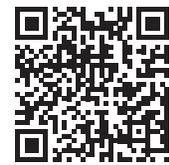


基于人工智能LLM技术的虚拟患者系统构建与临床教学应用



周志浩¹, 宋佳琳², 刘嘉成³, 周心悦⁴, 胡汉昆⁴

1. 东南大学附属中大医院急诊科 (南京 210009)
2. 东南大学医学院 (南京 210009)
3. 武汉大学计算机学院 (武汉 430072)
4. 武汉大学中南医院药学部 (武汉 430071)

【摘要】目的 构建一种基于人工智能大语言模型 (large language model, LLM) 技术、可用于医学教育的新型虚拟患者 (virtual patient, VP) 系统, 评价该系统在基层医生进修学习全科医学临床思维中的应用效果。**方法** 选取 2021 年 1 月至 2024 年 2 月在东南大学附属中大医院进修的基层社区医生为研究对象, 随机分为试验组和对照组, 分别采用基于 LLM 的 VP 系统教学、传统教学方法进行授课, 通过临床思维理论知识考核、临床思维能力考核、课程满意度调查评估教学效果, 并对结果进行相应的统计学分析。**结果** 共纳入 124 名基层社区医生, 其中试验组 60 例、对照组 64 例, 两组在一般基线资料上差异无统计学意义, 具有可比性。课程结束后, 试验组临床思维理论知识考核成绩显著高于对照组 (83.83 ± 3.15 vs. 79.92 ± 4.52 , $P < 0.01$), 且不及格率显著低于对照组 (0.00% vs. 9.38% , $P < 0.05$); 试验组在临床思维能力 3 个维度 (批判性、系统性、循证思维) 方面教学后分数均显著高于教学前, 而对照组仅在批判性思维维度上教学前后差异有统计学意义; 教学后试验组在系统思维、循证思维方面分数均显著高于对照组 ($P < 0.05$), 但在批判性思维上两组分数差异无统计学意义。试验组对授课的总体满意度也显著高于对照组 (93.33% vs. 85.48% , $P < 0.05$)。**结论** 基于 LLM 的 VP 系统提升了学员对临床思维理论知识的掌握程度, 也促进了其临床思维能力的培养, 该教学方法可为其他医学教育群体提供新的教学工具和思路。

【关键词】 人工智能; 大语言模型; 虚拟患者; 医学教育; 临床思维

Construction and clinical teaching application of virtual patient system: based on artificial intelligence LLM technology

ZHOU Zhihao¹, SONG Jialin², LIU Jiacheng³, ZHOU Xinyue⁴, HU Hankun⁴

1. Department of Emergency, Zhongda Hospital Southeast University, Nanjing 210009, China
2. School of Medicine, Southeast University, Nanjing 210009, China
3. School of Computer Science, Wuhan University, Wuhan 430072, China
4. Department of Pharmacy, Zhongnan Hospital of Wuhan University, Wuhan 430071, China

DOI: 10.12173/j.issn.1004-5511.202403091

基金项目: 科技创新 2030 “新一代人工智能”重大项目“标准化儿童患者模型关键技术与应用” (2021ZD0113400); 武汉大学学位与研究生教育教学改革研究一般项目

通信作者: 胡汉昆, 博士, 主任药师, 博士研究生导师, Email: huhankun@whu.edu.cn

周志浩, 主治医师, Email: xgg751116@126.com

Corresponding authors: HU Hankun, Email:huhankun@whu.edu.cn; ZHOU Zhihao, Email:xgg751116@126.com

【Abstract】Objective To construct a novel virtual patient (VP) system based on artificial intelligence large language model (LLM) technology for medical education, and to evaluate its effectiveness in improving the clinical thinking skills of primary care physicians in continuing education. **Methods** Primary care physicians from grassroots community hospitals who were in training at Zhongda Hospital Southeast University from January 2021 to February 2024 were selected as the research subjects and randomly divided into experimental group and control group. The VP system based on LLM and traditional teaching methods were used for instruction. The teaching effects were evaluated through theoretical knowledge assessment, clinical thinking ability assessment, and course satisfaction surveys, with the results subjected to corresponding statistical analysis. **Results** A total of 124 primary community doctors were included, with 60 in the experimental group and 64 in the control group. No significant statistical differences in the general baseline data between the experimental and control groups. The clinical thinking scores of the experimental group were significantly higher than those of the control group (83.83 ± 3.15 vs. 79.92 ± 4.52 , $P<0.01$), and the failure rate of the experimental group was significantly lower than that of the control group (0.00% vs. 9.38%, $P<0.05$). The experimental group showed a significant improvement in clinical thinking levels (critical, systematic, and evidence-based thinking) after teaching compared to before, while the control group only showed differences in critical thinking before and after teaching. The overall satisfaction with the course was also higher in the experimental group than in the control group (93.33% vs. 85.48%, $P<0.05$). **Conclusion** The VP system based on LLM enhanced the trainees' mastery of clinical thinking theoretical knowledge and also promoted the cultivation of their clinical thinking abilities. The teaching method provides new teaching tools and ideas for other medical education groups.

【Keywords】 Artificial intelligence; Large language model; Virtual patient; Medical education; Clinical thinking

虚拟患者 (virtual patient, VP) 在医学教育领域被证明是一种极具创新性和有效性的训练工具^[1-3]。它通过计算机技术,模拟患者从临床问诊、查体、辅检到诊断、治疗等全流程,提供一个零风险、低成本且贴近真实临床环境的实战训练平台,从而培养学员的临床思维能力^[4]。VP 技术自上世纪末起已在欧美国家得到广泛应用^[5],然而,VP 的构建过程却异常繁琐,需要投入大量的工作量和研发成本,导致目前市场上绝大多数 VP 产品所包含的案例数量与教学质量远未满足临床教学的实际需求。自 2022 年 11 月 ChatGPT-3.5 问世以来,大语言模型 (large language model, LLM) 迅速成为全球学术界和企业界瞩目的焦点,同时也促进了各行业的联合应用创新^[6-7]。

医疗行业日常运营场景产生大量的结构化或非结构化数据,LLM 凭借其快速阅读、精准解释海量数据的能力,能够高效地提取并整合这些数据中的关键信息。目前,LLM 在医疗领域的应用研究正逐渐深入,主要聚焦于提升医疗效率、降低医疗成本以及促进医学教育等方面^[8-11]。通过利用 LLM 对大量医学文献和案例进行深度学习和分析,可以提高知识传授和技能培养的高效性,从而深化医学教育改革与发展,满足新医科建设对医学人才的需求^[12]。

为了解决 VP 构建中海量文字与图形数据的读取与存储,以及基于自然语言生成技术的教学交互难题,本研究应用基于 LLM 的 VP 构建技术,旨在大幅提高 VP 的构建速度和效率,降低其构

建成本，并开展实际临床教学比较研究，以评价基于人工智能 LLM 技术的 VP 系统在临床实际教学中的应用效果及满意度。

1 资料与方法

1.1 基于 LLM 的 VP 系统构建

LLM 显著优势在于其卓越的人机交互能力，该能力使 LLM 能够精准洞察用户需求，并以高度结构化或半结构化的形式提供反馈，已有不少研究开始关注 LLM 与知识图谱的关联^[13-14]。本研究基于开源 LLM，通过微调开发适用于本研究任务的新模型—VP-LLM。VP-LLM 被用于从脱敏处理后的电子病历中抽取关键实体信息，用于构建诊疗流程图，该流程图是生成 VP 的核心知识底座，VP 生成流程图如图 1 所示。

1.1.1 实体识别

实体识别是自然语言处理（natural language processing, NLP）领域的一项关键技术，它专注于从输入文本中识别并标记出特定的实体，如疾病名称、症状描述、诊断结果、治疗措施等。近年来，基于 Transformer 架构的 LLM 在实体识别任务中展现了强大的性能。LLM 通过堆叠多个 Transformer 编码器，学习句子中不同位置间的复杂关系，从而实现对上下文信息的深度捕捉，该架构允许模型在处理序列数据时，自动学习到单词之间的依赖关系和语义联系。此外，LLM 还引入了自注意力机制，使模型能够更有效地处理长文本输入，减少信息丢失，并提升对序列内部结构的理解能力。

在 VP-LLM 系统中，本研究团队采用一种结

构化的方式来处理和解析医疗文本。根据教学大纲，将在临床实际工作中的典型病例、疑难病例等进行整理，纳入诊疗过程中所需的各种问题及答案（病史、查体、辅助检查等），并交于 LLM 自行进行实体识别纳入数据库。图 2 显示了部分实体类型名称列表。

1.1.2 诊疗流程图可视化编辑

对电子病历中的关键实体进行识别，能将实体按照诊断信息进行分类。这一过程中，问诊、查体、辅助检查等相关实体被视作诊断实体的支持或鉴别依据，为诊断提供了必要的背景信息和证据。同时，治疗实体则作为诊断实体的后续措施，进一步指导治疗方案的选择。通过这种细致的分类和逻辑构建，能够初步搭建起病历的诊疗逻辑框架。

本研究团队基于此研发了一套在线诊疗流程图可视化编辑平台，将识别出的实体与诊疗流程图进行关联确认，如图 3 所示。医学教培人员通过该平台能针对构建的 VP 底层诊疗流程图进行审核与编辑，并根据临床实际情况调整或补充诊疗流程，使其更加贴近实际临床生活，确保流程的准确性和实用性，这也可以提高学员们的互动性和主动性。

1.1.3 VP 生成与互动

基于在线诊疗流程图可视化编辑平台，研究团队获取了详尽的疾病诊疗流程图，该流程图完整展示了从病史采集到最终诊断，再到治疗方案制定与实施的全流程知识。为了验证 VP 系统的可行性和实用性，本项目设计并开发了一套 VP 工作站（图 4）。该工作站将疾病诊疗流程图与

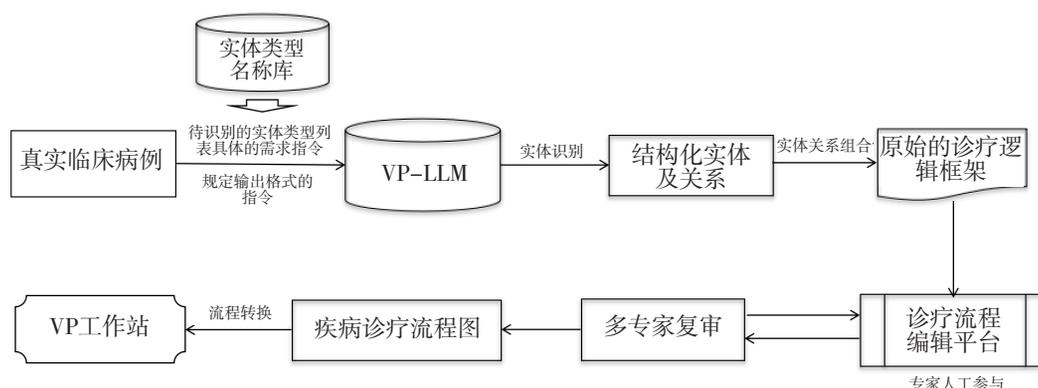


图1 VP生成流程图

Figure 1. VP generation flowchart

病史实体		症状实体		体格检查实体		辅检实体			
病史分类	病史类型	症状分类	症状类型	体格检查分类	体格检查类型	辅检大类	辅检小类		
一般情况	性别	发热	发热缓急	生命体征	体温	实验室检查	临床血液检测	血常规	
	年龄		发热温度		呼吸			网织红细胞	
	民族		持续时间		脉搏			红细胞沉降率	
	婚姻状况		发热部位	血压				
现病史	发病时间	咳嗽	咳嗽缓急	全身状态	意识状态		排泄物/分泌物/体液	生化	粪便常规
	起病急缓		持续时间		语调/语态				精液常规
	病因/诱因		声音大小		面容/表情			
	症状部位		咳嗽程度	体位/姿态/步态				
既往史	过往疾病	腹痛	腹痛部位	头颈部	视诊		血检与止血检测	血液学	血浆肉皮素
	过往外伤		疼痛性质		触诊				血小板黏附功能
	过往手术		发作缓急		听诊	凝血功能			
	过往过敏		持续时间				
个人史	居住地	气喘	气喘缓急	心脏	视诊	病理检查	体与组织和细胞病理	周围神经活检测	
	教育程度		持续时间		触诊			经皮穿刺肺活检测	
	职业		声音大小		听诊			鼻腔取活检测	
	烟酒嗜好		喘痰程度	叩诊				
家族史	父母疾病	汗出异常	汗量多少	腹部	视诊	电子显微镜技术	体液细胞学病理	骨髓组织活检测	
	近亲疾病		汗粘稠度		触诊			
	遗传病		汗出时间		听诊			
	健康情况		汗出部位	叩诊				

图2 部分实体类型名称库
Figure 2. Partial entity type name library

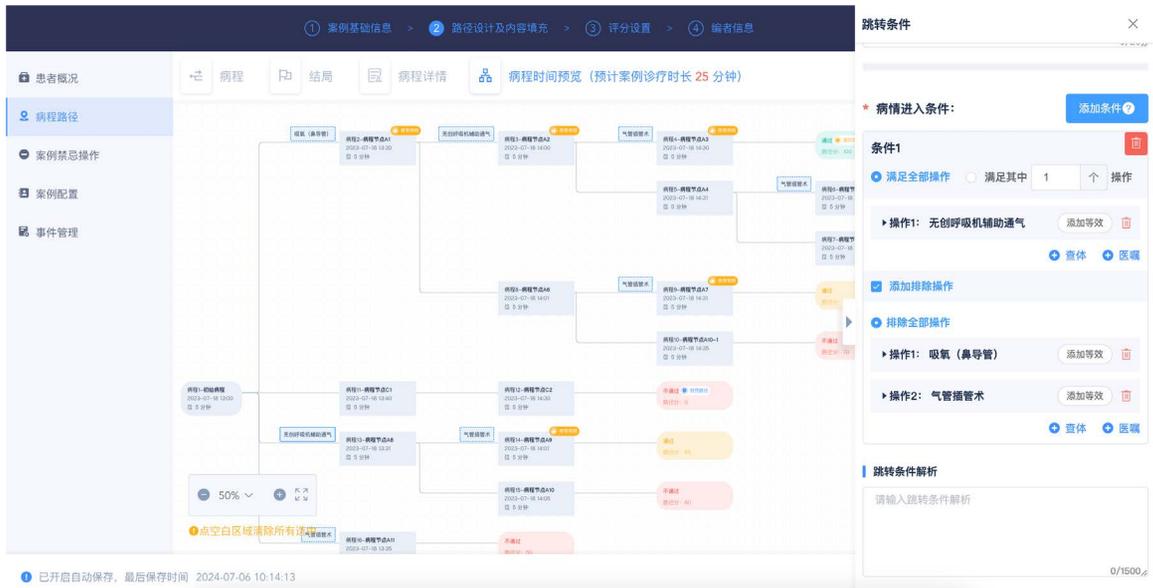


图3 在线诊疗流程可视化编辑平台界面
Figure 3. Platform interface online diagnosis and treatment process visualization editing

VP-LLM 相结合，实现了在线模拟诊疗的功能。通过此套系统，学员可以在模拟的临床情境中进行诊疗决策，而 VP 工作站将根据学员的决策实时反馈病情的动态变化，因此学员可以根据自己对于当前疾病的判断进行诊疗（询问相关病史、检查相应部位的症状和体征、进行必要的辅助检查等），并观察 VP 的病情发展，从而更加深入地理解采用不同诊疗策略对患者疾病的影响。这种互动式的模拟体验，不仅提高了学员对疾病诊疗流程的理解和熟悉程度，也提高了学员们对于疾病理论知识的掌握度，进而使其在实际临床工

作中更加自信。

1.2 临床教学应用效果评价

1.2.1 研究对象

本回顾性研究以 2021 年 1 月至 2024 年 2 月在东南大学附属中大医院进修学习全科医学专业（进修共计 1 年时间，上半年为理论教学，下半年为临床实践教学）的基层社区医生为研究对象。纳入标准：①年龄小于 45 岁；②具有《医师资格证书》和《医师执业证书》；③具备医学本科及以上学历，并应有相应专业 2 年及以上临床工作经验或具有专科学历，应有相应专业 5 年及以

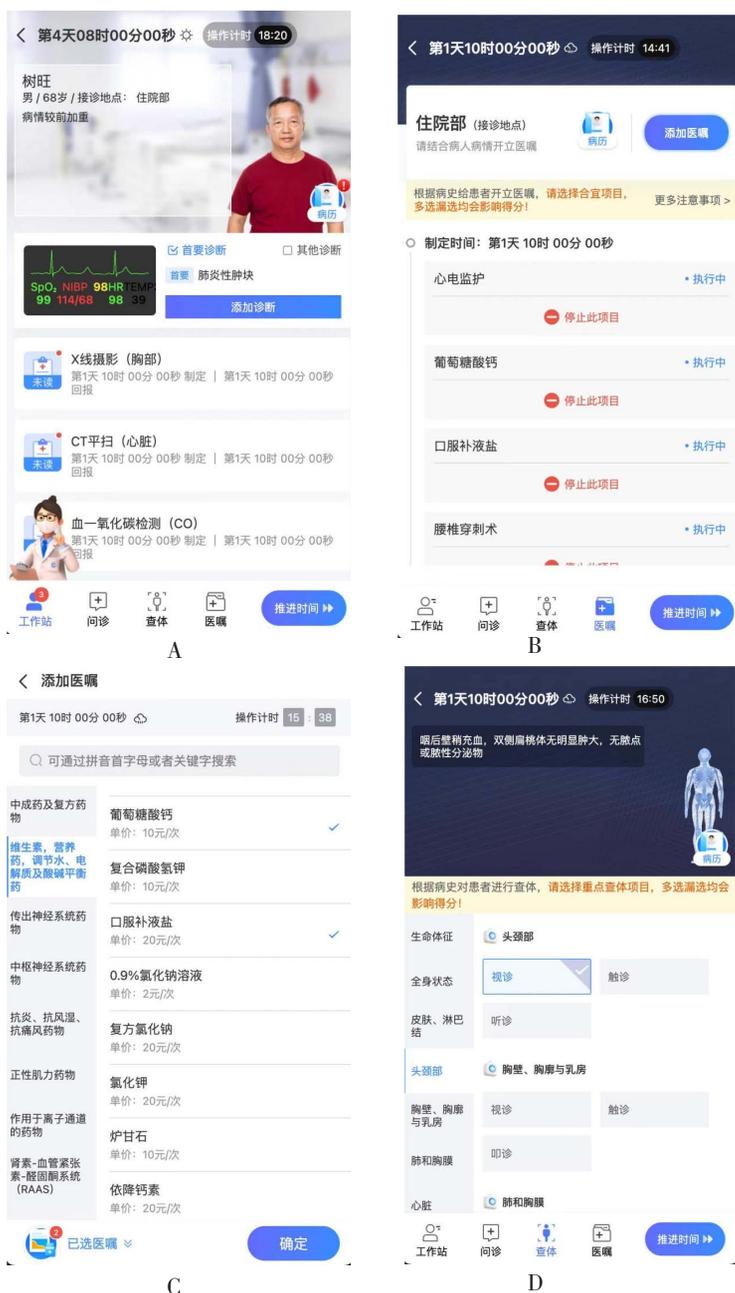


图4 VP系统互动界面

Figure 4. Interactive interface of VP system

注：A. VP模拟工作站；B. VP问诊；C. VP查体；D. VP医嘱。

上临床工作经验。排除标准：①年龄大于 45 岁；②无执业相关证书；③不具备相应的学历或工作经验。所有学员均对本研究内容知情同意。通过相关样本量计算公式，预计失访率约 10%，则至少纳入 110 名研究对象，本研究最终纳入 124 名。

1.2.2 研究方法

对符合纳排标准的学员按照随机数字表法将其随机分为试验组和对照组，然后进行全科医学相关理论知识的摸底考试。试验组采用 LLM 联合 VP 系统教学形式，对照组采用传统 PBL 教学形

式，分别同时进行 32 课时的全科医学教学训练。课程的教学目标、选择的案例、相关讲解训练的内容及授课教师保持一致。课程名称为《全科医学临床思维》，授课教师均为“疾病诊断与临床思维综合课程”教研组教师，并长期从事全科医学相关教学工作。具体实施过程如下：

试验组：由授课教师提前告知病例及问题，小组内成员自主查阅文献及讨论，课堂通过 VP 模拟患者真实就诊情景，成员在 VP 系统中自主进行该病例的诊疗（问诊、查体、采用辅助检查、

治疗手段等), 后台将其诊疗思路及过程登记上传至教师端, 学员操作结束后由授课教师进行相关知识点的系统讲解。

对照组: 由授课教师提前告知与试验组相同的病例及问题, 小组内成员自主查阅文献及讨论, 课堂上组员分别进行病例讲解及诊疗思路汇报, 最终教师进行答疑解惑并系统讲解该疾病的相关知识。

课程结束后, 两组进行相同的诊疗测试(参考国家卫生健康委《住院医师规范化培训结业临床实践能力考核标准方案 2022 版(全科)》)和相关专业理论知识考核(授课教师统一命题), 将考试成绩进行统计分析, 实践和理论考核满分各 100 分, 分别占比 50%, 构成临床思维理论知识考核总分。此外, 在评估相关文献后, 采用德尔菲专家函询法自行编制《医学生临床思维能力评估量表》^[15]在教学前后对两组社区医生进行调查。量表包括批判性思维、系统思维、循证思维 3 个维度, 每个维度的项目采用 Likert 5 级赋分法, “非常好” “较好” “一般” “较差” “非常差” 分别赋值 5、4、3、2、1 分, 得分越高, 临床思维水平越高。量表满分为 120 分, 换算为百分制后, 80~100 分为临床思维能力非常好、60~79 分为临床思维能力较好、40~59 分为一般、20~39 分为较差、0~19 分为非常差。相关量表 Cronbach' α 系数为 0.76。相关课程结束后采用“问卷星”小程序, 设置“疾病诊断与临床思维综合课程问卷调查(学生版)”对学员进行匿名调查, 对 VP 提升学员临床思维情况进行满意度调查, 内容包括一般特征、教学效果评价、临床诊疗能力、临床思维培养及提高等方面, 各题以 0~10 分计分, 6 分以上视为满意。

1.3 统计学分析

采用 SPSS 24.0 软件进行统计学分析, 计量资料以均数和标准差 ($\bar{x} \pm s$) 表示, 组内比较采用配对样本 t 检验, 组间比较采用独立样本 t

检验; 计数资料以例数和百分比 ($n, \%$) 表示, 组间比较采用 χ^2 检验, 计数资料理论频数不符合 χ^2 检验条件时使用 Fisher 确切概率法。 $P < 0.05$ 表示差异具有统计学意义。

2 结果

2.1 一般资料

共纳入 124 名基层社区医生, 其中试验组 60 例、对照组 64 例, 两组在年龄、性别、学历、摸底考试成绩等方面差异无统计学意义 ($P > 0.05$), 具有可比性, 见表 1。

2.2 临床思维能力及课程理论成绩比较

两组课程结束后的临床思维理论知识考核结果显示, 试验组成绩优于对照组 (83.83 ± 3.15 vs. 79.92 ± 4.52 , $t = -5.598$, $P < 0.01$), 且不及格率显著低于对照组 (0.00% vs. 9.38% , $P = 0.028$)。在临床思维能力 3 个维度上, 试验组教学后得分均显著高于教学前 ($P < 0.05$), 但对照组仅在批判性思维维度上教学前后差异有统计学意义 ($P < 0.05$), 而在系统思维和循证思维方面教学前后差异无统计学意义; 教学后, 试验组在系统思维、循证思维方面分数均显著高于对照组 ($P < 0.05$), 但在批判性思维上两组分数差异无统计学意义, 详见表 2。

2.3 临床诊疗能力满意度比较

共有 122 人完成了对临床诊疗能力和临床思维的提升满意度的匿名问卷调查。相较于对照组, 试验组在问诊能力、诊断能力、疾病治疗、逻辑思维能力满意度等方面显著提高, 差异均具有统计学意义 ($P < 0.05$), 在查体技能、疾病知识学习满意度方面略有提高, 但两组差异无统计学意义; 试验组对课程的总体满意度显著高于对照组 (93.33% vs. 85.48%), 差异具有统计学意义 ($P = 0.04$), 详见表 3。

表1 两组一般情况比较 ($n, \%$)

Table 1. Comparison of general situations between the two groups ($n, \%$)

特征	试验组 ($n=60$)	对照组 ($n=64$)	t/χ^2 值	P 值
年龄(岁)*	35.33 ± 7.15	33.95 ± 6.51	0.16	0.26
性别			0.03	0.72
男	29 (48.30)	33 (51.60)		
女	31 (51.70)	31 (48.40)		
学历(本科占比)	59 (98.30)	62 (96.90)	0.05	0.60
摸底考试成绩(分)*	79.82 ± 4.61	81.01 ± 4.11	0.27	0.13

注: *为计量资料, 以均数和标准差 ($\bar{x} \pm s$) 表示。

表2 两组临床思维能力教学前后比较 ($\bar{x} \pm s$)
Table 2. Comparison of clinical thinking ability and theoretical knowledge between the two groups ($\bar{x} \pm s$)

指标	试验组 (n=60)			对照组 (n=64)			t ₃ 值	P ₃ 值
	分数	t ₁ 值	P ₁ 值	分数	t ₂ 值	P ₂ 值		
批判性思维		-2.17	0.04		-3.83	0.01	0.53	0.60
教学前	13.75 ± 0.17			12.25 ± 0.59				
教学后	16.23 ± 5.07			15.89 ± 3.80				
系统思维		-2.60	0.02		-2.13	0.06	3.37	<0.01
教学前	43.89 ± 6.78			42.47 ± 5.33				
教学后	47.04 ± 4.45			46.75 ± 3.06				
循证思维		-2.21	0.03		-1.52	0.15	2.32	0.03
教学前	23.43 ± 2.66			24.88 ± 3.25				
教学后	28.99 ± 3.55			28.40 ± 4.40				

注: t₁值、P₁值为试验组教学前后组内比较; t₂值、P₂值为对照组教学前后组内比较; t₃值、P₃值为教学后试验组、对照组组间比较。

表3 两组临床诊疗能力及课堂满意度比较 (n, %)
Table 3. Comparison of clinical diagnosis and treatment abilities and classroom satisfaction between the two groups (n, %)

项目	试验组 (n=60)	对照组 (n=62)	χ ² 值	P值
问诊能力	58 (96.67)	55 (88.71)	4.67	0.03
查体技能	50 (83.33)	50 (80.65)	0.24	0.62
诊断能力	58 (96.67)	54 (87.10)	6.14	0.01
疾病治疗	57 (95.00)	54 (87.10)	3.98	0.04
逻辑思维能力	56 (93.33)	52 (83.87)	4.43	0.03
疾病知识学习	57 (95.00)	55 (88.71)	2.78	0.09
总体满意度	56 (93.33)	53 (85.48)	3.97	0.04

3 讨论

国家“互联网+”教育战略的提出,使得基于互联网的教学模式成为可能^[16]。本研究采用人工智能 LLM 技术的 VP 系统,并将其应用于临床实际教学中。该系统是以 LLM 为基础,凭借其快速阅读、精准解释海量数据的能力,结合诊疗过程中产生的大量数据,高效提取数据中的关键信息,从而构建 VP。利用其人机交互能力,可使学员在应用过程中体会真实的临床实践过程。

本研究结果表明,试验组理论考核成绩优于对照组,且不及格率显著低于对照组,说明 LLM 联合 VP 系统的应用在一定程度上可提高学员对于临床理论知识的掌握程度。既往的 PBL 案例需要教师结合实际生活场景和学习目标将临床真实病例进行改写,这无疑会受职业思维的限制,撰写一份能够体现生活化、情景化的整合案例并非

易事^[17]。有研究显示,VP 为安全、重复的培训提供了更多机会,产生更多的自主学习和更多增强临床推理的病例^[18]。VP 系统让学员中“知识不足者”经个性化互动训练从而补充完善知识体系,通过其成熟的教学模型,实现课中趣味教学、课后高效训练等功能,还可以把各个用户端诊治过程及时反馈给教师,使之进行对应的调整并加强相关方面的培养,实现真正意义上的“因材施教”,从而全面提高学员的医学理论知识水平。

在临床思维能力方面,本研究中两组临床思维能力均得到了一定的提升,教学后试验组在系统思维、循证思维方面分数均显著高于对照组,但在批判性思维上两组分数差异无统计学意义,说明 LLM 联合 VP 系统的应用可以通过人机交互的方式提高学员对于临床诊疗过程中的知识整合和逻辑互通,可在较大程度上提高每位学员在系统整合医学理论知识以及锻炼医学循证思维方面

的能力。LLM 作为人工智能的一种形式，联合 VP 系统的应用可帮助学员更有效地简化和转换复杂的患者信息，从而增强记忆和优化诊疗过程。基于可以主动回忆和探索的学习策略，有望提高医学理论知识的回忆和保留从而增强学员的理论基础^[19]。而 PBL 教学法以问题为中心，要求学员通过查阅资料、小组讨论、组间互评等方式，对学员自身的自主学习能力、语言表达能力可能做不到“面面俱到”。但 LLM 联合 VP 系统却可以上传每位学员平时和考核时的诊疗思路，更有针对性地发现和提出存在的问题。

VP 系统还可根据学员的表现，自动生成评价表，从问诊、查体能力、诊断和治疗能力等方面综合评估其表现，不仅有系统自动生成的评价表，也有学员自身的评价表，这些多维度的量化反馈可提供科学的评价和指导^[20]。本研究结果也显示，基于 VP 的教学形式，有效提升了学员自身的临床思维能力，尤其是问诊能力、诊断能力、疾病治疗能力和逻辑思维能力，说明相较于传统的 PBL 教学模式，LLM 联合 VP 具备更多的病例来源和场景模拟，提高了每位学员对于疾病的认知程度，并且教师角色由“知识传递”变成“学习促进”，学员获取知识的策略由“被动学习”转变为“主动学习”。

全科医生日常工作中需要在没有完善的检验/检查设备的情况下完成对患者的诊疗，因此其必须具备全面且扎实的全科医学知识、逻辑缜密的临床医学思维^[21]。传统 PBL 教学中，教师无法照顾到每位学员的逻辑思维，只能选择少数积极互动的学员进行 PBL 互动；而 VP 教学可及时获取每位学员对所学全科医学知识的掌握情况及临床思维情况。VP 教学巩固了学员的全科医学专业知识，更重要的是，通过个性化交互训练提高了每个学员系统性掌握全科医学知识的能力，这是对传统教学中仅侧重于总结性评价的完美补充。通过基于 LLM 的 VP 系统进行诊疗全流程交互训练，可将全科医学知识前后串联起来，调动起学员学习的积极主动性，也能及时了解自身知识掌握的程度，更好地系统性地理解与掌握知识，这在教学结束后的总体满意度调查结果中也得以验证。

全科医生临床决策能力贯穿疾病诊疗始终，在面对社区患者时，由于患者病种可覆盖人体各

系统，因此对全科医生有关临床思维与决策能力的持续培养非常重要^[22]。诊疗工作不仅要求医生具有扎实的理论知识、操作技能、医学伦理能力，而且需要秉持良好的思维能力，基于循证医学进行临床决策。在 VP 教学场景，学员沉浸在虚拟的临床案例中，互动的场景、虚拟的患者，激发学员主动思考和临床思辨，提升了临床决策能力。

综上，基于 LLM 技术构建的 VP 系统，能够较好地应用于临床教学，无论在主观上还是客观上均获得了学员的认可，可成为巩固全科医生产理论知识、提升其临床实践能力的重要工具。此外，在全科医学的培训教学中引入基于 VP 的教学思维和方法，可以有效弥补传统临床思维教学的弊端，让学员主动参与到教学活动中，真正实现“因材施教”，以达到提升基层医生临床思维能力的目的。本研究也存在一定局限性，如课后的反馈和满意度调查存在少量缺失，这可能与受试者配合程度有关。未来，研究团队会继续研究更为高效的基于 LLM 的 VP 系统，并更大范围地验证 VP 在其它医学培训教育中的应用效果，并将该方法推广至其他医学教育群体中，从而高质量地推动医学人才的培养。

参考文献

- 1 Elnaga HHA, Ahmed MB, Fathi MS, et al. Virtual versus paper-based PBL in a pulmonology course for medical undergraduates[J]. BMC Med Educ, 2023, 23(1): 433. DOI: [10.1186/s12909-023-04421-y](https://doi.org/10.1186/s12909-023-04421-y).
- 2 Mistry D, Brock CA, Lindsey T. The present and future of virtual reality in medical education: a narrative review[J]. Cureus, 2023, 15(12): e51124. DOI: [10.7759/cureus.51124](https://doi.org/10.7759/cureus.51124).
- 3 Chang HY, Chen CH, Liu CW. The effect of a virtual simulation-based educational application on nursing students' belief and self-efficacy in communicating with patients about complementary and alternative medicine[J]. Nurse Educ Today, 2022, 114: 105394. DOI: [10.1016/j.nedt.2022.105394](https://doi.org/10.1016/j.nedt.2022.105394).
- 4 Pottle J. Virtual reality and the transformation of medical education[J]. Future Healthc J, 2019, 6(3): 181-185. DOI: [10.7861/fhj.2019-0036](https://doi.org/10.7861/fhj.2019-0036).
- 5 Berman N, Fall LH, Smith S, et al. Integration strategies

- for using virtual patients in clinical clerkships[J]. *Acad Med*, 2009, 84(7): 942–949. DOI: [10.1097/ACM.0b013e3181a8c668](https://doi.org/10.1097/ACM.0b013e3181a8c668).
- 6 van Dis EAM, Bollen J, Zuidema W, et al. ChatGPT: five priorities for research[J]. *Nature*, 2023, 614(7947): 224–226. DOI: [10.1038/d41586-023-00288-7](https://doi.org/10.1038/d41586-023-00288-7).
- 7 Dwivedi YK, Kshetri N, Hughes L, et al. Opinion paper: "So what if ChatGPT wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy[J]. *Int J Inf Manage*, 2023, 71: 102642. DOI: [10.1016/j.ijinfomgt.2023.102642](https://doi.org/10.1016/j.ijinfomgt.2023.102642).
- 8 Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge[J]. *Nature*, 2023, 620(7972): 172–180. DOI: [10.1038/s41586-023-06291-2](https://doi.org/10.1038/s41586-023-06291-2).
- 9 余燕芳, 夏亮亮, 李翼鸿, 等. 基于知识图谱和大语言模型的终身学习资源库供给生态构建研究 [J]. *远程教育杂志*, 2024, 42(1): 104–112. [Yu YF, Xia LL, Li YH, et al. Research on the supply ecology of lifelong learning resource database based on knowledge graph and large language model[J]. *Journal of Distance Education*, 2024, 42(1): 104–112.] DOI: [10.15881/j.cnki.cn33-1304/g4.2024.01.012](https://doi.org/10.15881/j.cnki.cn33-1304/g4.2024.01.012).
- 10 Benítez TM, Xu Y, Boudreau JD, et al. Harnessing the potential of large language models in medical education: promise and pitfalls[J]. *J Am Med Inform Assoc*, 2024, 31(3): 776–783. DOI: [10.1093/jamia/ocad252](https://doi.org/10.1093/jamia/ocad252).
- 11 Safranek CW, Sidamon-Eristoff AE, Gilson A, et al. The role of large language models in medical education: applications and implications[J]. *JMIR Med Educ*, 2023, 9: e50945. DOI: [10.2196/50945](https://doi.org/10.2196/50945).
- 12 成颖, 周四元, 乔逸. 新医科背景下案例驱动-双师同堂结合教学模式与传统教学模式的比较 [J]. *药物流行病学杂志*, 2023, 32(10): 1194–1200. [Cheng Y, Zhou SY, Qiao Y. Comparison between case-driven-dual-teacher cooperative teaching model and traditional teaching method under the background of new medical science[J]. *Chinese Journal of Pharmacoepidemiology*, 2023, 32(10): 1194–1200.] DOI: [10.19960/j.issn.1005-0698.202310014](https://doi.org/10.19960/j.issn.1005-0698.202310014).
- 13 李源, 马新宇, 杨国利, 等. 面向知识图谱和大语言模型的因果关系推断综述 [J]. *计算机科学与探索*, 2023, 17(10): 2358–2376. [Li Y, Ma XY, Yang GL, et al. Survey of causal inference for knowledge graphs and large language models[J]. *Journal of Frontiers of Computer Science and Technology*, 2023, 17(10): 2358–2376.] DOI: [10.3778/j.issn.1673-9418.2307065](https://doi.org/10.3778/j.issn.1673-9418.2307065).
- 14 Jovanović M, Campbell M. Connecting AI: merging large language models and knowledge graph[J]. *Computer*, 2023, 56(11): 103–108. DOI: [10.1109/MC.2023.3305206](https://doi.org/10.1109/MC.2023.3305206).
- 15 宋俊岩. 医学生临床思维能力评价指标体系的建立及实证研究 [D]. 青岛: 青岛大学, 2015. [Song JY. A study on the evaluation index system of clinical thinking ability of medical students and demonstration[D]. Qingdao: Qingdao University, 2015.] <https://d.wanfangdata.com.cn/thesis/D705674>.
- 16 杨琳. 基于“互联网+教育”的网络传播课程线上教学探索研究 [J]. *新闻研究导刊*, 2024, 15(9): 117–119. [Yang L. Exploration and research on online teaching of network communication course based on 'internet plus education'[J]. *Journal of News Research*, 2024, 15(9): 117–119.] DOI: [10.3969/j.issn.1674-8883.2024.09.034](https://doi.org/10.3969/j.issn.1674-8883.2024.09.034).
- 17 丁文婧, 沈海波, 徐丁. ChatGPT 在泌尿外科 PBL 教学中的应用探索与实践 [J]. *中国医学教育技术*, 2024, 38(3): 287–292. [Ding WJ, Shen HB, Xu D. Exploration and practice of applying ChatGPT in PBL teaching of urology[J]. *China Medical Education Technology*, 2024, 38(3): 287–292.] DOI: [10.13566/j.cnki.cmet.cn61-1317/g4.202403006](https://doi.org/10.13566/j.cnki.cmet.cn61-1317/g4.202403006).
- 18 Hamilton A. Artificial intelligence and healthcare simulation: the shifting landscape of medical education[J]. *Cureus*, 2024, 16(5): e59747. DOI: [10.7759/cureus.59747](https://doi.org/10.7759/cureus.59747).
- 19 Arango-Ibanez JP, Posso-Nuñez JA, Díaz-Solórzano JP, et al. Evidence-based learning strategies in medicine using AI[J]. *JMIR Med Educ*, 2024, 10: e54507. DOI: [10.2196/54507](https://doi.org/10.2196/54507).
- 20 郝洁, 彭庆龙, 丛山, 等. 基于提示学习的医学量表问题文本多分类研究 [J]. *中国循证医学杂志*, 2024, 24(1): 76–82. [Hao J, Peng QL, Cong S, et al. A study on multi-class classification of medical questionnaire item texts based on prompt learning[J]. *Chinese Journal of Evidence-Based Medicine*, 2024, 24(1): 76–82.] DOI: [10.7507/1672-2531.202307139](https://doi.org/10.7507/1672-2531.202307139).
- 21 Stolper E, Van Royen P, Jack E, et al. Embracing complexity with systems thinking in general practitioners' clinical reasoning helps handling uncertainty[J]. *J Eval Clin Pract*, 2021, 27(5): 1175–1181. DOI: [yxxx.whuznhmedj.com](https://doi.org/10.1111/</p></div><div data-bbox=)

[jep.13549](#).

- 22 MacFarlane AE. Optimising individual and community involvement in health decision-making in general practice consultations and primary care settings: a way

forward[J]. Eur J Gen Pract, 2020, 26(1):196–201. DOI: [10.1080/13814788.2020.1861245](#).

收稿日期: 2024 年 03 月 26 日 修回日期: 2024 年 06 月 24 日
本文编辑: 桂裕亮 曹 越

引用本文: 周志浩, 宋佳琳, 刘嘉成, 等. 基于人工智能LLM技术的虚拟患者系统构建与临床教学应用[J]. 医学新知, 2024, 34(7): 833–842. DOI: 10.12173/j.issn.1004–5511.202403091
Zhou ZH, Song JL, Liu JC, et al. Construction and clinical teaching application of virtual patient system: based on artificial intelligence LLM technology[J]. Yixue Xinzhi Zazhi, 2024, 34(7): 833–842. DOI: 10.12173/j.issn.1004–5511.202403091