

多智能体协同驱动复杂任务解决的学习环境框架构建及分析方法研究——以医 学生临床推理为例

Building an Environment Framework and Process-Analytic Methods for Multi-Agent Collaboration-Driven Complex Task Solving: A Case Study of Clinical Reasoning in Medical Students

许家奇^{1*}, 唐陆祺¹, 林常敏², 陈娜平^{2*}

¹ 北京大学教育学院

² 汕头大学医学院

*xjq@pku.edu.cn; s_chennp@stu.edu.cn

【摘要】 生成式人工智能为模拟复杂真实任务情境、培养学习者高阶能力提供了新范式。然而，在人机协同解决问题的过程中，学习者如何与由多个 AI 智能体组成的“虚拟团队”进行深度交互、动态调整其认知策略，仍是当前研究的盲区。本研究面向医学生临床推理练习，构建由 AI 病人、AI 体格查汇报员、AI 辅助检查汇报员、AI 助手与 AI 带教医生组成的多智能体协同学习环境，构建融合行为日志、眼动、面部表情、出声思维与刺激回忆的多模态学习分析框架。研究期望为探究人与多智能体协同中的信息搜寻、假设建构与反思修正等认知与元认知策略提供新的方法论视角与分析思路。

【关键词】 多智能体协同；复杂任务解决；临床推理；教学环境设计

Abstract: Generative artificial intelligence has introduced a new paradigm for simulating authentic complex-task contexts and fostering learners' higher-order competencies. However, in human-AI collaborative problem solving, how learners engage in deep interaction with a virtual team composed of multiple AI agents and dynamically adapt their cognitive strategies remains a major blind spot in current research. Focusing on medical students' clinical reasoning practice, this study develops a multi-agent collaborative learning environment comprising an AI patient, a AI physical-examination reporter, an AI auxiliary-test reporter, an AI assistant, and an AI clinical tutor. We further propose a multimodal learning analytics framework that integrates behavioral logs, eye-tracking data, think-aloud protocols, stimulated recall, and facial expressions. This work aims to offer a novel methodological lens and analytic approach for investigating the cognitive and metacognitive strategies, such as information seeking, hypothesis construction, and reflective revision, that emerge in human-multi-agent collaboration.

Keywords: multi-agent collaboration, complex task solving, clinical reasoning, instructional environment design

1. 前言

生成式人工智能正以其强大的情境生成与自然语言交互能力，深刻重塑教育领域，尤其为模拟真实世界的复杂任务提供了高保真度。传统教学模式在培养如临床问诊、工程设计等开放式、高阶认知技能时，常面临情境固化、成本高昂与规模化瓶颈等问题(Chernikova et al., 2020)。利用人工智能构建的学习环境既有助于降低构建与反馈供给的成本，也能扩展学习者演练与反思的空间。然而，早期的智能学习环境设计易陷入“知识投喂”的误区，学习者容易把关键推理环节外包给 AI 系统，从而产生“元认知懒惰”(Fan et al., 2025)。多智能体协同为破解这一困境提供了有效的路径。它不再将生成式人工智能视作单一的答案输出端，而是通过

配置多个具备不同职责与交互规则的智能体,模拟现实任务中分工协作的“虚拟团队”,从而建构更具生态效度的社会性与认知性复杂情境(于济凡 et al., 2024)。在医学教育中,这意味着可以构建一个包含“AI 病人”“AI 带教医生”等角色的虚拟诊室。这种设计既能利用“AI 助手”分担学习者在信息检索上的认知负荷,又能借助“AI 带教医生”提供苏格拉底式的提问,激发其元认知反思,从而在减负与增智之间取得平衡。

本研究提出并阐述面向临床问诊复杂任务解决的多智能体协同环境构建框架,同时提出用于解析人-多智能体协同学习过程的多模态分析方法,将行为日志、眼动、出声思维、表情及刺激回忆访谈等过程数据对齐,捕捉学习者在信息搜寻、假设建构、证据权衡与元认知监控中的动态演化,为深度理解并优化人-多智能体协同学习过程提供一套系统性的研究范式。

2. 研究现状

2.1. 多智能体协同在教育中的应用

随着生成式人工智能的飞速发展,智能体技术正深刻重塑人机交互的范式,在智慧学习环境中展现出巨大的应用潜力。这一演进过程大致可分为两个阶段,一是单智能体作为学习助手的发展阶段。基于大语言模型的智能体凭借其强大的自然语言理解、生成与推理能力,已经能够扮演智能导师、学习同伴等多种角色(黄昌勤 等, 2025)。与依赖预设知识库和规则的传统智能导学系统不同,现代智能体能够提供更具有启发性、即时性和个性化的反馈,甚至可以感知学习者的部分行为特征来动态调整教学内容(Cheng et al., 2024)。这标志着智能体正从一个被动的信息查询工具,转变为一个能够与学习者进行深度认知对话的伙伴。二是多智能体协同构建学习场景阶段。通过为不同智能体赋予独特的角色、目标与行为模式,能够在虚拟环境中模拟复杂的社会互动。一方面,多智能体协同能通过分布式协作支持复杂问题的解决,让学习者在观摩和参与中学习团队合作与策略制定(郭鑫 等, 2023);另一方面,通过创设逼真的社会交互情境,为学习者提供演练沟通、协商等社会情感技能的场所(翟雪松 等, 2024)。

2.2. 学习者与多智能体协同的分析方法

要深入理解并优化智慧学习环境中学习者与智能体的协作过程,须依托科学、多维度的分析方法。首先,主观报告法是探究学习者体验与心理感受的传统工具。但其不足在于依赖事后回忆,难以捕捉动态交互中的即时反应。为弥补这一缺陷,研究者引入了出声思维法与刺激回忆法。前者要求学习者实时出声表达其思维过程,有助于揭示策略形成与困难点(Lim et al., 2021);后者则通过回放交互录像或眼动轨迹,引导学习者对关键事件进行反思和解释,实现客观行为与内在心智的精确匹配。其次,生理计算方法为揭示隐性认知过程提供了实时客观的证据。眼动追踪通过注视点、扫视路径、瞳孔变化等指标,量化学习者对智能体信息的关注模式及界面搜索效率;脑电技术则凭借毫秒级精度反映大脑的认知负荷与情绪波动,不同频段的能量变化可指示注意水平与情感效价(王忠民 等, 2022)。最后,基于多模态协同分析,即整合问卷、访谈、眼动、行为日志等多源数据能够挖掘数据间的互补关联,构建学习过程的立体化、全景式视图。

3. 多智能体协同学习环境构建及分析方法框架

3.1. 多智能体协同学习环境的主要内容

本研究的多智能体学习环境框架以“复杂任务链构建 + 角色分工互补 + 学习者可控介入”为设计主线,面向问诊任务设计包括病史采集、检查资料获取、诊断决策与汇报撰写的连续流程,并嵌入结构化推理外显工具以支持过程记录与反思,主要内容如下:

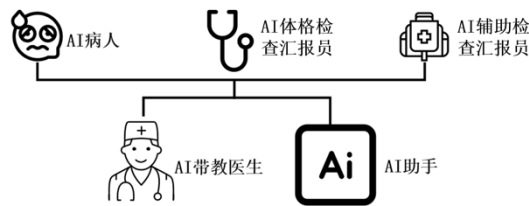


图1 多智能体协同学习环境

(1) AI病人(情境呈现者)。基于真实脱敏病例数据和提示词工程构建。具备个性化的人设、情绪反应和病史记忆，负责在病史采集阶段与学生进行自然语言对话，提供非结构化的临床线索。(2) AI带教医生(元认知脚手架)。AI带教医生基于临床推理对比分析表，以及学生跟病人、AI体格/辅助检查汇报员的历史聊天记录综合进行反馈。通过嵌入反思问题与提示反馈等多种方式，AI带教医生鼓励学生主动回顾与监控自己的推理路径，识别思维中的盲点与漏洞。(3) AI助手。定位为认知任务的卸载工具。它承担的是信息检索和数据处理的工作，而非代替学生的推理与决策。通过这种设计，学生将降低在任务处理中的认知负荷，将更多精力专注于需要深度思考的问题。(4) AI体格/辅助检查汇报员。模拟诊疗流程中的检查过程，学生根据初步判断主动开具检查单，智能体反馈相应的检查结果。

3.2. 多模态数据分析框架

为了系统性地解码学习者在多智能体环境中的复杂认知过程，本研究构建了一个整合数据采集、处理与融合分析的多模态数据分析框架，如图2所示。该框架旨在将外显的行为操作与内隐的认知及情感状态进行时序关联，从而精细刻画学习者的动态学习轨迹。

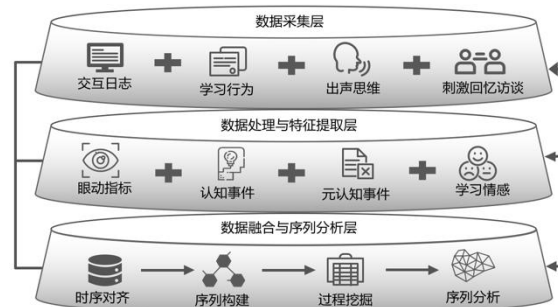


图2 多模态数据分析框架

3.2.1 数据采集层

本框架的核心在于同步捕获一个围绕问诊任务的、富含上下文的多源数据档案。首先涵盖两类基础数据：高层级的交互日志和行为追踪数据。其次，同步生成一份复合视频，包括学习者的屏幕活动、实时的眼动轨迹以及面部视频。任务结束后，进行刺激回忆访谈，即与学习者共同回看这份复合视频，让他们对关键决策点背后的思考过程进行解释与澄清。

3.2.2 数据处理与特征提取层

此阶段旨在将原始数据转化为可供分析的结构化特征。首先，从眼动数据中提取注视时长、注意力转移路径及瞳孔直径变化等量化客观指标。其次，在基于多源证据的认知过程编码中，通过对出声思维、眼动轨迹和交互行为视频的互证，将学习过程操作化为一系列认知、元认知和情感事件。多维度编码体系主要包括：(1) 基础操作，如信息查询、浏览、记录等基本行为；(2) 认知维度，包括假设的生成与检验、证据的搜集与评估、基于证据链的推理与决策。(3) 元认知维度，侧重于监控活动，如评估理解程度、识别认知冲突等。(4) 情绪维度，通过面部表情和语调识别出的困惑、专注、顿悟等情感状态。

3.2.3 数据融合与序列分析层

将特征数据流在统一的时间轴上进行时序对齐，形成高密度事件序列。随后，运用过程

挖掘技术对比不同学习者的策略路径差异。在微观层面,采用延迟序列分析等方法,量化分析特定事件间的转移概率和典型模式。本框架得以将学生的行为、注意、情感和反思融为一体,构建出关于其如何在人与多智能体协同中学习与思考的、有深度证据支持的精细化画像。

4. 讨论及展望

本研究面向医学生临床推理这一高不确定、强情境依赖的复杂任务,提出多智能体协同学习环境和多模态学习分析的研究方案。与以往将生成式人工智能作为单一答疑工具的做法不同,本研究强调通过角色分工与信息互通来重建近似真实的临床互动结构。AI病人负责呈现叙事线索并支持追问,体格检查与辅助检查汇报员提供客观证据,医学知识助手承担检索与解释以降低外在操作负担,而带教医生智能体聚焦推理过程的外显、反思与分层引导。这样的设计取向把“做任务”与“想任务”组织在同一工作流中,为大规模、可复盘的临床推理训练提供可实施路径。本研究提出多模态数据分析框架,以支撑对“人-多智能体协同过程”的过程性研究。通过时序对齐行为、眼动、出声思维与情感数据,打造可解释的证据链。

本研究为未来的人与多智能体协同学习开启了更广阔的探索空间。当前的多模态框架主要服务于研究验证,而未来的核心挑战在于,能否将分析得出的关键过程特征,如注意力不集中、证据-假设脱节等低效模式,转化为实时运行的学习者诊断模型。这将使“AI带教医生”等智能体能够基于对学习者的当前认知状态的深度理解,触发真正个性化、时效性的引导与反馈,从而构建一个能够动态响应学习过程的闭环系统。

基金项目:本文为北京市自然科学基金青年项目“智慧学习环境中融合视觉健康的具身化人-智能体交互机制与关键技术研究”(项目编号:4264119);2025年度广东省本科高校教学质量与教学改革工程建设项目(粤教高函〔2026〕4号,678)的阶段性研究成果。

参考文献

- Cheng, Y., Zhang, C., Zhang, Z., Meng, X., Hong, S., Li, W., Wang, Z., Wang, Z., Yin, F., Zhao, J., & He, X. (2024). *Exploring large language model based intelligent agents: Definitions, methods, and prospects* (No. arXiv:2401.03428).
- Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., Fischer, F., & DFG Research group COSIMA. (2020). Facilitating diagnostic competences in higher education—A meta-analysis in medical and teacher education. *Educational Psychology Review*, 32(1), 157–196.
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489–530.
- Lim, L., Bannert, M., van der Graaf, J., Molenaar, I., Fan, Y., Kilgour, J., Moore, J., & Gašević, D. (2021). Temporal assessment of self-regulated learning by mining students' think-aloud protocols. *Frontiers in Psychology*, 12.
- 于济凡, 李睿森, 李曼丽, & 刘惠琴. (2024). 多智能体协同交互的高临场感在线学习环境构建. *现代教育技术*, 34(12), 17–26.
- 王忠民, 赵玉鹏, 郑榕林, 贺炎, 张嘉雯, & 刘洋. (2022). 脑电信号情绪识别研究综述. *计算机科学与探索*, 16(4), 760–774.
- 翟雪松, 季爽, 焦丽珍, 朱强, & 王丽英. (2024). 基于多智能体的人机协同解决复杂学习问题实证研究. *开放教育研究*, 30(3), 63–73.
- 郭鑫, 王微, 青伟, 李剑, & 何召锋. (2023). 基于强化学习的多智能体泛化性研究. *计算机技术与发展*, 33(4), 114–119.

黄昌勤, 钟益华, 王希哲, 韩中美, & 魏同权. (2025). 从单智能体到多智能体: 大模型智能体支持下的激励型学习活动设计与实证研究. 华东师范大学学报 (教育科学版), 43(5), 44-56.