

· 特约专稿 ·

通信作者



洪申达, 北京大学健康医疗大数据国家研究院新体制助理教授、副研究员、博士生导师。2019 年博士毕业于北京大学智能科学系。研究方向为医疗时序数据的人工智能算法研究及其在临床和智能可穿戴的应用、AI 数字健康。主持国家自然科学基金、CCF-智谱大模型创新基金、CCF-腾讯犀牛鸟基金等项目。以第一或通讯作者在 NEJM AI、The Lancet Digital Health、npj Digital Medicine、Cell Patterns 等期刊和 ICLR、ICML、NeurIPS 等会议发表论文 60 余篇, 被引超过 6000 次。担任 npj Digital Medicine、Health Data Science 期刊副编辑。相关成果已取得医疗器械注册证。

家庭 AI 心电仪的用户与大语言模型对话分析研究

黄舜 章德云 魏爽 洪申达

【摘要】 背景 便携式心电图(ECG)设备的普及在用户与复杂的医疗数据之间造成了解读鸿沟。具备 AI 大语言模型(Large Language Model, LLM)的家庭 AI 心电仪成为弥合此鸿沟的潜在工具。本研究旨在通过分析大规模真实世界数据, 探究用户在获取便携式 ECG 结果后与 AI 大模型的用户行为, 包括对话意图、留存性及参与度。**方法** 本研究对 2024 年 4 月 16 日至 2025 年 8 月 18 日期间, 来源于“问心无恙”家庭 AI 心电仪的 480,532 条真实世界对话记录进行了一项回顾性研究。数据经过清洗与预处理后, 被整合为 109,017 个独立的对话会话。我们对 1% 的对话会话(1091 组)进行随机抽样, 采用“大模型辅助+人工审核”的方法分析用户意图。同时, 我们对全体用户并得到 AI 解读结果后所表现出的用户行为的留存率与对话交互轮数进行了统计分析。**结果** 用户意图分析显示, “解读检验结果”(81.30%)、“治疗建议”(51.15%)和“病因询问”(41.89%)是最高频的三类意图。超过 87% 的对话包含两种及以上的多维意图。用户留存率分析发现, 54.21% 的用户进行了多次(≥ 2 次)对话。从交互深度来看, 大部分对话为 2~5 轮的中度互动(63.5%), 而仅有 1 轮的浅层询问占 24.38%。**结论** 研究证实, AI 大语言模型在帮助用户跨越便携式 ECG 数据解读鸿沟方面正发挥着核心作用。用户表现出以结果解读为核心、多意图交织的复杂交互模式, 并具备一定的长期使用意愿。这些发现为优化数字健康服务提供了实证依据, 同时也警示了平台需高度关注用户对“治疗建议”的高频需求所伴随的潜在风险。

【关键词】 大语言模型; 便携式心电图; 解读鸿沟; 用户行为; 真实世界研究

[中图分类号] R541.7 R540.4 [文献标识码] A [文章编号] 1005-0272(2025)05-321-07

[引用格式] 黄舜, 章德云, 魏爽, 等. 家庭 AI 心电仪的用户与大语言模型对话分析研究 [J]. 临床心电学杂志, 2025, 34(5): 321-327.

A study on user conversations with large language models in home AI-ECG device HUANG Shun^{1, 2}, ZHANG Deyun³, WEI Shuang⁴, HONG Shenda^{1, 2}. 1. National Institute of Health Data Science, Peking University,

基金项目: 国家自然科学基金(编号: 62102008)

作者单位: 100191 北京, 北京大学健康医疗大数据国家研究院(黄舜, 洪申达); 100191 北京, 北京大学医学部医学技术研究院(黄舜, 洪申达); 230088 安徽 合肥, 安徽心之声医疗科技有限公司(章德云); 102200 北京, 中央财经大学外国语学院(魏爽)

作者简介: 黄舜, 硕士, 主要从事临床决策支持 AI 模型开发、真实世界验证与人机交互的研究。章德云, 硕士, 主要从事人工智能生理信号算法、智能可穿戴研究。魏爽, 博士, 中央财经大学外国语学院助理教授, 主要从事医疗话语语用研究。洪申达, 博士, 北京大学助理教授, 主要从事医疗人工智能算法、智能可穿戴研究。

通信作者: 洪申达, E-mail: hongshenda@pku.edu.cn

Beijing 100191, China; 2. Institute of Medical Technology, Peking University Health Science Center, Beijing 100191, China; 3. Anhui HeartVoice Medical Technology, Anhui Hefei 230088, China; 4. School of Foreign Studies, Central University of Finance and Economics, Beijing 102200, China.

【Abstract】 Background The widespread adoption of portable electrocardiogram (ECG) devices has created an interpretation gap between users and complex medical data. Home-based AI ECG devices powered by large language models (LLMs) have emerged as potential tools to bridge this gap. This study aimed to investigate user behaviors—including conversational intents, retention, and engagement—after receiving portable ECG results and interacting with an LLM, using large-scale real-world data. **Methods** We conducted a retrospective study of 480, 532 real-world conversation records collected from the Wenxin Wuyang home AI ECG device between April 16, 2024, and August 18, 2025. After data cleaning and preprocessing, these were consolidated into 109,017 unique conversation sessions. A random 1% sample (1,091 sessions) was selected for user intent analysis using a combined "LLM-assisted + human review" approach. Additionally, we analyzed user retention and the number of conversational turns following receipt of AI-generated ECG interpretations. **Results** The user intent analysis revealed that the three most frequent categories were "interpreting test results" (81.30%), "seeking treatment advice" (51.15%) and "inquiring about professional terms" (41.89%). Over 87% of dialogues contained two or more multi-dimensional intentions. The user retention analysis showed that 54.21% of users engaged in multiple (≥ 2) dialogue sessions. Regarding interaction depth, the majority of dialogues were moderate interactions of 2~5 turns (63.5%), while shallow, single-turn inquiries accounted for 24.38%. **Conclusion** This study confirms that AI large language models play a core role in helping users overcome the interpretation gap of portable ECG data. Users exhibit complex, multi-intent interaction patterns centered on result interpretation and demonstrate a willingness for long-term engagement. These findings provide empirical evidence for optimizing digital health services while also highlighting the potential risks associated with the high user demand for "treatment advice," which platforms must address with caution.

【Keywords】 Large Language Models; Portable ECG; Interpretation Gap; User Behavior; Real-World Study

1 介绍

便携式与可穿戴健康监测设备的普及与发展,尤其是便携式心电图(ECG)设备的兴起,使得原本局限于临床环境中的生理数据采集能力被赋予给了普通消费者。用户由此从过去的被动信息接收者,逐渐转变为自身健康管理的主动参与者。这种赋能正是数字健康革命的核心愿景之一^[1-2]。

然而,当用户面对复杂的医疗数据时,往往缺乏必要的医学训练来理解这些数据或图形所蕴含的临床意义^[3]。数据获取能力与数据解读能力之间由此形成了巨大的落差,即所谓的“解读鸿沟”^[4]。尽管现有的便携式设备通常接入云平台 and AI 计算,可以将结果自动化地呈现给用户^[5-6],但当用户在看到异常结果时,或希望进一步理解背后的数据含义与相关专业概念,乃至寻求额外帮助时,这种鸿沟依然难以真正消除^[4]。

在此背景下,大语言模型(LLM)作为一种新兴解决方案逐渐进入人们的视野。用户与 LLM 的对话,成为其弥合解读鸿沟的关键途径之一,甚至展现出同理心并回应用户^[7]。这一行为模式的出现,一方

面为提升公众健康素养与实现个性化健康管理带来了前所未有的机遇^[8];另一方面,这种交互模式也带来了新的挑战。业界与学界普遍关注 LLM 回答的准确性、安全性,及其作为非医疗建议的固有限制所带来的潜在风险^[9-10]。因此,在深入评估这些具体风险之前,分析用户与这类工具的实际交互方式与核心诉求,便成为一项基础且必要的前提。

然而,针对用户在使用便携式设备后与 LLM 展开的此类特定对话,其实际交互情况和用户行为分析,目前仍缺乏大规模的实证研究。现有研究更多聚焦于 LLM 在专业临床场景中的应用,或从宏观层面探讨其在医疗领域中的多种任务^[11-12]。因此,本研究聚焦于普通用户在使用便携式心电图仪获取 ECG 数据并得到 AI 解读结果后所表现出的用户行为,具体分析其与 LLM 对话中的用户意图、对话频次及交互模式。

2 方法

2.1 数据来源和预处理

本项目是一项基于真实世界对话数据的回顾性研究。数据集来源于“问心无恙”家庭 AI 心电图数据库,该数据库收录了 2024 年 4 月 16 日至 2025 年 8

月 18 日期间,中国境内用户与“问心无恙”AI 大模型交互产生的 480,532 条对话记录。

为确保数据质量与分析的有效性,我们对原始数据进行了预处理。首先,移除了所有由系统单方面发起且用户未作任何回应的对话记录,并采用规则与关键词匹配的方法,滤除了内容明显无意义的测试性和无效对话(例如仅包含“Test”、“测试”或单个标点符号的条目)和信息不全的对话。

随后,为了更好地分析连续的交互行为,我们将数据以对话会话为单位进行分组,即把同一用户 ID 在 30 分钟内产生的一轮或多轮对话定义为一个独立的对话组。若用户两次发言的时间间隔超过 30 分钟,则后者被视为一个新会话的开始。本研究方案已通过北京大学医学部伦理委员会的审查与批准(批准号:IRB00001052-23071),并获准豁免签署知情同意书。

2.2 随机抽样与用户意图分析

为了深入探究用户与 AI 大模型交互的核心意图,我们对预处理后的对话组数数据集进行了 1% 的随机抽样,该比例旨在保证样本代表性的同时,兼顾人工审核的成本与可行性。我们采用“大模型辅助+人工审核”的混合方法对抽样数据进行意图分类。具体而言,我们首先利用智谱 AI 的 GLM-4.5 模型,对

每个对话会话中用户的所以提问进行自动分类。接着,由研究人员对模型进行分类结果进行审核,以确保最终分类结果的准确性和可靠性。在完成所有抽样数据的意图分类后,我们采用描述性统计方法,分析了不同用户意图的频率分布情况。

2.3 用户留存与参与度分析

为评估用户对 AI 大模型的长期使用情况与单次交互的深度,我们从用户留存率与用户参与度两个维度进行了分析。用户留存率被定义为在研究周期内发起过一次以上(≥ 2 次)独立对话会话的用户比例,通过计算每个独立用户 ID 对应的对话会话数量进行统计。用户参与度则以单次对话会话中的交互轮数作为核心衡量指标,其中一个交互轮次被定义为一次用户提问及一次模型应答。基于交互轮数,我们将用户参与度划分为四个等级:1 轮为浅层询问,通常是简单的信息获取;2~5 轮为中度互动,用户存在数次追问或澄清;6~9 轮为深度互动,用户表现出持续的探索意愿;10 轮及以上则为高参与度互动,对话进行了广泛或深入的挖掘。

3 结果

3.1 数据清洗与分组

数据清洗与分组结果如图 1 所示。在最初的

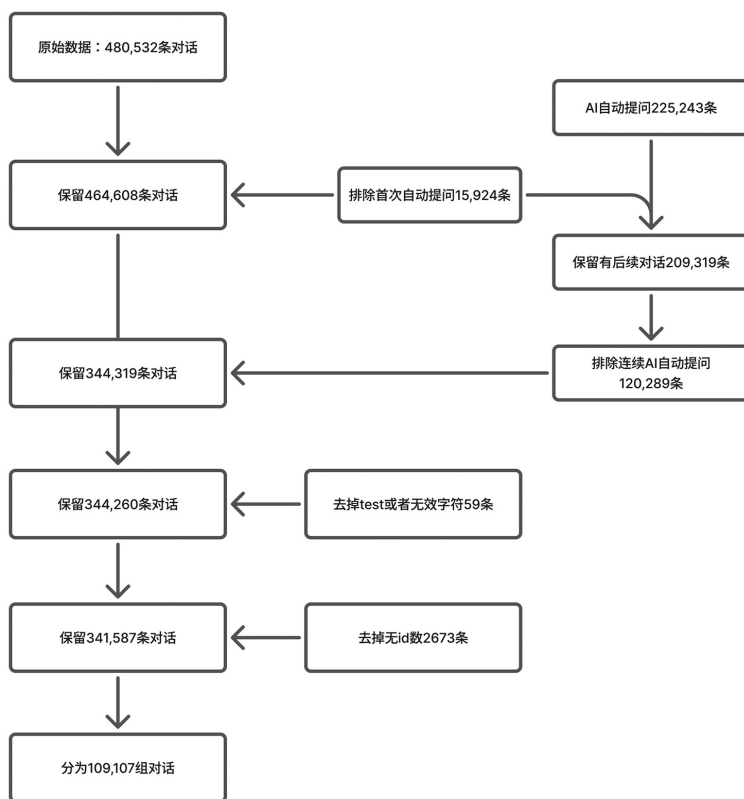


图 1 数据清洗与分组流程

480,532 条对话记录中,系统单次发起的提问占 15,924 条,系统连续发起的提问为 120,289 条。此外,还发现到 59 条无效对话以及 2,673 条信息不全对话,均在清理过程中予以剔除。

经过上述步骤,最终保留 341,587 条有效对话记录。在此基础上,以“同一用户在 30 分钟内的对话”为一轮进行划分,从而将保留数据进一步组织为 109,017 组对话。

3.2 用户意图分析

在对对话组进行 1% 的随机抽样后,最终获得了 1091 组对话。基于 GLM-4.5 模型的分类结果,所有

提问内容被划分为八类,分别为:解读检验结果、专业名词询问、治疗建议、病因询问、预防建议、症状询问、生活/饮食注意以及其他。经人工复核后,具体分布如图 2 所示:其中“解读检验结果”数量最多,为 887 组,占比 81.30%;其次为“治疗建议”,558 组,占比 51.15%;“病因询问”457 组,占比 41.89%;“专业名词询问”347 组,占比 31.81%;“其他”(包括不相关话题或未能归类的内容)345 组,占比 31.62%;“症状询问”236 组,占比 21.63%;“预防建议”145 组,占比 13.29%;而“生活/饮食注意”共 144 组,占比 13.20%。

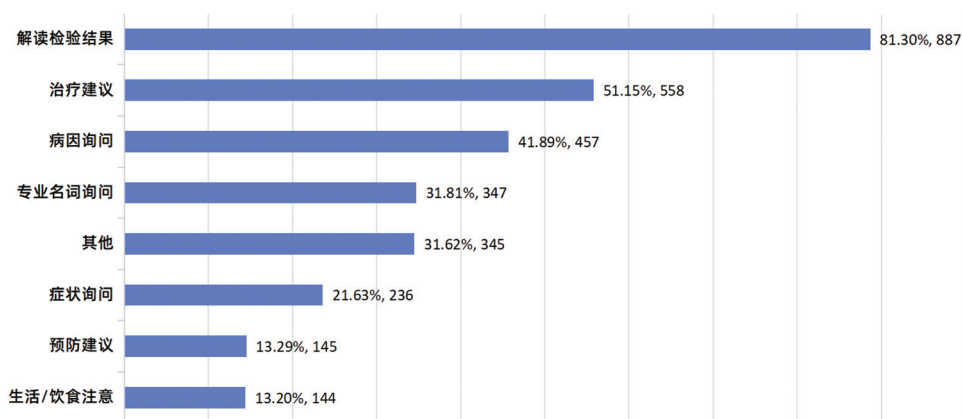


图 2 抽样提问类型分布和统计

进一步从对话组的类型分布来看(表 1),用户往往在一次对话中探讨多个维度的问题。统计结果显示:仅包含 1 种提问类型的对话为 236 组,占比 21.63%;包含 2 种类型的为 325 组,占比 29.79%;包含 3 种类型的为 283 组,占比 25.94%;包含 4 种类型的为 162 组,占比 14.85%;而包含 5 种及以上类型的则有 85 组,占比 7.79%。

表 1 抽样提问维度统计

提问维度	数量(组)	占比(%)
1 种提问类型	236	21.63
2 种提问类型	325	29.79
3 种提问类型	283	25.94
4 种提问类型	162	14.85
5 种及以上提问类型	85	7.79

3.3 用户留存率与参与人数

根据用户 ID 的统计结果,共计筛选出 26,419 位用户。从总体分布来看,单次对话的用户数为 12,097 人,占有所有用户的 45.79%;而进行多次对话的用户则有 14,322 人,占比为 54.21%。从平均水平来看,每位

用户与模型的交互次数为 4.13 次,标准差为 8.19,中位数为 2 次,其最高对话次数达到 580 次,显示出显著的离群情况。

对于交互轮数来说,平均每组对话包含 3.13 轮,标准差为 2.81,中位数为 2 轮,而单组对话的最大轮数可达 108 轮。其中仅有 1 轮对话的组数为 26,578 组,占总对话数的 24.38%,显示出相当比例的浅层互动;在 2 至 5 轮之间的中度互动占比最高,共有 69,230 组,占比达 63.5%,表明绝大部分对话属于有限往返的交流;而包含 6 至 9 轮的组数为 9,838 组,占比 9.03%,表现为一定程度的深入探讨;至于 10 轮及以上的高度参与对话,则仅有 3,371 组,占比最低,仅为 3.09%。

4 讨论

本研究首次通过大规模真实世界数据,系统地分析了普通用户在获得便携式心电图(ECG)结果后并得到 AI 解读结果后所表现出的用户行为,包括意图及参与模式。研究结果清晰地揭示了,大语言模型在弥合普通用户与复杂医疗数据之间的解读鸿沟方面,正扮演着至关重要的角色。

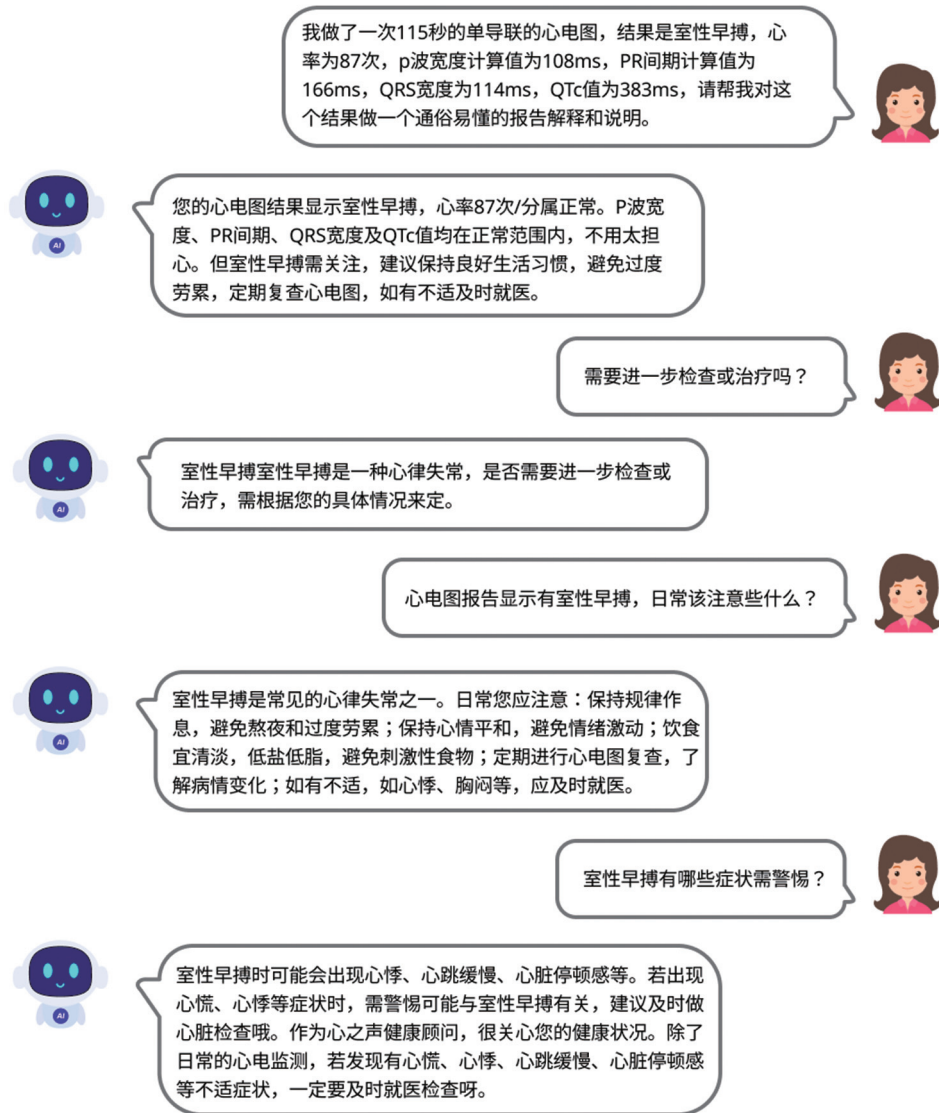


图 3 对话案例展示

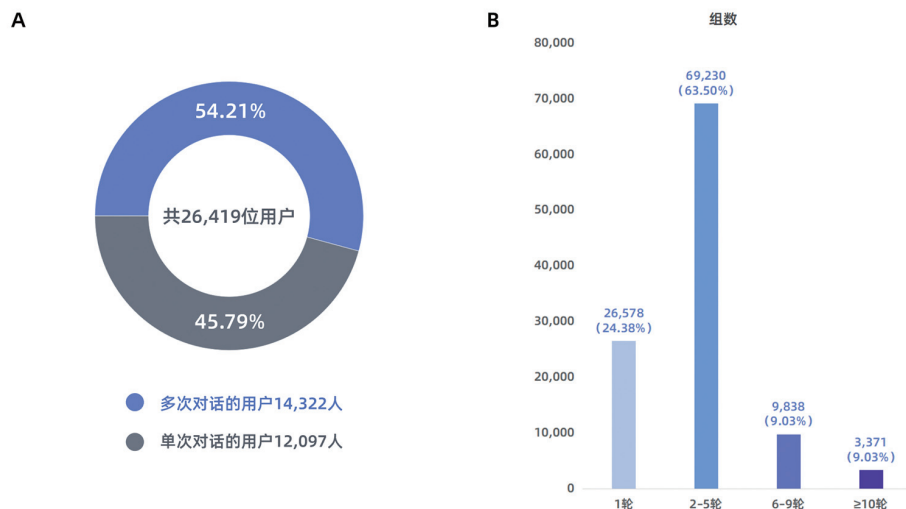


图 4 A: 用户留存率统计;B: 对话轮数统计

4.1 核心发现与解读

在高达 81.30%的对话意图集中于“解读检验结果”,这直接印证了用户在获取生理数据后,迫切需要专业化的解读支持。这表明,AI 对话系统已成为用户获取即时、个性化健康信息解读的首选工具之一。与此同时,“治疗建议”(51.15%)和“病因询问”(41.89%)紧随其后,说明用户的需求并未停留在简单的结果层面,而是呈现出向深度理解和寻求行动指导延伸的趋势。

一个值得关注的现象是,超过 78%的对话会话包含两种或以上的意图类型。这表明用户的咨询过程是多维度的、非线性的。他们倾向于在一次连续的对话中,从解读报告出发,逐步深入到名词解释、病因探寻、预防措施乃至生活方式的调整。这种复杂的交互模式,对模型的多轮对话能力、上下文理解能力以及知识整合能力提出了极高的要求,也凸显了将 LLM 作为一站式健康咨询入口的巨大潜力。

在用户粘性方面,超过半数(54.21%)的用户进行了多次对话,这说明该 AI 模型并非一次性查询工具,而是成功地在一定程度上扮演了长期健康伴侣的角色,用户愿意在产生新的疑问时再次求助。然而,从交互深度来看,大部分对话(63.5%)属于 2~5 轮的中度互动,而深度互动(≥ 6 轮)的比例相对较低。这可能反映了当前用户的主流使用习惯:他们倾向于通过数轮追问解决核心疑惑,但较少进行非常深入或广泛的探索。同时,24.38%的“浅层问询”也说明,仍有大量用户仅将其用于快速获取答案。

4.2 研究的意义与启示

本研究的发现对于数字健康服务的设计与优化具有重要的实践意义。首先,它强调了以“结果解读”为核心功能的重要性,开发者应持续优化模型对 ECG 报告、关键指标及异常提示的理解与解释能力。其次,鉴于用户意图的多样性与关联性,模型应被设计为能够主动引导和拓展对话,例如在解释完相关疾病后,可以主动提供相关病因、预防建议或生活注意事项的选项,形成更智能、更全面的服务闭环。

此外,用户对“治疗建议”的高频次需求是一个需要高度警惕的信号。尽管大语言模型在技术上可以生成看似合理的建议,但可能并潜藏着巨大的安全风险^[13]。这提示平台方必须在产品设计中加入更强、更明确的风险提示与非医疗建议声明,并探索如何安全、合规地将用户引导至专业的医疗服务系统。

4.3 研究的局限性与未来研究方向

本研究也存在一定的局限性。首先本研究是一项回顾性研究,我们分析了用户的提问行为,但未能评估模型回答的准确性、安全性以及用户对回答的满意度。其次,我们未能获取用户的具体人口学信息(如年龄、性别、教育背景)及临床背景(如既往病史),这些变量可能对用户提问行为产生重要影响。而意图分析基于 1%的抽样数据,尽管保证了可行性,但仍可能存在一定的抽样偏差。

未来的研究可以从以下几个方面展开:一是对模型回答的质量进行系统性评估,特别是其准确性、完整性和安全性,这是衡量该技术应用价值的关键。二是开展前瞻性研究或用户调研,通过问卷、访谈等形式,深入了解用户的使用动机、满意度以及 AI 的回答对其健康认知和行为的实际影响。三是结合用户画像数据,分析不同特征的用户群体在交互行为上是否存在差异,从而为提供更加个性化、精准的服务提供依据。最后,探索人机协同的新模式,研究如何将 AI 对话系统更有效地整合到现有的医疗服务流程中,使其成为医生工作的有效补充,而非潜在的风险来源。

总而言之,本研究通过对大规模真实世界对话数据的分析,证实了 AI 大语言模型在帮助用户跨越便携式 ECG“解读鸿沟”方面发挥着积极且核心的作用。用户展现出以结果解读为核心、多意图交织的复杂交互模式,并表现出一定的长期使用意愿。这些发现为优化现有数字健康产品、提升用户健康素养提供了宝贵的实证依据,同时也警示我们必须高度关注其在应用中伴随的伦理与安全挑战。

参考文献

- [1] HUGHES A, SHANDHI M M H, MASTER H, et al. Wearable devices in cardiovascular medicine [J]. *Circ Res*, 2023, 132(5): 652-670.
- [2] FU Z J, HONG S D, ZHANG R, et al. Artificial-intelligence-enhanced mobile system for cardiovascular health management[J]. *Sensors (Basel)*, 2021, 21(3):773.
- [3] SMITH B, MAGNANI J W. New technologies, new disparities: The intersection of electronic health and digital health literacy[J]. *Int J Cardiol*, 2019, 292: 280-282.
- [4] CANALI S, SCHIAFFONATI V, ALIVERTI A. Challenges and recommendations for wearable devices in digital health: Data quality, interoperability, health equity, fairness [J]. *PLoS Digit Health*, 2022, 1(10): e0000104.

- [5] HONG S D, FU Z J, ZHOU R B, et al. CardioLearn: a cloud deep learning service for cardiac disease detection from electrocardiogram [C]// In Companion Proceedings of the Web Conference 2020. Taipei, Taiwan, 2020.
- [6] SIONTIS K C, NOSEWORTHY P A, ATTIA Z I, et al. Artificial intelligence -enhanced electrocardiography in cardiovascular disease management[J]. Nat Rev Cardiol, 2021, 18(7): 465-478.
- [7] AYERS J W, POLIAK A, DREDZE M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum[J]. JAMA Intern Med, 2023, 183(6): 589-596.
- [8] HUO B, BOYLE A, MARFO N, et al. Large language models for chatbot health advice studies: a systematic review [J]. JAMA Netw Open, 2025, 8(2): e2457879.
- [9] BEDI S, LIU Y T, ORR-EWING L, et al. Testing and evaluation of health care applications of large language models: a systematic review[J]. JAMA, 2025, 333(4): 319-328.
- [10] JOHRI S, JEONG J, TRAN B A, et al. An evaluation framework for clinical use of large language models in patient interaction tasks[J]. Nat Med, 2025, 31(1): 77-86.
- [11] WANG Z F, WANG H Y, DANEK B, et al. A perspective for adapting generalist AI to specialized medical AI applications and their challenges[J]. NPJ Digit Med, 2025, 8(1): 429.
- [12] IQBAL U, TANWEER A, RAHMANTI A R, et al. Impact of large language model (ChatGPT) in healthcare: an umbrella review and evidence synthesis[J]. J Biomed Sci, 2025, 32(1):45.
- [13] MESKÓ B, TOPOL E J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare [J]. NPJ Digit Med, 2023, 6(1):120.

(收稿日期:2025-09-15)

作者·读者·编者

《临床心电学杂志》官方网站及投审稿系统 上线通知

尊敬的各位教授、各位主任、广大作者及读者：

你们好！为更好地服务心电学学术交流与成果传播，《临床心电学杂志》全新官方网站即将正式上线，网站内置的智能投审稿系统也将同步开放，现特此通知：

1. 上线及使用时间：官方网站与智能投审稿系统，预计于 2025 年 11 月上旬同步开放访问、登录及投稿功能，届时可通过官网开展全流程投稿操作。

2. 核心功能：登录官网后，不仅可通过智能投审稿系统完成稿件提交、进度查询、编辑沟通、录用通知接收等操作，还可便捷获取杂志最新刊期内容、最新期刊资讯动态、往期文献检索等专业服务，助力学术研究与交流。

3. 过渡安排：网站上线前，如有紧急投稿需求，可暂通过原联系方式（投稿邮箱 lcxzxz_2046@126.com）与编辑部对接；官网上线后，将同步发布详细的网站导航及投审稿系统操作指引，确保顺利使用。

最后感谢您长期以来对《临床心电学杂志》的支持与关注，期待通过全新官方平台，与您共同推动心电学领域学术发展！

《临床心电学杂志》编辑部

2025 年 10 月 26 日