

基于学生视角的大语言模型生成护理硕士跨学科连续照护模拟案例的评价研究

张桃桃, 程 妍, 徐蔡洁, 商 丽*

湖北医药学院护理学院, 湖北 十堰

收稿日期: 2026年7月27日; 录用日期: 2026年9月21日; 发布日期: 2026年9月29日

摘 要

目的: 本研究旨在系统比较DeepSeek R1、GEMINI和豆包三种大语言模型(LLM)生成三幕式连续照护案例的质量差异。方法: 采用标准化提示词驱动三款LLM各生成5例覆盖常见临床场景的案例, 由23名护理硕士研究生基于验证过的评价指标体系进行盲法评价, 涵盖案例结构、内容(含跨学科协作与循证思维)及应用效果维度。结果: DeepSeek R1、GEMINI和豆包三种模型在综合总分依次为 (258.43 ± 67.68) 、 (245.96 ± 74.01) 、 (245.43 ± 74.56) , 无显著统计学差异($P > 0.788$), 且所有二级指标也均无统计学差异($P > 0.05$); 然而, 跨学科协作深度仅达中等水平, 在该维度的得分DeepSeek R1、GEMINI和豆包依次为 (25.57 ± 7.04) 、 (24.87 ± 6.70) 、 (23.91 ± 7.67) 。结论: 本研究所测试的三种LLM在结构化提示下生成案例的整体质量趋于一致且基本满足教学需求, 可作为高效的护理模拟案例辅助生成工具, 但在跨学科互动真实性与伦理安全性上仍需人工审校。未来研究应聚焦提示词工程优化以增强情境深度, 并通过真实教学场景试验验证其对学生临床推理能力的实际影响, 从而构建“人机协同”的护理教育新范式。

关键词

大语言模型(LLM), 护理专业学位研究生(MNS), 跨学科, 模拟案例

An Evaluation Study on the Generation of Interdisciplinary Continuous Care Simulation Cases for Master of Nursing by LLM from a Student Perspective

Taotao Zhang, Yan Cheng, Caijie Xu, Li Shang*

School of Nursing, Hubei University of Medicine, Shiyan Hubei

*通讯作者。

文章引用: 张桃桃, 程妍, 徐蔡洁, 商丽. 基于学生视角的大语言模型生成护理硕士跨学科连续照护模拟案例的评价研究[J]. 创新教育研究, 2026, 14(9): 807-814. DOI: 10.12677/ces.2026.149747

Abstract

Objective: This study aims to systematically compare the quality differences in generating three-act continuous care case scenarios by three large language models (LLMs): DeepSeek R1, GEMINI, and DouBao. **Methods:** Standardized prompts were used to generate five cases each from the three LLMs, covering common clinical scenarios. Twenty-three master's students in nursing conducted blinded evaluations using a validated assessment framework, evaluating case structure, content (including interdisciplinary collaboration and evidence-based thinking), and application effectiveness. **Results:** The overall composite scores of DeepSeek, GEMINI, and DouBao were 258.43 ± 67.68 , 245.96 ± 74.01 , and 245.43 ± 74.56 , respectively, with no statistically significant differences ($P > 0.788$); all secondary indicators also showed no significant differences ($P > 0.05$). However, the depth of interdisciplinary collaboration was only moderate, with scores of 25.57 ± 7.04 , 24.87 ± 6.70 , and 23.91 ± 7.67 for DeepSeek R1, GEMINI, and DouBao, respectively. **Conclusion:** Current mainstream LLMs produce cases of consistent overall quality under structured prompts, meeting basic teaching requirements and serving as efficient auxiliary tools for nursing simulation case generation. However, manual review remains necessary to ensure the authenticity of interdisciplinary interactions and ethical safety. Future research should focus on optimizing prompt engineering to enhance contextual depth and validate its practical impact on students' clinical reasoning abilities through real-world teaching scenario experiments, thereby establishing a new paradigm of "human-machine collaboration" in nursing education.

Keywords

Large Language Model (LLM), Master of Nursing (MNS), Interdisciplinary, Simulation Case

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着全球老龄化进程的加速以及慢性病共病趋势的日益严峻,医疗卫生服务的需求已从传统的院内急性照护,向“院内-出院-社区/居家”一体化的连续照护模式转变。这种转变对护理人才的培养提出了更高要求,不仅需要其掌握扎实的临床推理与照护计划制定能力,更需具备在多学科团队中进行有效协作的跨学科素养[1]。护理专业学位研究生(MNS)作为高级临床护理实践者的主要储备力量,其跨学科连续照护能力的培养已成为当前护理教育的核心议题之一[2][3]。模拟教学作为一种有效的体验式学习策略,在培养护理硕士研究生高阶思维能力、临床决策能力及团队协作能力方面扮演着至关重要的角色[4][5]。然而,传统的模拟教学案例开发过程高度依赖于临床专家的知识与经验,存在耗时长、资源需求大、案例主题单一和情境固化等固有瓶颈,难以高效、全面地覆盖连续照护的完整轨迹,尤其是在居家照护与社区康复场景的呈现上存在显著短板[6]。近年来,以大语言模型(LLMs)为代表的生成式人工智能技术能够通过海量文本数据上进行训练,展现出先进的自然语言理解和内容生成能力[7]。近期一项研究以相同提示词在5个生成式AI平台生成老年案例,发现尽管案例总体内容多被认为较准确,但案例情境高度集中在传统医疗机构场景,未出现居家照护案例[8]。因此,本研究构建护理硕士跨学科连续照护,包

括院内 - 出院 - 社区/居家三幕式模拟案例的标准化生成, 比较不同大语言模型生成案例的质量差异, 并评价生成案例在教学可用性、安全合规性、跨学科协作要素与连续照护轨迹覆盖方面的表现。

2. 对象与方法

2.1. 研究对象

选择大语言模型 DeepSeek、GEMINI、豆包(Doubao), 上述 3 种模型在自然语言理解与生成任务中均具有代表性, 且在国内教育场景中的可及性存在差异, 比较其结果有助于为护理教育机构选择适宜的生成工具提供参考。

2026 年 3 月~4 月期间以招募的形式, 便利选取湖北医药学院护理学院的护理专业硕士研究生 23 名为评价者。纳入标准: 护理硕士专业学位研究生; 具备至少 6 个月临床轮转经历; 已完成护理专业课程学习; 自愿参与本研究。排除标准: 因外出实习、请假等原因无法全程参与本研究评价过程者。

2.2. 研究方法

2.2.1. 案例生成

案例生成提示词的设计基于 MNS 核心能力培养要求, 结合美国护理联盟模拟教学设计标准及跨学科连续照护实践框架, 经 5 名具有模拟教学经验的护理教育专家进行独立审阅和修订。提示词结构包含案例主题设定、患者基本信息、三幕式情境递进要求、跨学科团队参与节点、需嵌入的临床决策冲突点及伦理考量要素等核心模块, 以确保生成案例符合教学目标导向。

案例生成流程遵循标准化操作程序。首先, 将统一提示词分别输入 3 种大语言模型, 设定温度参数为 0.7 以平衡创造性与一致性, 最大输出令牌数设定为 4096 以确保三幕式完整叙事。每款模型独立生成 5 例跨学科连续照护案例, 主题涵盖老年共病管理、肿瘤患者过渡期护理、慢性伤口延续照护、脑卒中后康复及终末期症状控制五个临床常见场景, 以检验模型在不同复杂度情境下的生成稳定性。生成案例经初步格式整理后, 进入去标识化与盲法处理环节。

2.2.2. 案例去标识化与盲法

案例去标识化采用双重处理策略。第一层为自动脱敏, 利用正则表达式匹配并替换患者姓名、具体住址、医保编号等直接标识符, 将医院名称统一编码为 A 医院、B 社区卫生服务中心等泛化表述, 时间信息转换为相对病程天数而非具体日期。第二层为人工复核, 由 2 名未参与后续评价的研究者独立审查案例文本, 确保无间接标识信息残留, 如罕见病种组合、特殊职业背景等可能指向真实个案的描述特征。盲法实施方面, 所有生成案例经随机编码后, 评价者仅获知案例主题为跨学科连续照护情境, 不透露具体生成模型来源, 以避免品牌先入效应干扰质量判断。

2.2.3. 评价工具

本研究采用采用欧阳子瑶等[9]构建的 MNS 教学临床案例库评价指标体系进行评估。该体系基于三维质量结构模型构建, 包含 3 项一级指标、12 项二级指标、72 项三级指标, 涵盖案例结构、案例内容、案例应用效果 3 个维度, 采用德尔菲法与优序图法确定指标权重, 专家权威系数为 0.888~0.902, 肯德尔和谐系数为 0.157~0.258 ($P < 0.001$), 指标重要性评分均值为 4.23~5.00 分, 评价者根据案例逐条按照 Likert 5 分法进行评价, 1~5 分代表“非常不同意”到“非常同意”, 得分越高表明案例质量越优, 可系统评估案例的规范性、科学性、循证性、跨学科融合度、连续照护设计水平及教学应用价值。

2.2.4. 评价实施

为提高评价结果的一致性, 所有评价者在正式评价前接受统一标准化培训。培训由研究团队成员实

施, 内容包括研究目的与流程介绍、评价指标体系结构及各级指标内涵解析、评分标准说明、典型案例示范评价以及常见评分偏差讨论等。培训过程中, 研究者重点解释案例结构、案例内容、跨学科协作、循证思维及教学应用效果等关键评价维度的判断依据, 并组织评价者针对示例案例进行预评价。研究团队根据预评价结果, 对评分差异较大的指标进行集中讨论, 进一步明确评分边界和判断标准, 形成统一评价规范后开展正式评价。

本研究采用在线随机分组抽签工具进行评价任务标准化分配。研究共纳入 15 例模拟教学案例和 23 名护理硕士研究生作为评价对象, 分别编号为 C1~C15 和 S1~S23。为降低评价疲劳和重复评价偏倚, 每名研究生随机评价 3 例不同案例, 采用案例均衡轮次分配机制, 最终形成 69 份评价任务, 其中 9 例案例接受 5 次评价, 6 例案例接受 4 次评价, 确保不同案例获得相对均衡的评价次数。评价者仅获知案例主题及评价任务, 不知晓案例对应的大语言模型来源, 实现盲法评价。

评价过程通过线上问卷平台完成, 系统设置必填项校验, 避免数据遗漏。全部评价工作均在 2 周内的规定时间内完成, 评价者可自主安排评价时间, 无时间压力干扰评价判断。所有评价结束后, 由研究者统一导出原始评分数据, 并建立匿名化数据库。数据处理流程包括: (1) 数据完整性检查, 剔除未完成全部评价条目的记录; (2) 逻辑一致性核查, 检查异常评分及重复提交情况; (3) 数据编码, 将评价者编号、案例编号及各指标评分进行统一编码; (4) 形成最终分析数据库, 用于后续统计分析。所有数据处理过程均由两名研究者独立核对, 以保证数据准确性和可追溯性。

2.3. 统计分析

采用 SPSS 26.0 软件进行数据分析。连续变量采用均数 $\pm$ 标准差($\bar{x} \pm s$)进行描述, 分类变量采用频数和百分比进行描述。首先对数据分布特征进行检验, 根据数据分布情况选择相应统计方法比较不同大语言模型生成案例质量差异。本研究采用 Kruskal-Wallis 检验比较 DeepSeek R1、GEMINI 及豆包三组案例综合评分及各维度评分差异, 并采用 Mann-Whitney U 检验进行两两比较。

为评价不同评价者之间评分的一致性, 采用 Fleiss' Kappa 系数评价多评价者分类一致性。根据评价结果计算总体评价者间一致性, 并报告 Kappa 值及其 95%置信区间(95%CI)。Kappa 值解释参考 Landis 和 Koch 标准: <0.00 表示一致性较差, $0.00\sim0.20$ 表示轻微一致, $0.21\sim0.40$ 表示一般一致, $0.41\sim0.60$ 表示中等一致, $0.61\sim0.80$ 表示较好一致, $0.81\sim1.00$ 表示高度一致。统计检验均采用双侧检验, 以 $P<0.05$ 为差异具有统计学意义。

2.4. 伦理学考虑

所有评价者签署知情同意书, 明确告知案例由 AI 生成且经过脱敏处理, 评价数据仅用于学术研究并以编码形式呈现。研究过程中如发现生成案例存在明显临床错误或伦理争议情境, 立即启动专家复核程序, 必要时终止该案例的评价流程, 以确保参与者的专业判断不受误导性信息影响。

3. 结果

3.1. 评价者间一致性分析

采用 Fleiss' Kappa 检验评价 23 名护理硕士研究生对 LLM 生成案例评分的一致性。结果显示, 本研究总体 Fleiss' Kappa 值为 0.68 (95%CI: 0.62~0.74), 提示不同评价者之间的评分一致性为较好一致, 评分结果可靠, 可开展后续质量差异分析。其中案例结构维度的一致性最高, Kappa 值为 0.73 (95%CI: 0.66~0.79), 案例内容维度 Kappa 值为 0.66 (95%CI: 0.59~0.72), 案例应用效果维度 Kappa 值为 0.63 (95%CI: 0.57~0.69), 各维度一致性均达到较好及以上水平。

3.2. 评价者一般资料

本次共发放评价资料 23 份，回收有效评价数据 23 份，有效回收率为 100%。23 名护理硕士研究生中，男生 2 名，女生 21 名；年龄 22~28 岁，平均(24.31±1.42)岁；研究生就读年级为研一 4 名，研二 10 名，研三 9 名；临床轮转时长 6~23 个月，平均(12.65±3.74)个月；有慢病照护实践经历者 16 名，参与过模拟教学活动者 19 名(见表 1)。

Table 1. Basic information of MNS graduate students (n = 23)

表 1. MNS 研究生基本情况(n = 23)

项目	分类	n	百分比/%
性别	男	2	8.7
	女	21	91.3
年级	研一	4	17.39
	研二	10	43.48
	研三	9	39.13
研究方向	慢性病管理	8	34.78
	急危重症管理	7	30.43
	长期照护	8	34.78
读研形式	非全日制	3	13.04
	全日制	20	86.96

3.3. 不同 LLM 生成案例的整体质量分析

由表 2 可见，3 种模型生成案例的综合总分分别为：DeepSeek (258.43 ± 67.68)、GEMINI (245.96 ± 74.01)、豆包(245.43 ± 74.56)，组间差异无统计学意义(F = 0.239, P = 0.788)。经 LSD 多重比较显示，任意两组间差异均无统计学意义(P > 0.05)。

3.4. 不同 LLM 生成案例的案例结构、案例内容、案例应用效果分析

由表 2 可见，跨学科协作深度仅达中等水平，在该维度的得分 DeepSeek R1R1、GEMINI 和豆包依次为(25.57 ± 7.04)、(24.87 ± 6.70)、(23.91 ± 7.67)，说明本研究所测试的三种 LLM 在生成深度跨学科融合的教学案例时，尚存在能力短板。

DeepSeek 在案例撰写规范性、内容系统性、跨学科特性、循证思维、能力素养等多个维度的平均分略高，说明它生成的案例在结构严谨性和内容深度上，整体表现更稳定一些，但优势并不明显，不具备统计学意义(P > 0.05)。

Table 2. Overall quality and dimensional analysis of different LLM-generated cases (n = 69, $\bar{x} \pm s$)

表 2. 不同 LLM 生成案例的整体质量及各维度分析(n = 69, $\bar{x} \pm s$)

一级指标	二级指标	DeepSeek R1	GEMINI	豆包	F 值	P 值
1 案例结构指标	1.1 案例具有完整性(4 条)	14.35 ± 4.51	14.52 ± 5.01	13.04 ± 4.86	0.653	0.524
	1.2 案例编排合理性(4 条)	14.78 ± 4.47	13.35 ± 4.74	14.17 ± 4.65	0.558	0.575
	1.3 案例撰写规范性(5 条)	18.48 ± 5.52	16.96 ± 6.18	16.91 ± 5.26	0.57	0.568
	结构总分	47.61 ± 13.88	44.83 ± 15.41	44.13 ± 14.29	0.369	0.693

续表

2 案例内容指标	2.1 案例选题真实有代表性, 与教学目标一致(6 条)	22.13 ± 6.31	19.43 ± 6.29	21.22 ± 6.71	1.042	0.358
	2.2 案例具有系统性和设计性(7 条)	25.78 ± 7.06	24.26 ± 7.53	23.39 ± 8.53	0.564	0.572
	2.3 案例具有跨学科特性(7 条)	25.57 ± 7.04	24.87 ± 6.70	23.91 ± 7.67	0.31	0.735
	2.4 案例遵循科学性原则(5 条)	18.22 ± 5.57	16.91 ± 5.42	18.30 ± 4.89	0.497	0.61
	2.5 案例基于循证思维(3 条)	11.17 ± 3.79	10.87 ± 3.97	10.39 ± 3.90	0.295	0.746
	内容总分	102.87 ± 27.75	96.35 ± 28.91	97.22 ± 30.47	0.341	0.712
3 案例应用效果指标	3.1 提高学生职业道德素养(7 条)	24.96 ± 6.77	24.57 ± 8.41	24.43 ± 8.41	0.027	0.973
	3.2 提高学生知识素养(3 条)	10.74 ± 2.78	10.65 ± 3.63	10.43 ± 3.76	0.048	0.953
	3.3 提高学生能力素养(14 条)	51.09 ± 14.77	48.61 ± 16.08	48.74 ± 15.36	0.188	0.829
	3.4 案例具有实践性(7 条)	21.17 ± 4.98	20.96 ± 4.62	20.48 ± 5.08	0.122	0.886
	效果总分	107.96 ± 26.98	104.78 ± 30.32	104.09 ± 30.62	0.114	0.893
总体总分		258.43 ± 67.68	245.96 ± 74.01	245.43 ± 74.56	0.239	0.788

4. 讨论

本研究结果显示, 在统一的结构化提示词引导下, DeepSeek R1、GEMINI、豆包 3 种模型生成的案例在综合总分及结构、内容、应用效果等维度上均无显著统计学差异($P > 0.05$), 表明本研究所测试的三种 LLM 在遵循标准化指令时, 其输出质量趋于一致, 均能基本满足教学需求。该结果凸显了当前生成式 AI 在遵循标准化框架生成结构化教育内容时的通用性和稳定性[6]。这与当前报道的不同 LLM 在对肿瘤标志物解读、回答专业问题等方面有差异等研究结果不同[10][11], 但亦有研究尚未指出不同模型在护理领域中的生成的健康教育方案有统计学差异[12]。本研究中 3 种模型虽然保证了教学案例的基本质量和规范性, 但也反映出, 不同模型在满足护理教学特定需求时, 均未表现出突破性的优势, 后续仍需针对护理跨学科连续照护场景对模型进行领域微调与指令优化, 才能进一步提升生成质量。

本研究发现, 3 种模型在跨学科协作深度维度的得分仅处于中等水平, 这与跨学科连续照护本身的复杂属性有关, 该场景要求案例不仅整合护理、医疗、康复、社工等多角色职责[13], 还要清晰呈现不同照护阶段的衔接节点与协作冲突, 对大语言模型的逻辑整合能力与领域知识广度都提出了较高要求, 现有通用模型的知识储备与场景整合能力, 在支撑深度跨学科内容的上有待提升, 这也为后续相关研究指明了方向, 可通过在提示词中明确嵌入各学科具体职责清单、分阶段协作流程模板等领域专属信息, 或接入专科学识库辅助生成, 来弥补现有模型的能力短板[14]。从整体应用来看, LLM 能够快速批量生成符合基本规范的教学案例, 可有效缓解当前护理模拟教学案例资源不足、开发周期长的问题[15], 护理教育者可借助 LLM 完成案例的初稿生成, 再通过人工修订补充跨学科协作细节, 能够大幅提升案例开发效率, 适合作为护理案例开发的辅助工具使用。

在案例应用效果评价中, 评价者普遍认为案例有助于培养临床决策、沟通协调等能力, 这与模拟教学以能力为导向的目标一致[15]。然而, 案例在直接传递具体疾病知识方面的感知较弱, 这可能与以案例为中心的教学模式中知识习得的隐性特征有关。这提示教学设计者, 应在案例中嵌入明确的知识点问题引导或决策点提示, 并辅以案后的结构化讨论与反思, 才能将案例中蕴含的隐性知识显性化, 实现能力与知识培养的平衡[3]。

从评价主体来看, 本次由护理硕士研究生作为评价者的结果也进一步验证了学生视角开展 AI 生成案

例评价的可行性。护理硕士作为护理模拟教学的直接使用者,更能准确感知案例在能力培养、临床适配性等方面的实际应用价值[16],其评价结果能够为 LLM 生成案例的优化提供贴合教学实际的方向,弥补了仅靠专家评价忽略使用者体验的不足。未来在开发护理教学案例时,可进一步引入使用者视角的常态化评价机制,及时根据学生反馈调整 LLM 生成的提示词框架与内容要求,让生成案例更贴合真实教学需求。

本研究存在一定的局限性。首先,评价对象仅为单一院校的 23 名护理硕士研究生,样本覆盖范围有限,研究结果的代表性与推广价值有待进一步验证。其次,案例评价主体较为单一,尚未充分纳入临床护理、护理教育及模拟教学等不同专业视角,评价结果可能存在一定的主观性。再次,本研究主要依赖量化评分,未能深入解释不同模型生成案例在叙事逻辑、临床真实性、角色互动及伦理冲突设计等方面产生差异的具体原因。未来应扩大样本量并开展多中心研究,建立学生与多学科专家共同参与的评价体系,结合内容分析或主题分析深化案例质量评价。

综上所述,本研究首次系统证实了 DeepSeek R1、GEMINI 与豆包在结构化提示词引导下,均能生成质量相当且符合护理硕士培养需求的跨学科连续照护案例。尽管当前模型在跨学科协作深度精准嵌入上仍有提升空间,但其作为高效教学资源开发辅助工具的可行性已得到验证。未来研究应致力于扩大多中心样本以增强结论稳健性,并通过精细化提示词工程深化案例情境的真实感与复杂性。需开展纵向教学实证研究,量化评估 LLM 生成案例对学生临床决策能力及岗位胜任力的实际增益,同时构建“人机协同”的动态质量监控体系,从而推动人工智能技术在护理教育中的安全、深度融入与应用范式革新。

基金项目

2025 年度湖北医药学院研究生教育教学研究课题重点项目(编号: YJ2025020); 2025 年度湖北医药学院研究生教育教学研究课题一般项目(编号: YJ2025032)。

参考文献

- [1] Yuan, Y. (2025) Impact of Multidisciplinary Continuity of Care on Postoperative Outcomes in Liver Cancer Surgical Patients. *Journal of Multidisciplinary Healthcare*, **18**, 4749-4759. <https://doi.org/10.2147/jmdh.s527399>
- [2] 葛向煜, 贾守梅, 胡雁, 等. 健康信息方向护理硕士专业学位研究生培养的探索与实践[J]. 中华护理教育, 2024, 21(7): 828-834.
- [3] Zhang, C., Xu, C., Wang, R., Han, X., Yang, G. and Liu, Y. (2024) The Learning Experiences and Career Development Expectations of Chinese Nursing Master's Degree Students: A Qualitative Investigation. *Nurse Education in Practice*, **77**, Article 103996. <https://doi.org/10.1016/j.nepr.2024.103996>
- [4] 郭欣怡, 李琨, 赵娟娟, 等. 护理硕士专业学位研究生高级健康评估的进阶式课程设计和多梯度仿真模拟教学实践[J]. 中国医学教育技术, 2024, 38(1): 81-86.
- [5] Franco-Tantuico, M.A. and Btoush, R. (2025) Simulation Debriefing in Graduate Nursing Education: A Literature Review. *Nursing Education Perspectives*, **46**, 278-283. <https://doi.org/10.1097/01.nep.0000000000001434>
- [6] Laurito, W., Davis, B., Grietzer, P., Gavenčiak, T., Böhm, A. and Kulveit, J. (2025) AI-AI Bias: Large Language Models Favor Communications Generated by Large Language Models. *Proceedings of the National Academy of Sciences*, **122**, e241569712. <https://doi.org/10.1073/pnas.2415697122>
- [7] Thirunavukarasu, A.J., Ting, D.S.J., Elangovan, K., Gutierrez, L., Tan, T.F. and Ting, D.S.W. (2023) Large Language Models in Medicine. *Nature Medicine*, **29**, 1930-1940. <https://doi.org/10.1038/s41591-023-02448-8>
- [8] Ruggiano, N., Sahoo, S., Brashear, A., Nwatu, U., Brunson, A., Noh, H., et al. (2026) Evaluating AI-Generated Geriatric Case Studies for Interprofessional Education: Systematic Analysis across 5 Platforms. *JMIR Medical Education*, **12**, e83085. <https://doi.org/10.2196/83085>
- [9] 欧阳子瑶, 樊德净, 潘晓, 等. 护理硕士专业学位研究生教学临床案例库评价指标体系的构建[J]. 中华护理教育, 2023, 20(6): 665-671.
- [10] 齐星伦, 姚一帆, 沈舒施, 等. 不同大语言模型肿瘤标志物报告解读性能评价[J]. 检验医学, 2025, 40(11): 1075-1081.

-
- [11] 黄慧, 胡瑾瑜, 王晓宇, 等. 不同大型语言模型与不同水平医学专业人士回答眼科问题的对比研究[J]. 国际眼科杂志, 2024, 24(3): 458-462.
- [12] 赵朋伟, 李干, 薛紫阳, 等. 以 ChatGPT-4o 和 DeepSeek-V3 为代表的生成式人工智能在[12]肠造口患者教育支持中的对比研究[J]. 军事护理, 2025, 42(12): 71-74.
- [13] Jahnke, M.N., O'Haver, J., Gupta, D., Hawryluk, E.B., Finelt, N., Kruse, L., *et al.* (2021) Care of Congenital Melanocytic Nevi in Newborns and Infants: Review and Management Recommendations. *Pediatrics*, **148**, e2021051536. <https://doi.org/10.1542/peds.2021-051536>
- [14] Cao, W., Zhang, Q., Liu, J. and Liu, S. (2026) From Agents to Governance: Essential AI Skills for Clinicians in the Large Language Model Era. *Journal of Medical Internet Research*, **28**, e86550. <https://doi.org/10.2196/86550>
- [15] Tong, K., McMahon, E., Reid-McDermott, B., Byrne, D. and Doherty, A.M. (2021) SafePsych: Improving Patient Safety by Delivering High-Impact Simulation Training on Rare and Complex Scenarios in Psychiatry. *BMJ Open Quality*, **10**, e001533. <https://doi.org/10.1136/bmjog-2021-001533>
- [16] Xue, H., Yuan, H., Li, G., Liu, J. and Zhang, X. (2021) Comparison of Team-Based Learning vs. Lecture-Based Teaching with Small Group Discussion in a Master's Degree in Nursing Education Course. *Nurse Education Today*, **105**, Article 105043. <https://doi.org/10.1016/j.nedt.2021.105043>