need to break through the data bottleneck[EB/OL]. (2025-5-19)[2025-05-26]. <https://epaper.scdaily.cn/shtml/scrb/20250519/326554.shtml>.

[责任编辑 顾雪竹]