http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00027
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

基于临床信息状态跟踪的人工智能辅助多轮预问诊研究

唐文珺[1]

[1]中国医学科学院北京协和医学院基础医学研究所，北京 100005

中图分类号：**R319**；**TP18**

[摘要] 目的 评价基于临床信息状态跟踪的人工智能辅助多轮预问诊系统（clinical information state tracking-based medical pre-consultation system，ClinIST-Med）的信息采集效果及交互效率。**方法** 前瞻性采集 2026 年 5 月海南医科大学附属医院和哈尔滨医科大学附属医院 4 个专科 80 例患者的门诊对话及病历，以患者初始自述为共同输入，重建后续交互。ClinIST-Med 通过临床信息需求图谱跟踪信息状态，由规则引擎选择追问目标，GPT-4o 生成问题。比较 ClinIST-Med、GPT-4o 自由追问和固定问诊的信息采集指标、交互轮次及报告临床可用性。**结果** ClinIST-Med 的信息完整度、关键信息覆盖率和临床可用性评分均高于自由追问组及固定问诊组；重复提问率和平均交互轮次均低于 2 种对照方法。**结论** 在基于真实门诊资料的交互重建中，ClinIST-Med 提高了病史信息采集效果和报告临床可用性，减少了重复提问及交互轮次。

关键词：人工智能；大语言模型；病史采集；预问诊

10.12201/bmr.202611.00027V1

Artificial Intelligence-Assisted Multi-Round Pre-Consultation Based on Clinical Information State Tracking

Tang Wenjun[1]

[1]Institute of Basic Medical Sciences & School of Basic Medicine， Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China

Abstract: **Objective** To evaluate the information acquisition performance and interaction efficiency of a clinical information state tracking–based medical pre-consultation system (ClinIST-Med). **Methods** Outpatient dialogues and medical records were prospectively collected from 80 patients across four specialties at the Affiliated Hospital of Hainan Medical University and the Affiliated Hospital of Harbin Medical University in May 2026. Each patient’s initial spontaneous statement served as the common input for reconstructing subsequent interactions. ClinIST-Med tracked information states using a clinical information requirement map, selected follow-up questioning targets through a rule engine, and generated questions using GPT-4o. ClinIST-Med, unconstrained GPT-4o questioning, and

唐文珺
t13078781939@163.com

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00027
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

fixed questioning were compared in terms of information acquisition metrics, interaction rounds, and the clinical usability of the resulting reports. **Results** ClinIST-Med achieved higher information completeness, key information coverage, and clinical usability scores than both comparator approaches, with a lower repeat-question rate and fewer interaction rounds. **Conclusion** In interactions reconstructed from real outpatient data, ClinIST-Med improved medical history acquisition and report usability while reducing repeated questions and interaction rounds.

Keywords: artificial intelligence; large language model; medical history taking; pre-consultation

10.12201/bmr.202611.00027V1

唐文珺
t13078781939@163.com

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00027
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

1 引言

真实门诊患者自述常存在信息碎片化、表达口语化及关键病史遗漏，医生需进一步核对症状、病程、既往史和用药情况，由此增加基础信息核对和重复询问的时间。预问诊将部分病史采集前移，可在患者接诊前形成初步病史信息，为后续门诊沟通提供基础[1, 2]。

现有预问诊主要采用固定问卷和基于大语言模型（large language model，LLM）的动态问诊。固定问卷能够覆盖预设病史要素，但通常按照固定顺序提问，难以根据患者已经提供的信息动态调整。LLM 能够根据对话内容动态生成自然语言问题，但信息采集质量与交互效率仍需兼顾[3, 4]。Luo 等发现，LLM 智能体的眼科预问诊病史质量评分高于住院医师，但交互时间更长[8]。一项纳入 24 个专科 2069 名患者的随机对照试验显示，聊天机器人的应用与医生问诊时间缩短及医生感知的诊疗协调改善相关[9]。国内医学大模型在知识问答、文本生成和临床辅助等方面的研究也为预问诊提供了技术基础[5-7]。

然而，预问诊的关键不仅是生成合理的问题，还包括判断已有病史信息是否充分，并据此确定后续追问的内容与优先级。如果缺少对临床信息采集状态的持续跟踪，即使能够生成自然语言问题，也难以保证多轮交互始终围绕尚未明确的关键病史要素展开。因此，结合临床信息状态判断与追问目标选择，是优化预问诊信息采集的关键。

本研究构建基于临床信息状态跟踪的医学预问诊系统（clinical information state tracking–based medical pre-consultation system，ClinIST-Med），通过临床信息需求图谱将病史要素划分为完整、部分完整和缺失 3 种状态，根据临床优先级确定追问目标，再由 GPT-4o 生成符合患者理解水平的自然语言问题。基于真实门诊医患对话数据重建交互过程，将 ClinIST-Med 与 GPT-4o 自由追问及固定问诊进行比较，评价信息完整度、关键信息覆盖率、重复提问率、平均交互轮次及生成报告的临床可用性。

10.12201/bmr.202611.00027V1

2 资料与方法

2.1 研究对象

本研究经海南医科大学伦理委员会和哈尔滨医科大学附属第二医院伦理委员会批准（批准号分别为 HYMLL-2026-109、KY2026-157），于 2026 年 5 月在海南医科大学附属医院和哈尔滨医科大学附属医院前瞻性采集真实门诊医患对话数据。纳入标准：① 年龄 18～75 岁；② 意识清楚，具备自主表达能力；③ 具有完整的门诊病史资料，可用于确定病例适用的临床信息要素及评价参照。排除标准：① 存在认知障碍或严重精神障碍，无法完成正常交流；

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00027
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

② 存在明显听力或语言表达障碍；③ 录音质量不符合分析要求，有效语音占比<70%；④ 拒绝签署知情同意书。

研究人员取得患者书面知情同意，采用科大讯飞 SR502 型录音笔现场录制门诊医患对话，并通过两院医院信息系统调取对应患者本次就诊的门诊病历，与录音逐一匹配。共纳入 80 例，其中海南医科大学附属医院 60 例，包括肝胆外科、内分泌科和心内科各 20 例；哈尔滨医科大学附属医院 20 例，均来自肿瘤化疗科。

2.2 数据预处理

真实世界门诊对话可能存在语音转写误差、信息遗漏、表达不完整及语义不明确等噪声问题，为减少语音识别和人工转写误差，由 2 名经过培训的医学研究生独立转写录音，对分歧结合原始录音复核，由第 3 名研究者裁定。转写文本及门诊病历均在分析前完成去标识化处理，删除或掩码姓名、身份证号、联系方式及医院登记号。

为模拟患者首次使用预问诊系统的初始状态，将医生正式接诊前、未经系统引导的患者完整主动陈述作为 3 种方法的共同初始输入，平均 156 字，范围 42～387 字。

本研究未要求同一患者重复完成 3 次真实预问诊，而是依据其完整医患对话及医生诊疗记录，建立病例的临床信息参照标准，整理可确认的病史要素，用于重建后续回答并评价信息采集情况。针对各方法生成的问题，按其涉及的临床信息要素检索原始记录，仅将能够明确支持回答的内容判为有效匹配；主题相同但不足以确定信息状态的内容不予匹配。无明确证据时不推测补充，也不将未匹配判为患者无相应病史。

3 种方法采用相同的信息来源和匹配原则，独立更新临床信息状态，减少信息来源及处理规则差异对比较的影响。每次运行均重置交互状态，仅保留患者初始自述及当前方法已获得的交互内容，避免不同方法间的信息记忆相互干扰。

2.3 临床信息需求图谱与规则库构建

依据各专科临床诊疗指南、医学系统命名法临床术语和《疾病和有关健康问题的国际统计分类》第十次修订本，结合门诊病史采集流程，构建临床信息需求图谱及规则库。根据专科及主要就诊问题确定病例适用信息元，包括症状部位、起病时间、持续时间及演变、性质、加重或缓解因素、伴随症状、既往史和用药史等。

信息元划分为完整（complete，C）、部分完整（partial，P）和缺失（missing，M）3 种状态。C 表示该信息元所需的关键临床参数均已获得；P 表示已提及相关信息，但缺少影响判断的关键参数；M 表示尚未获得相关信息。

10.12201/bmr.202611.00027V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00027
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

各专科 2 名高级职称医师采用 1～5 分独立评价信息元的临床重要性，分歧由第 3 名高级职称医师裁定。评分的加权 kappa 值为 0.87。信息元及其权重见附表 1。

2.4 ClinIST-Med 系统架构与运行机制

ClinIST-Med 由临床信息状态跟踪引擎（clinical information state tracking engine，CIST-Engine）和大语言模型生成模块组成，分别负责信息状态更新与目标选择、自然语言问题生成。

每轮交互后，CIST-Engine 通过实体识别和信息槽位匹配更新各信息元的状态。根据临床优先级，从 P、M 状态信息元中选择至多 3 个追问目标；C 状态信息元不再入选，但生成问句仍可能涉及已完整获得的信息。当患者补充的信息满足相应信息元的完整性要求时，系统将其状态更新为 C；当新增信息仍不足以满足完整性要求时，则保留 P 或 M 状态。获得补充回答后重新判定状态，并选择下一轮目标。

将目标信息元的定义、状态、相关对话及安全约束输入 GPT-4o（gpt-4o-2024-05-13）。模型参数设置为 temperature=0.2、top_p=0.8、max_tokens=150。提示词要求模型围绕指定追问目标，每轮生成至多 3 个简短、通俗的临床问题，每个问题对应 1 个目标信息元，不得主动扩展至其他病史内容，不得引入未经当前对话支持的临床假设，也不得提供诊断或治疗建议。

满足以下任一条件时终止交互：① 病例适用的关键临床信息元均达到 C 状态；② 连续 3 轮未获得新的有效临床信息；③ 达到 12 轮交互。交互结束后生成结构化预问诊报告。

2.5 对照方法与评价指标

设置 GPT-4o 自由追问和固定问诊 2 种对照方法。自由追问采用相同的患者初始自述，由 GPT-4o 根据自身交互历史自主提问，不设信息元目标约束，模型版本及生成参数与 ClinIST-Med 一致。固定问诊采用 2 名临床专家依据专科信息需求图谱制定的问卷，按固定顺序提问，不根据初始自述动态跳转。3 种方法采用相同的交互终止条件，即病例适用的关键临床信息元均达到 C 状态、连续 3 轮未获得新的有效临床信息或达到最大 12 轮交互中的任一条件时终止。每种方法独立运行并重置交互状态。

本研究选择 GPT-4o 作为自由追问的生成式对照，主要考虑其通用语言理解与生成能力，以及通过应用程序编程接口实现多轮交互实验的可行性。研究固定使用 GPT-4o 并统一生成参数，以保证实验条件的一致性和可复现性。固定问诊作为预设问题驱动的对照。本研究重点比较不同问诊策略在临床信息采集中的表现，而非对不同大语言模型进行整体性能排名。

10.12201/bmr.202611.00027V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00027
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

主要评价指标包括信息完整度、关键信息覆盖率、重复提问率和平均交互轮次。信息完整度为交互结束时处于 C 或 P 状态的信息元数占病例适用信息元总数的百分比；关键信息覆盖率为权重≥4 分的适用信息元中达到 C 状态的比例。重复提问率为再次询问此前已完整获得信息的问题数占总问题数的百分比，按实际问句逐题判定；交互轮次为初始自述后至终止交互完成的问答轮数。

由 3 名门诊临床专家依据统一标准，采用 5 分李克特量表独立评价 3 种方法生成的结构化预问诊报告。所有报告随机编码，评价者不知系统名称及生成过程，以减少主观偏倚。为降低真实世界对话中信息缺失、表达不完整及语义不明确等噪声对评分的影响，专家对照患者初始自述、完整医患对话及医生诊疗记录，区分原始资料未提及或无法确认的信息与系统对已有信息的遗漏、错误归纳或不当补充。评价涵盖信息完整性、内容准确性、临床相关性、组织清晰度及对后续问诊的辅助价值，具体标准见附表 2。

2.6 统计学方法

采用 SPSS 26.0 进行统计分析。计量资料经正态性检验后以均数±标准差表示。同一批病例的 3 种方法比较采用重复测量方差分析，两两比较采用 Bonferroni 校正；不满足参数检验条件时采用相应非参数检验。采用双侧检验，以 $P<0.05$ 为差异有统计学意义。

2.7 整体框架示

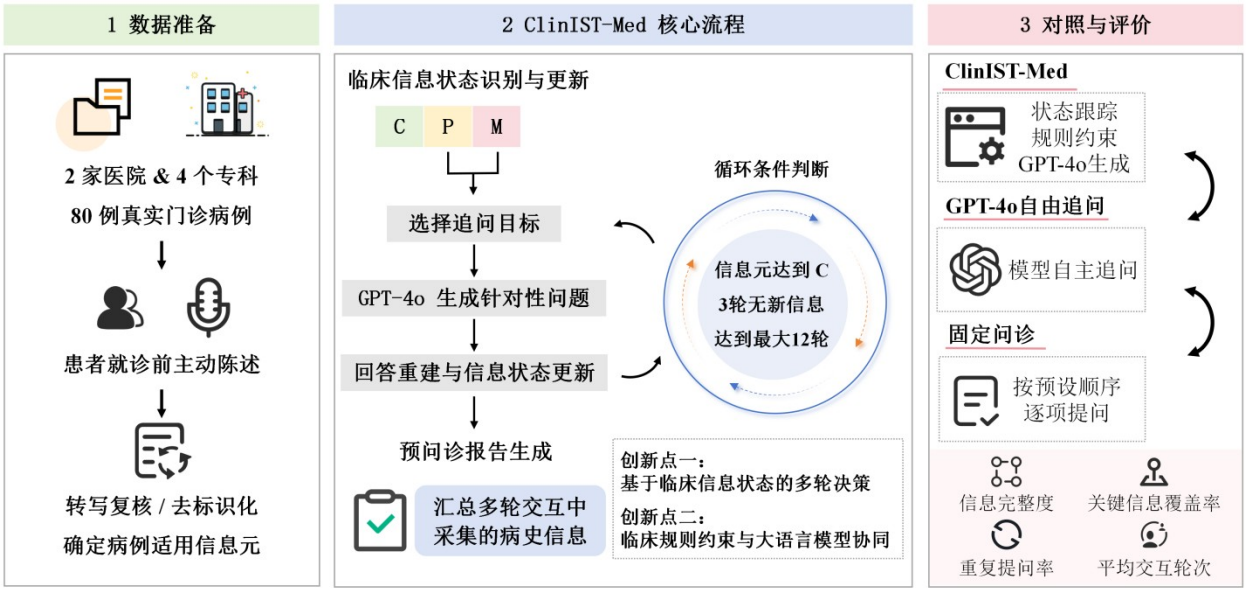


图 1 整体框架示意图

10.12201/bmr.202611.00027V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00027
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

3 结果

3.1 病例分布

本研究纳入的 80 例门诊病例在地区与专科维度的分布详见表 1。

表 1 2 家医疗机构及临床科室病例分布［$\bar{x} \pm s$，n（%）］

地区	科室	年龄（岁）	总数	男性	女性
东北	肿瘤化疗科	58.8±10.4	20（25.0）	8（40.0）	12（60.0）
华南	肝胆外科	56.7±10.2	20（25.0）	6（30.0）	14（70.0）
华南	内分泌科	53.4±10.6	20（25.0）	8（40.0）	12（60.0）
华南	心内科	46.2±28.2	20（25.0）	11（55.0）	9（45.0）
合计	--	--	80（100.0）	33（41.3）	47（58.7）

3.2 3 组预问诊方法的效果及效率对比

ClinIST-Med 的信息完整度和关键信息覆盖率分别为 93.8 ± 4.2%和 95.1 ± 3.8%，均高于 2 种对照方法；重复提问率和平均交互轮次分别为 5.8 ± 1.9%和 7.1 ± 1.2 轮，均低于 2 种对照方法。各指标总体差异均有统计学意义（P<0.001），两两比较结果见表 2。

表 2 3 组预问诊方法的信息采集效果及交互效率比较 （$\bar{x} \pm s$）

注：*与 ClinIST-Med 组比较，P<0.05；#与 GPT-4o 自由追问组比较，P<0.05。

组别	信息完整度（%）	关键信息覆盖率（%）	重复提问率（%）	平均交互轮次（轮）
自由追问组	84.7±6.5*	87.1±5.9*	14.6±3.8*	9.3±1.8*
固定问诊组	89.8±5.1*#	91.4±4.6*#	18.3±4.2*#	10.2±2.1*#
ClinIST-Med	**93.8±4.2**	**95.1±3.8**	**5.8±1.9**	**7.1±1.2**

3.3 临床信息状态分布与更新分析

表 3 展示了 ClinIST-Med 组中信息状态的初始分布与更新情况，在初始 P、M 信息元中，分别有 347、396 个最终达到 C 状态，另有 209 个 M 信息元转为 P 状态。被追问数按至少被直接追问 1 次的不同信息元计数，同一信息元多次追问仅计 1 个。

表 3 初始状态与后续追问行为

初始状态

初始信息元数（n）

10.12201/bmr.202611.00027V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00027
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

	被追问数（*n*）	追问率（%）	最终为 C 数（*n*）	最终为 P 数（*n*）	最终为 M 数（*n*）
C	427	0	0.0	427	0	0
P	496	431	86.9	347	126	23
M	687	621	90.4	396	209	82

10.12201/bmr.202611.00027V1

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00027
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

3.4 临床可用性评价与亚组分析

ClinIST-Med、自由追问和固定问诊的临床可用性评分分别为 4.42 ± 0.48、3.83 ± 0.52、4.08 ± 0.45 分，总体差异有统计学意义（P<0.001）。专家评分的组内相关系数（intraclass correlation coefficient，ICC）为 0.86，95%置信区间为 0.80～0.91。ClinIST-Med 在肝胆外科、内分泌科、心内科和肿瘤化疗科的信息完整度分别为 94.1%、92.7%、95.8%、92.6%。

4 讨论

4.1 临床信息状态跟踪对信息采集效果及交互效率的影响

本研究中，ClinIST-Med 的信息完整度和关键信息覆盖率较自由追问组分别提高 9.1%和 8.0%，较固定问诊组分别提高 4.0%和 3.7%，说明 ClinIST-Med 能够覆盖更多适用的病史信息元，并更充分地获取关键临床信息。其原因可能在于，固定问诊按照预设顺序提问，难以根据已有信息调整后续内容；自由追问虽然能够动态生成问题，但缺少明确的信息状态约束。ClinIST-Med 则结合信息状态和临床优先级确定目标，将追问集中于尚未完整获得的重要信息。

状态转移结果中，347 个初始 P 信息元和 396 个初始 M 信息元最终达到 C 状态，209 个 M 信息元转为 P 状态，体现了对已有信息的完善和对缺失内容的补充。这也说明，信息采集既包括扩大病史覆盖范围，也包括补齐影响判断的关键参数。部分信息元在交互结束时仍未达到完整状态，说明基于已有门诊资料重建回答的方式仍受到原始记录内容的限制。

ClinIST-Med 的重复提问率较自由追问组和固定问诊组分别降低 8.8%和 12.5%；平均交互轮次分别减少 2.2 和 3.1 轮。结果说明，系统优先补充尚未完整采集的信息，避免重复追问已完整获得的内容，有助于提高信息采集效果并减少交互轮次。

10.12201/bmr.202611.00027V1

4.2 预问诊报告的临床可用性及应用意义

ClinIST-Med 生成报告的临床可用性评分高于自由追问组和固定问诊组，说明更充分的病史采集可能有助于形成更便于临床查阅的报告。尤其是关键信息覆盖率的提高，意味着更多对当前临床判断具有较高重要性的信息被完整获取，为医生进一步核实病史提供了基础。3 名专家评分的组内相关系数为 0.86，95%置信区间为 0.80～0.91，支持评价结果的一致性。

4 个专科的信息完整度均超过 90%，显示基于专科需求开展状态跟踪和目标选择在本研究病例中均可实施。ClinIST-Med 将“确定采集内容”与“生成问题表达”分开，使临床信息需求直接参与追问决策，形成信息识别、状态更新、目标选择和报告生成的连续流程。

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00027
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

5 结论

在基于真实门诊资料的交互重建中，ClinIST-Med 较 GPT-4o 自由追问和固定问诊提高了信息完整度、关键信息覆盖率及报告临床可用性，减少了重复提问和交互轮次，支持临床信息状态跟踪用于诊前病史采集。

本研究纳入 2 家三甲医院 4 个专科 80 例患者，每个专科 20 例，样本及专科覆盖有限。采用同一批病例重建交互减少了病例差异，但不能充分反映真实患者对不同问题的理解与应答；原始资料未记录或无法匹配的信息也会限制回答重建及评价。临床信息需求图谱和规则库依赖专家知识，其跨专科适用性及动态更新仍需检验。

后续将扩大医疗机构、专科和病例复杂程度的覆盖，开展真实患者参与的前瞻性研究，评价信息采集效果、实际交互时间及对门诊接诊和临床决策的辅助价值。

伦理声明：本研究经海南医科大学伦理委员会和哈尔滨医科大学附属第二医院伦理委员会批准，批准号分别为 HYMLL-2026-109 和 KY2026-157。所有患者均自愿参与并签署知情同意书，数据已做去标识化处理。

利益冲突：所有作者声明不存在利益冲突。

10.12201/bmr.202611.00027V1

参考文献

[1] Bickmore T W, Puskar K, Schlenk E A, et al. Maintaining reality: relational agents for antipsychotic medication adherence[J]. Interact Comput, 2010, 22(4): 276-288.

[2] Hindelang M, Sitaru S, Zink A. Transforming health care through chatbots for medical history-taking and future directions: comprehensive systematic review[J]. JMIR Med Inform, 2024, 12: e56628.

[3] Tu T, Schaekermann M, Palepu A, et al. Towards conversational diagnostic artificial intelligence[J]. Nature, 2025, 642(8067): 442-450.

[4] Kaplan A, Aslantaş Kaplan B, Aydilek A, et al. Spine patient history-taking: a two phase study of a prompt-engineered large language model assistant[J/OL]. Eur Spine J, 2026[2026-11-15]. https://doi.org/10.1007/s00586-026-10094-0.

http://www.biomedrxiv.org.cn/article/doi/10.12201/bmr.202611.00027
此预印本（未经同行评审）的作者拥有该文稿的版权，biomedRxiv拥有永久保存权。任何人未经允许不得重复使用。

[5] 阮彤, 卞俣昂, 余广涯, 等. 医学大语言模型研究与应用综述[J]. 中国卫生信息管理杂志, 2023, 20(6): 853-861.

[6] 王绮婷, 曹博宇, 林浩云, 等. 大语言模型在职业卫生领域的评估[J]. 广东医学, 2025, 46(11): 1628.

[7] 张玉晨, 王传涛, 夏海梁, 等. 基于大语言模型的医疗智能助手应用研究进展[J]. 协和医学杂志, 2025, 16(6): 1511-1518.

[8] Luo M J, Lai Y, Wang W, et al. Standardized pre-consultation by a large language model agent vs ophthalmology residents: a randomized clinical trial[J/OL]. NPJ Digit Med, 2026[2026-11-15]. https://doi.org/10.1038/s41746-026-03232-x.

[9] Tao X, Zhou S, Ding K, et al. An LLM chatbot to facilitate primary-to-specialist care transitions: a randomized controlled trial[J]. Nat Med, 2026, 32(3): 934-942.

10.12201/bmr.202611.00027V1