GAI问诊模型在诊断学教学中的实践探索与效果分析

——与传统床旁教学的对照研究

杨泽利[1]，时晓芳[1]，梁文霞[1]，汪羚利[2]

1.北京市延庆区医院（北京大学第三医院延庆医院），首都医科大学延庆教学医院全科医学科，北京 102100；
2.北京大学第三医院血液内科，北京 100191

【摘要】 本研究旨在探讨生成式人工智能问诊模型在诊断学教学中替代传统床旁教学（基于病房患者）的可行性及其效果。研究选取首都医科大学延庆教学医院2023、2024级“3＋2”助理全科医师规范化培训的医学生为研究对象，共73人。其中，2023级学生及2024级甲组学生共38人设为对照组；2024级乙组及丙组学生共35人为试验组。试验组采用基于Claude 2的生成式人工智能（generative artificial intelligence，GAI）问诊模型进行问诊训练，对照组则采用传统床旁教学（基于病房患者）进行问诊实践。通过教师评价和学生自评两种方式，比较两组教学效果。研究结果显示，教师评价结果显示，试验组学生在问诊流程[（91.40±2.89）分 vs.（89.34±4.12）分]和病史采集[（91.31±2.35）分 vs.（88.55±3.45）分]两个维度的得分均高于对照组，差异具有统计学意义（均$P<0.05$）。在临床思维方面，两组得分亦存在差异[（90.34±2.21）分vs.（89.29±3.30）分，$P=0.048$]。其中，试验组在人文关怀维度的提升最为突出，得分显著高于对照组[（97.86±1.46）分 vs.（93.71±2.88）分，$P<0.001$]。学生自评结果显示，两组在专业知识掌握、临床思维能力、医患沟通能力等方面两组差异无统计学意义（均$P>0.05$）。研究结果表明，基于Claude 2的GAI问诊模型能够有效提升学生在问诊流程规范性、病史采集完整性及人文关怀行为演练方面的表现，学生对该教学方法及效果的满意度也更高，其整体教学效果与传统床旁教学相当，可作为现有教学资源的有益补充进行推广应用。

【关键词】 人工智能；诊断学；传统床旁教学；问诊教学

【中图分类号】 R-4；G642.2 **【文献标志码】** A **【文章编号】** 1004-5287(2026)04-0436-06

【DOI】 10.13566/j.cnki.cmet.cn61-1317/g4.202604002

Practice exploration and effectiveness analysis of GAI consultation models in Diagnostics teaching
——A comparative study with traditional bedside teaching

YANG Zeli[1], *SHI Xiaofang*[1], *LIANG Wenxia*[1], *WANG Lingli*[2]

1. *Department of General Medicine, Beijing Yanqing District Hospital (Peking University Third Hospital Yanqing Hospital), Capital Medical University Yanqing Teaching Hospital, Beijing* 102100, *China*;
2. *Department of Hematology, Peking University Third Hospital, Beijing* 100191, *China*

【Abstract】 This study aimed to evaluate the feasibility and effectiveness of replacing traditional bedside teaching (based on ward patients) with a generative artificial intelligence consultation model in Diagnostics education. A total of 73 medical students enrolled in the “3＋2” assistant general practitioner standardized training program from the class of 2023 and 2024 at Capital Medical University Yanq-

基金项目：北京大学第三医院教学研究专项课题（2024bysyxykzp02）

收稿日期：2025-10-14

作者简介：杨泽利，主治医师，本科，研究方向为老年患者常见病和多发病的诊治及管理。

通信作者：汪羚利，主治医师，博士，研究方向为血液科的常见病。E-mail：lingli@bjmu.edu.cn

ing Teaching Hospital were selected. Students from the class of 2023 and group A of the class of 2024 (n=38) served as the control group, while groups B and C of the class of 2024 (n=35) constituted the experimental group. The experimental group received consultation training using a GAI (generative artificial intelligence) consultation model based on Claude 2, whereas the control group underwent traditional bedside teaching with ward patients. Teaching outcomes were assessed through both teachers' evaluation and students' self-evaluation. Teachers' evaluation demonstrated that the experimental group outperformed the control group in consultation procedure [(91.40±2.89) vs. (89.34±4.12)] and history-taking [(91.31±2.35) vs. (88.55±3.45)], with both differences reaching statistical significance (both P<0.05). A between-group difference was also observed in clinical reasoning [(90.34±2.21) vs. (89.29±3.30), P=0.048]. Notably, the most pronounced improvement in the experimental group was found in humanistic care [(97.86±1.46) vs. (93.71±2.88), P<0.001]. However, students' self-evaluation revealed no statistically significant differences between the two groups in terms of professional knowledge mastery, clinical reasoning ability, or doctor-patient communication skills (all P>0.05). According to the results, the GAI consultation model based on Claude 2 effectively improved students' performance in consultation procedure standardization, history-taking completeness, and humanistic care practice, while also receiving higher student satisfaction. Its overall teaching effectiveness was comparable to that of traditional bedside teaching, suggesting that it can serve as a valuable supplementary tool to existing teaching resources for broader application.

【Keywords】 artificial intelligence; Diagnostics; traditional bedside teaching; diagnostic interview teaching

问诊是临床诊断的核心技能，其教学效果直接影响医学生的临床胜任力。已有研究表明，传统课堂教学存在理论与实践脱节现象，仅37.6%的学生能独立完成系统性病史采集[1]。为弥补这一缺陷，标准化病人（standardized patient，SP）教学模式由Barrows教授系统阐述后被广泛推广[2]，国内学者尝试通过医学生互扮SP提升教学参与度，但在考核中发现症状模拟失真率高达21.3%[3]。临床数据显示，SP模式使问诊考核平均分提升了18.7分，却需要教师额外投入46.2%的备课时间用于病例编写和SP培训[4]。为解决SP培训成本高、学生实践机会不足的瓶颈，计算机问诊系统通过数字化病例库和自动化交互（如文本问答与语音识别技术），实现了问诊训练的时空解耦与标准化评估[5]，但交互式人工智能系统仍存在15.4%的逻辑跳转错误[6]。基于虚拟标准化病人（virtual standardized patient，VSP）的思维训练系统创新性地将症状关联性分析融入教学，使病史采集完整度提升至82.4%[7]。在此基础上，本研究结合生成式人工智能（generative artificial intelligence，GAI）技术，构建了动态病例生成模型——通过分析真实病例库模拟疾病自然进程，并依托医学本体库实现的语义理解引擎完成症状关联分析与逻辑验证，确保问诊中症状演化的病理合理性[8]，为问诊教学提供新型数字化解决方案。

首都医科大学延庆教学医院（以下简称“我院”）作为助理全科医师规范化培训的教学基地，学生将来主要工作于医疗条件艰苦、检查设备不完善的基层卫生院，因此临床技能尤为重要。我院此前采用“集中讲授＋分组床旁（基于患者）练习”的问诊教学模式，但面临患者资源匮乏和实践机会不足等问题，调查显示学生普遍反映练习时间不充分、学习效果欠佳。国外研究表明ChatGPT与Claude 2在医学模拟教学中能够帮助医学生提升医患沟通能力和临床技能水平[9]。通过虚拟患者，学生可练习诊断、沟通和决策，避免因临床经验不足对真实患者造成风险。例如，ChatGPT能模拟真实病例场景，提供个性化学习体验，显著提升学习效率并确保安全性，尤其在教学资源有限的地区，可弥补传统教学不足，帮助学生顺利从预临床阶段过渡到临床阶段[10]。国内研究表明GAI在诊断学问诊教学中有显著潜力。该技术融合自然语言处理与深度学习，辅助生成诊断内容、模拟问诊场景并提供个性化学习支持，从而革新传统的医学教学模式[11]。此外，GAI还拓展至临床技能培训领域，通过虚拟仿真训练有效提升学生的技能操作能力和临床决策水平[12]。

在模型的选择上，本研究对当前主流的GAI模型进行了综合评估。与以ChatGPT-4为代表的通用大模型相比，Claude 2在长文本处理与事实准确性方面表现更为突出，这对于模拟长病程、复杂病史的患者至关重要。例如，Claude 2支持更长的上下文窗口，能一次性解析并关联复杂病例信息，而ChatGPT-4在处理超长文本时可能出现细节遗忘[13]。在数据可靠性方面，Claude 2的“幻觉”率在多项第三方评测中低于ChatGPT-4，尤其在生物医学命名实体识别（分数达92.1%）等任务上，其精确

性显著优于LLaMA-2等开源模型[14]。与国内主流大模型相比，Claude 2在遵循复杂指令、保持角色一致性及伦理安全性方面更有优势，其采用的Constitutional AI框架经过严格红队测试，错误输出率显著降低，更适合应用于医疗等高风险教学场景[15]。因此，综合考量长文本处理能力、低幻觉率与伦理安全性，Claude 2是构建高保真、安全可靠的虚拟标准化病人的更优选择。因此，本研究旨在评估GAI问诊模型在诊断学教学中的应用效果与潜力。鉴于SP在基层教学机构中并非常规配置，而病房患者床旁教学是当前广泛使用的主流实践，本研究选择将GAI问诊模型与后者进行对照比较，以期务实评估GAI模型在补充和优化现有教学资源方面的实用价值。

1 资料与方法

1.1 研究对象与分组

以首都医科大学延庆教学医院2023级和2024级“3+2”助理全科医师规范化培训的73名医学生为研究对象。将2024级学生随机分为甲、乙、丙三组。为扩充样本量并构建对照，将2023级全体学生与2024级甲组合并设为对照组（n=38），2024级乙组与丙组合并为试验组（n=35）。鉴于所有学生均在相同教学体系与师资下接受培训，且本研究主要关注教学模式的即时效果，该分组设计合理。

1.2 纳入与排除标准

纳入标准：①于我院全科医学科接受助理全科医师规范化培训的医学生；②对研究内容知情且自愿参与。排除标准：①出科考核成绩或满意度调查问卷等关键研究资料缺失者；②未能全程参与本次教学研究安排的培训或评估活动者；③未按照教学方案完成规定训练或评估任务者；④同期参与其他可能影响本研究关键结局指标（如问诊技能）的教学研究者。

1.3 研究方法

对照组（传统床旁教学，基于病房患者）采用传统诊断学见习教学模式，具体流程为：首先教师讲解问诊理论要点、注意事项及技巧（同试验组），随后安排学生进入病房，对真实住院患者进行病史采集。为确保教学可比性，学生问诊的病例类型与试验组设计的8个典型病例在主要疾病谱上保持一致。教学环节以教师对学生表现的总结评价作为结束。

试验组采用GAI问诊模型教学模式，具体流程为：①由教师讲授问诊相关理论知识、问诊注意事项及技巧等（同对照组）；②采用Anthropic公司发布的Claude 2模型（版本v2.1），通过限制条件模拟患者角色，角色设定为“真实住院患者（非AI程序）”，回答内容中禁用医学术语（如“腹痛”需表述为“肚子一阵阵拧着疼”），设定温度参数0.7、最大生成长度500 tokens以控制对话逻辑，并通过10例测试病例验证对话逻辑符合率≥90%；③设计8个典型病例（消化科4例：胃食管反流病、肝硬化失代偿期；呼吸科2例：COPD急性加重、肺炎；心血管科2例：稳定性心绞痛、心力衰竭），病例资料来源于我院真实病案（经脱敏处理），每个病例均包含主诉、现病史（含3个以上时间节点）、既往史（含药物过敏史）、个人史（吸烟/饮酒量化数据）、家族史（一级亲属疾病史）等结构化数据，提交至Claude 2进行多轮迭代训练（每病例训练5轮，每轮包含完整问诊对话，并根据教师反馈调整提示词，直至模型输出与预设病例逻辑一致率达90%以上）；④由学生扮演医生，通过文字对话框与Claude 2进行交流，模拟问诊过程；⑤问诊结束后，学生对照设定的病例资料，寻找遗漏部分，总结不足。为减少评价偏倚，教师在进行出科考试评分时采用单盲设计，即教师不知晓学生所属组别。

1.4 评价方法

评估指标包括出科考试成绩和教学满意度调查两部分。出科考试采用百分制，从问诊流程、病史采集、临床思维、病例书写、综合应变能力和人文关怀六个维度进行考核。满意度调查涵盖自我评价、对指导教师、教学方法及教学效果的评价，采用Likert 5级量表（很满意、满意、一般、不满意、很差），其中“很满意”和“满意”合并为“总体满意”，其余选项归为“总体不满意”。所有调查均采用匿名方式进行。

1.5 统计分析方法

使用SPSS 22.0软件建立数据库，进行统计学分析。计数资料以n（%）表示，采用χ^2检验进行组间比较；计量资料用($\bar{x}\pm s$)表示，两组间比较用独立样本t检验，以$P<0.05$表示差异有统计学意义。

2 结果

2.1 两组研究对象一般情况

本研究共纳入73名助理全科医师规范化培训学生，其中试验组35人，对照组38人。试验组男性13人（37.14%），女性22人（62.86%）；年龄20～33岁，平均年龄(27.14±3.25)岁。对照组男性12人(31.58%)，女性26人(68.42%)；年龄19～37岁，平均年龄(28.74±5.28)岁。两组学生的性别、年龄及入学成绩比较，差异均无统计学意义（均$P>0.05$）。尽管对照组包含不同年级学生，基线比较显示两组在性别、年龄及入学成绩上无统计学差异（均$P>0.05$），表明两组具有可比性。

表1　两组研究对象的一般情况

特征	试验组(n=35)	对照组(n=38)	χ^2/t	P
男性[n(%)]	13(37.14)	12(31.58)	0.251	0.645
年龄[($\bar{x}\pm s$),岁]	27.14±3.25	28.74±5.28	−1.656*	0.104
入学成绩[($\bar{x}\pm s$),分]	85.27±3.46	84.93±3.69	0.902*	0.372

注：*表示t值。

2.2　两组学生出科考成绩比较

在教师评价结果显示，试验组在问诊流程[(91.40±2.89)分 vs.(89.34±4.12)分]和病史采集[(91.31±2.35)分vs.(88.55±3.45)分]上的得分高于对照组，差异有统计学意义(均P<0.05)；在临床思维[(90.34±2.21)分 vs.(89.29±3.30)分]和人文关怀[(97.86±1.46)分 vs.(93.71±2.88)分]方面，试验组也高于对照组，差异有统计学意义(均P<0.05)；在病例书写与综合应变能力上，两组成绩相近，差异无统计学意义(均P>0.05)(见表2)。

表2　两组学生出科考试成绩比较

考核项目	试验组[n=35,($\bar{x}\pm s$),分]	对照组[n=38,($\bar{x}\pm s$),分]	t	P
问诊流程	91.40±2.89	89.34±4.12	2.585	0.003
病史采集	91.31±2.35	88.55±3.45	3.472	0.001
临床思维	90.34±2.21	89.29±3.30	1.687	0.048
病例书写	90.90±2.73	90.34±3.12	0.925	0.179
综合应变能力	90.97±2.98	89.29±3.89	1.868	0.069
人文关怀	97.86±1.46	93.71±2.88	6.891	<0.001

2.3　两组学生满意度调查

学生自评结果显示，试验组学生在对教学方法(88.6% vs. 60.5%)与教学效果(82.9% vs. 52.6%)的总体满意度上高于对照组，差异有统计学意义(均P<0.01)。在对自我评价的满意度上，试验组(71.4%)也高于对照组(39.5%)，差异有统计学意义(P=0.006)。两组学生对指导老师的满意度差异无统计学意义(试验组80.00% vs.对照组76.3%，P=0.704)。如表3所示。

表3　两组学生满意度调查比较

调查维度	试验组[n=35,n(%)]	对照组[n=38,n(%)]	χ^2	P
教学方法	31(88.6)	23(60.5)	7.443	0.006
教学效果	29(82.9)	20(52.6)	7.543	0.006
自我评价	25(71.4)	15(39.5)	7.511	0.006
指导老师	28(80.0)	29(76.3)	0.145	0.704

3　讨论

临床问诊是医学生必备的核心技能，教学质量直接决定医学生将来的医疗水平。SP教学模式虽能高度模拟患者，并且互动性强，但其成本高昂、标准化程度统一难度大和时间、空间灵活性差，无法满足大规模、重复训练的需求。近些年来，AIGC技术的突破为教育领域开辟了新路径[16]。本研究将GAI问诊模型引入医学生问诊教学中，采用类随机对照试验，比较GAI问诊模型与传统床旁教学模式的优势和局限性，为构建“数智化”临床技能教学新生态提供实证与理论参考。本研究的对照组采用了在实际教学中广泛应用的病房患者床旁教学，而非经过严格训练的SP，这种设计旨在比较GAI模型与当前基层教学实践中更常见的传统教学方法的差异，其结论主要适用于评估GAI模型在补充或优化现有床旁教学方面的潜力。

本研究结果显示，采用GAI问诊模型的试验组在问诊流程、病史采集及人文关怀方面的教师评价得分均高于采用传统床旁教学的对照组，而在病例书写与综合应变能力上两组差异无统计学意义。同时，学生对GAI教学方法与效果的满意度也更高。试验组在问诊流程规范性和病史采集完整性上的突出表现，可能与GAI模型提供的结构化、可重复训练环境有关。GAI模型通过预设的结构化病例和即时、标准化的文本反馈，有效降低了学生在初始学习阶段的外部认知负荷(如记忆病例细节、担心问诊遗漏)，使其能更专注于问诊程序性知识(如提问顺序、关联性追问)的获取与内化，这与AIGC有助于规范医学生学习路径的研究一致[17]。其核心价值体现在三方面：其一，在教学资源配置上，GAI模型实现了从依赖人力到依托算法的转变，解决了SP招募、培训、薪酬及状态波动等问题[18]，让教师能专注于教学设计与个体化指导；其二，在学习模式上，GAI提供的可重复的训练场景，完美契合了刻意练习的核心要素——在明确目标指导下进行反复、专注的练习，并即时获得反馈以修正错误[19]；其三，在学习动机上，试验组88.6%的学生接受这种新型教学模式，其互动强、反馈快、无压力的特点有效激发了数字时代学生的学习内驱力。这种模式帮助学生在低压力环境下快速形成问诊的肌肉记忆，从而在教师评价的标准化考核中展现出更高的

操作规范性。

本研究中,试验组在“人文关怀”维度上的得分显著高于对照组($P<0.001$),揭示了GAI教学环境在技能习得初级阶段的独特价值。其内在机制并非GAI能够模拟或传递真实情感,而在于它为学生构建了一个心理舒适区 。在社会认知理论框架下,学生的行为受环境、个人认知与行为三者交互影响。面对真实患者时,学生常因担心犯错、冒犯或能力不足而产生表现焦虑 ,这严重抑制了其尝试运用关怀性语言和行为的意愿,GAI作为无感情的互动对象,可以有效弥补这一劣势。在这种安全的模拟环境中,学生得以卸下心理负担,将注意力从避免出错转向主动学习,更专注于练习和内化关怀性的标准化语言与对话模式。教师在进行评分时,其观察焦点也自然落在了学生是否主动、规范地执行了这些被教导的人文关怀行为上。因此,GAI模型在此处的核心价值,是作为人文关怀外显行为 的标准化训练器,高效地帮助学生完成了从“知识知晓”到“行为演练”的关键转化。这与技能获得模型中,初学者通过分解步骤和模仿来学习复杂技能的观点相吻合。

同时,本研究还发现了GAI模型在培养高阶临床能力,特别是“综合应变能力”方面的局限性。试验组与对照组在该维度上无显著差异($P>0.05$)。这可以从情境学习理论和适应性专长的角度进行剖析。首先,真实的临床决策发生在复杂、不确定且充满独特性的实践共同体中,需要医生根据动态变化的非结构化信息(如患者微表情、语气、突发体征、家属情绪等)进行即时判断与调整[20]。当前的GAI模型尽管能模拟标准病情发展,但本质上仍是在有限参数下运行的、相对封闭的文本交互系统。它无法再现真实临床情境中信息的不完备性、矛盾的突发性以及人际互动的微妙性,因此难以有效训练学生的情境感知力和适应性应变能力。其次,在培养临床推理的深度与灵活性上,AI也存在不足。虽然模型能基于预设逻辑回应,但它缺乏对疾病谱系深度关联和鉴别诊断复杂性的真实模拟能力,限制了学生进行发散性思维和应对罕见或复杂共病情况的训练机会。这导致学生在面对高度不确定性或需整合多元信息进行创新性决策时,从GAI训练中迁移出的能力有限。此外,GAI系统在标准化与个性化之间存在固有张力。例如,在“病例书写”项上两组无显著差异,可能部分归因于GAI训练中隐含的模板化倾向,虽保证了信息完整性,却可能在不经意间抑制了学生根据实际情况进行个性化、精炼化书面表达的创造性。

综上所述,GAI问诊模型与基于病房患者的传统床旁教学模式并非简单的替代关系,而是具备不同的教学功能与适用阶段。未来的医学教育改革应致力于构建一个阶梯式、互补型的整合教学体系。这一体系可遵循技能获得与专长发展的渐进规律:初级阶段,利用GAI模型进行大规模、低成本、标准化的问诊流程与基础人文行为刻意练习,夯实技能基础,建立自信;中级阶段,引入VP,在更具真人互动复杂性的环境中,评估和锻炼学生的整合应用能力、实时应变能力及初步的共情互动;高级阶段,在教师指导下进行真实临床实践,解决复杂、个性化的临床问题,并深化职业身份认同与内化的人文素养。在技术层面,未来需开发融合多模态交互(语音、视觉)、整合真实临床大数据与专家推理路径的下一代GAI模型,并结合虚拟现实、增强现实技术,构建高沉浸感、高动态性的模拟场景,以突破GAI在复杂情境模拟和深度决策支持方面的瓶颈[21]。

当然,我们也必须客观地看待本研究的局限性。本研究作为单中心探索性研究,样本量较小,且所有研究对象均来自同一教学医院,这可能限制了研究结论的普适性。不同地区、不同层级医院的教学资源和学生背景存在差异,本研究的发现能否在其他环境中推演,尚需进一步验证。同时,本研究的教学干预周期较短,未能评估AI问诊模型对学生长期临床能力发展的影响。因此,未来的研究应致力于开展多中心、大样本、长周期的随机对照试验,以提供更高等级的循证医学证据,从而推动GAI教学领域的持续进步与完善。

4 结束语

本研究通过教学实践对比研究,证实了AI问诊模型在诸多方面较传统教学模式有较为显著的优势,如提升医学生问诊技能、提高教师教学效率及学习主动性等方面。该模型能够为学生提供标准化、可重复的训练条件,有助于提高其问诊流程规范性与病史采集完整性,并获得了学生的认可。然而,GAI问诊模型在复杂临床情境的综合应变能力、个体化临床决策深度等方面仍存在局限。为增强教学效果,GAI能力提升任重而道远,建议将该教学模型推广至更广泛的临床技能培训场景,从而为医学教育方法的系统化创新提供有益借鉴。

参考文献

[1] 曹湘玉, 胡柯, 黄民江, 等. 基于学习通平台的7E教学模式在诊断学问诊教学中的应用效果[J]. 广西医学, 2022, 44(21): 2581-2584.

[2] Barrows H S. An overview of the uses of standardized patients for teaching and evaluating clinical skills[J]. Academic Medicine, 1993, 68(6): 443-451.

[3] 段志军, 金艳玲, 杜建玲, 等. 将医学生作为标准化病人用于问诊考核的实验与分析[J]. 中国现代医学杂志, 2007, 17(21): 2683-2685.

[4] 高志荣, 李青松. 标准化病人在诊断学实训问诊教学中的应用和体会[J]. 中国高等医学教育, 2019(9): 101-102.

[5] 王朝英. 基于5P教学模式的英语语音教学设计研究[J]. 教育进展, 2023, 13(10): 7308-7313.

[6] 王成红, 唐晓鸿, 张柯, 等. 交互式人工智能虚拟病例系统在诊断学教学中的实践应用[J]. 中华医学教育探索杂志, 2021, 20(4): 388-391.

[7] 李翔, 陈晓云, 张慧群, 等. 基于VSP的病史采集思维训练系统在问诊技能教学中的应用研究[J]. 中国医学教育技术, 2023, 37(2): 186-190.

[8] 王贤良, 葛昭, 赵志强, 等. 多元化教学模式在中医心血管内科住院医师规范化培训中的实践[J]. 中国中医药现代远程教育, 2023, 21(17): 1-3.

[9] Kıyak Y S, Emekli E. Using large language models to generate script concordance test in medical education: ChatGPT and Claude[J]. Revista Española de Educación Médica, 2024, 5(3): 1-8.

[10] Rajmohan S. Revolutionizing medical education: Can ChatGPT boost subjective learning and expression?[J]. Journal of Medical Systems, 2023, 47(1): 61.

[11] 宋雨欣, 李子怡, 丁宁. 生成式人工智能对医学教育的影响探析[J]. 中国医学教育技术, 2024, 38(3): 281-286.

[12] 张雷, 王安生, 刘戈, 等. 基于人工智能视域下模拟操作在胸心外科研究生教学中的应用[J]. 蚌埠医学院学报, 2024, 49(7): 965-968.

[13] Shao W J, Zhang R, Ji P L, et al. Astronomical knowledge entity extraction in astrophysics journal articles via large language models[J]. Research in Astronomy and Astrophysics, 2024, 24(6): 140-155.

[14] Jahan I, Laskar M T R, Peng C, et al. A comprehensive evaluation of large Language models on benchmark biomedical text processing tasks[J]. Computers in Biology and Medicine, 2024, 171: 108189.

[15] Anthropic. Model card and evaluations for claude models[R/OL]. San Francisco: Anthropic, 2023. [2025-04-29]. https://www. anthropic. com/claude-2-model-card.

[16] Cross J, Kayalackakom T, Robinson R E, et al. Assessing ChatGPT's capability as a new age standardized patient: Qualitative study[J]. JMIR Medical Education, 2025, 11: e63353.

[17] 陈湘, 邓然, 吴川清. 生成式人工智能大型语言模型在医学教育实践的探讨[J]. 临床急诊杂志, 2024, 25(6): 310-314.

[18] Liu X N, Wu C L, Lai R, et al. ChatGPT: When the artificial intelligence meets standardized patients in clinical training[J]. Journal of Translational Medicine, 2023, 21(1): 447.

[19] Ramazanzade K, Ebrahimzade A. Comparison of the effect of two clinical education methods SNAPPS and conventional method on the medical students' clinical reasoning skills[J]. Journal of Birjand University of Medical Sciences, 2018, 25: 46-52.

[20] 黄泽平, 张文君, 李思敏, 等. 基于元宇宙的临床医学教育教学应用前景[J]. 中国高等医学教育, 2023(9): 1-4.

[21] 丛慧源, 玄明实, 朱丹丹, 等. 基于人工智能的中医标准化病人问诊模型开发与应用[J]. 中医教育, 2025, 44(4): 131-135.