

# DeepSeek 护理个案理解能力测试与应用展望

王东丽 邵 帅 丁 磊 陆玉梅 严 颖 丁菊红

(南通大学附属南通第三医院重症医学科 南通 226006)

**〔摘要〕** **目的/意义** 探讨 DeepSeek 对护理个案的理解能力, 展望其在护理教育和临床辅助决策中的应用前景。**方法/过程** 选取《护理综合案例分析》中 28 个护理个案的 120 道开放性问答题, 利用 DeepSeek - R1 - 671b 生成答案。两名评分者依据标准答案, 从准确性、全面性、实用性、简洁性和条理性 5 个维度, 对模型生成答案和答案解析进行独立评分, 并计算总体正确率。**结果/结论** DeepSeek 能较准确理解并回答护理个案问题, 在护理教育和临床辅助决策中有一定应用潜力。

**〔关键词〕** DeepSeek; 大语言模型; 护理个案; 模型评估

**〔中图分类号〕** R - 058 **〔文献标识码〕** A **〔DOI〕** 10.3969/j.issn.1673-6036.2026.02.015

## Evaluation of DeepSeek's Comprehension Ability in Nursing Case Analysis and Its Application Prospects

WANG Dongli, SHAO Shuai, DING Lei, LU Yumei, YAN Ying, DING Juhong

Department of Critical Care Medicine, The Third Affiliated Hospital of Nantong University, Nantong 226006, China

**〔Abstract〕** **Purpose/Significance** To explore DeepSeek's comprehension ability of nursing cases, and to look forward its application prospects in nursing education and clinical decision support. **Method/Process** A total of 120 open-ended questions from 28 nursing cases in the *Comprehensive Case Analysis of Nursing* are selected. The DeepSeek - R1 - 671b is used to generate answers. Two graders independently score the answers based on five dimensions: accuracy, comprehensiveness, practicality, conciseness and logical clarity, and the overall accuracy rate is calculated. **Result/Conclusion** DeepSeek can understand and answer the questions of nursing cases with reasonable accuracy and shows potential for application in nursing education and clinical decision support.

**〔Keywords〕** DeepSeek; large language model; nursing case; model evaluation

## 1 引言

大语言模型 (large language model, LLM) 基于深度学习技术, 通过海量数据训练, 具备文本信息

理解、生成与推理能力<sup>[1]</sup>。近年来 LLM 快速发展, 代表性模型有 ChatGPT、Gemini、豆包、文心一言和 DeepSeek 等。与国外 LLM 相比, DeepSeek 更适应中文语言环境, 能够更好地理解中文语境与文化背景, 从而更精准地满足我国本土化需求<sup>[2]</sup>。与其

**〔修回日期〕** 2025 - 12 - 07

**〔作者简介〕** 王东丽, 副主任护师, 发表论文 5 篇; 通信作者: 邵帅, 护师。

**〔基金项目〕** 南通大学临床医学专项科研基金 (项目编号: 2024HY016)。

他国产 LLM 相比, DeepSeek 处理连续性任务更稳定, 文本生成更自然, 推理能力更强<sup>[3]</sup>。在中国知网以“DeepSeek”为主题进行检索, 自 2025 年 1 月 20 日 DeepSeek - R1 发布至 2025 年 5 月 31 日, 已有 1 711 篇相关文献发表, 其中学术期刊论文 652 篇, 表明 DeepSeek 已成为人工智能领域研究热点之一。目前, LLM 已广泛应用于机器翻译、软件开发、创意写作、教育辅导等领域, 其在护理学领域也正得到快速推广与发展。为确保其在护理教育、临床等场景中应用的准确性与安全性, 应先行对其护理知识掌握情况进行评估与验证<sup>[4]</sup>。有研究<sup>[5]</sup>表明 LLM 在护理基础知识理解与应用方面表现优异; 但是对于 LLM 在复杂临床情境中的个案分析与高阶推理能力, 相关研究仍较少。通过评估 LLM 对护理个案的理解和推理能力, 可判断其在临床辅助决策、患者教育等高风险场景中的潜在价值<sup>[5-6]</sup>。本研究评估 DeepSeek 模型护理个案答题表现, 为在护理教育及临床辅助决策中的应用提供参考。

## 2 资料与方法

### 2.1 资料来源

以科学出版社出版的《护理综合案例分析》<sup>[7]</sup>为资料来源。该书收录 140 余例临床案例, 涵盖基础护理学、内科护理学、外科护理学、妇产科护理学、儿科护理学等护理学科, 每个案例设 3~5 个情境化开放性问答题。案例纳入标准: 题干完整, 问题有明确参考答案; 内容与临床护理实践相关, 能反映护理工作实际问题。排除标准: 题干模糊, 无法理解含义; 缺乏答案解析或答案争议较大; 与临床护理实践不相关。最终纳入 28 个案例, 共 120 道题目。

### 2.2 研究方法

**2.2.1 评价指标** 利用 DeepSeek 解答纳入题目。从准确性、全面性、实用性、简洁性和条理性 5 个维度评价 DeepSeek 生成答案与答案解析。每个维度 1—5 分, 1 分为最差, 5 分为最佳。5 个维度得分相加为总分, 总分范围为 5~25 分, 得分越高表示

答题质量越高<sup>[8]</sup>。由两名具有副主任护师职称的专家独立评分。两名评分者从事护理专业年限分别为 15 年和 17 年, 其中 1 名为硕士研究生导师。各维度最终得分取二者评分均值。如果准确性维度得分  $\geq 4$  分且未出现关键性理解错误, 判定为“回答正确”; 如果得分  $< 4$  分, 则判定为“回答不正确”。总体准确率 = 正确题目数/纳入总题数  $\times 100\%$ 。

**2.2.2 样本量计算** 使用 PASS15 软件进行样本量计算, 采用配对样本  $t$  检验 ( $t$ -Test Inequality 模型), 设定 Power 为 0.8,  $\alpha$  为 0.05, 得出所需样本量为 100<sup>[9]</sup>。本研究纳入 120 道题目, 超过最低样本量要求。

**2.2.3 盲法** 为减少评分主观性, 每道题由两名评分者独立评分。将 DeepSeek 生成答案与答案解析随机标记为“答案 A”和“答案 B”。评分者不知晓答案来源, 独立阅读案例及答案进行评分, 以确保公正客观。如果两名评分者单项评分差距达到 2 分及以上, 邀请第 3 名评分者参与评价, 并以 3 人平均分作为最终得分, 提高准确性和可靠性<sup>[10]</sup>。

**2.2.4 DeepSeek 模型的接入** 截至 2025 年 2 月 20 日, DeepSeek 已发布手机版、网页版、DeepSeek - V3 版以及 DeepSeek - R1 版。其中, DeepSeek - R1 - 671b 是当前参数规模最大 (6 710 亿参数)、性能最强的版本, 具备语言理解、逻辑推理与语义整合能力, 在复杂任务中准确性和稳定性更高。同时, 其知识覆盖面广, 有实时网络检索功能, 且数据更新至 2024 年 7 月, 能够胜任高精度信息处理任务, 尤其适用于高复杂度的专业任务与科研场景。DeepSeek - R1 - 671b 模型支持本地部署和通过应用程序接口 (application programming interface, API) 接入云端服务平台两种方式。由于该模型计算量大, 对硬件配置要求高, 本地部署难以满足稳定运行需求<sup>[11-12]</sup>。因此采用 API 接入方式, 在电脑端安装硅基流动公司提供的接口程序, 注册账号后, 通过 API 接入其超级计算机平台, 进行模型运算, 每生成 1 000 个字符费用约 0.011 元。

**2.2.5 输入指令调试** DeepSeek 可能受“幻觉”现象影响, 生成看似合理但实际不准确的答案<sup>[13]</sup>。为此设计输入指令时应避免模糊或多义性表述, 从

源头减少“幻觉”<sup>[14]</sup>。经反复实验与调试，确定输入指令分两步。第 1 步：命令 DeepSeek 分析护理个案，“你是一位护理领域专家，请对我提供的护理个案做详细分析”。第 2 步，命令 DeepSeek 基于个案信息回答具体问题。例如，对“该患者为何要进行肺功能检查”，输入指令为：“该患者为何要进行肺功能检查？（仅基于个案信息，不臆测，只回答问题本身即可）”。通过边界设置，可避免过度推测或解读，减少模型“幻觉”，提升作答相关性、评分效率与评分一致性<sup>[15]</sup>。

### 2.3 统计学方法

采用 SPSS 26.0 软件进行数据统计分析。计量资料符合正态分布时用独立样本 *t* 检验，不符合正态分布时用秩和检验；计数资料用  $\chi^2$  检验比较组间差异。统计学显著性水平设定为  $P < 0.05$  为有统计学意义<sup>[16]</sup>。

## 3 结果

### 3.1 DeepSeek 与答案解析组间得分比较

两组得分的正态分布检验结果显示，所有维度得分均不符合正态分布，因此采用秩和检验。数据分析显示，DeepSeek 在全面性、条理性和总分方面得分均高于答案解析 ( $P < 0.05$ )，在准确性、实用性和简洁性方面得分均低于答案解析 ( $P < 0.05$ )，见表 1。DeepSeek 在准确性、全面性、实用性、简洁性和总分维度的标准差大于答案解析 ( $P < 0.05$ )，见表 2。

表 1 DeepSeek 与答案解析得分比较  
[分,  $M (P_{25}, P_{75})$ ]

维度	答案解析	DeepSeek	Z	P
准确性	5 (5, 5)	4.75 (4, 5)	-5.029	<0.001
全面性	4.5 (4, 4.5)	5 (5, 5)	-10.906	<0.001
实用性	4 (4, 4.5)	4 (3.5, 4.5)	-2.557	0.011
简洁性	4 (4, 4.5)	4 (3, 4)	-4.415	<0.001
条理性	4 (4, 4.5)	5 (4.5, 5)	-10.797	<0.001
总分	21.5 (21, 22)	22 (21, 23)	-2.984	0.003

表 2 DeepSeek 与答案解析得分标准差比较 (分,  $\bar{x} \pm s$ )

维度	答案解析	DeepSeek	F	P
准确性	4.862 ± 0.267	4.550 ± 0.520	65.84	<0.001
全面性	4.287 ± 0.347	4.863 ± 0.343	11.76	<0.001
实用性	4.171 ± 0.358	3.954 ± 0.719	32.84	<0.001
简洁性	4.067 ± 0.330	3.713 ± 0.741	31.66	<0.001
条理性	4.133 ± 0.354	4.804 ± 0.357	0.78	0.380
总分	21.521 ± 0.795	21.883 ± 1.587	18.72	<0.001

### 3.2 DeepSeek 与答案解析得分组内比较

两组得分均不符合正态分布，因此采用秩和检验。经 Bonferroni 校正，DeepSeek 组在“简洁性 - 条理性”和“全面性 - 准确性”两个维度的比较中  $P > 0.05$ ，差异无统计学意义，其余各维度间的比较差异有统计学意义 (均  $P < 0.05$ )。答案解析组仅在“简洁性 - 全面性”“简洁性 - 准确性”及“条理性 - 准确性”3 个维度的比较中差异有统计学意义 (均  $P < 0.05$ )，其余维度间的比较差异无统计学意义，见表 3。DeepSeek 组的 *H* 值 (255.520) 略高于答案解析组 (246.633)。

表 3 DeepSeek 与答案解析组内得分比较

主维度	比较维度	答案解析		DeepSeek	
		P	P (Bonferroni 校正后)	P	P (Bonferroni 校正后)
简洁性	条理性	0.220	1	0.035	0.345
	实用性	0.053	0.533	<0.001	<0.001
	全面性	<0.001	<0.001	<0.001	<0.001
条理性	准确性	<0.001	<0.001	<0.001	<0.001
	实用性	0.481	1	<0.001	<0.001
	全面性	0.003	0.032	<0.001	<0.001
实用性	准确性	<0.001	<0.001	<0.001	<0.001
	全面性	0.025	0.252	0.001	0.009
	简洁性	<0.001	<0.001	<0.001	<0.001
全面性	准确性	<0.001	<0.001	0.361	1

### 3.3 DeepSeek 与答案解析答题正确率比较

以准确性维度得分  $\geq 4$  为“回答正确”，答案解析总体准确率为 100%，DeepSeek 总体准确率为 85.8%，差异有统计学意义 ( $\chi^2 = 18.3, P < 0.001$ )，见表 4。

表 4 DeepSeek 与答案解析答题正确率比较

评估项目	答案解析	DeepSeek	$\chi^2$	<i>P</i>
题目总数 (道)	120	120	18.3	<0.001
准确性得分 $\geq 4$ 的题目数 (道)	120	103		
准确性得分<4的题目数 (道)	0	17		
总体准确率 (%)	100	85.80		

## 4 讨论

### 4.1 DeepSeek 答题优势分析

DeepSeek 在全面性和条理性维度得分高于答案解析, 差异具有统计学意义。分析其原因, 主要包括两方面。一是 DeepSeek - R1 - 671b 模型数据库更新至 2024 年 7 月, 具备联网检索功能, 结合深度学习算法和大规模文本预训练, 能够高效获取护理学各学科的相关数据, 因此在全面性维度上得分高于答案解析<sup>[17]</sup>。例如, 在一例急性心肌梗死护理个案中, DeepSeek 回答涵盖患者病史、症状、诊断步骤、护理目标、治疗干预和护理评估 6 个方面。与答案解析相比, 治疗干预部分增加药物治疗、患者生活方式干预及心理护理等内容, 能更好地满足护理实践多维度需求。二是 DeepSeek 强大语言理解和推理能力使其能准确解读问题含义, 抓住关键信息进行合理推理, 在大规模文本数据中提取有效信息, 确保回答逻辑性和条理性。因此, DeepSeek 凭借信息检索、语言理解和推理优势, 在全面性和条理性维度得分优于答案解析<sup>[18]</sup>。

### 4.2 DeepSeek 答题不足分析

DeepSeek 在准确性、实用性和简洁性维度得分低于答案解析, 差异具有统计学意义, 其总体准确率也低于答案解析。分析其原因, 主要包括 3 方面。一是 DeepSeek 从数据库提取信息, 进行分析并得到答案, 对于开放式护理个案问题, 其回答易过度解读或与题目关联性较弱<sup>[19]</sup>。例如, 在一例急性心肌梗死护理个案中, DeepSeek 提供大量病理机制背景信息, 却缺少具体护理措施和临床干预, 导致实用性得分较低。而答案解析中多为实际护理干预和临床操作, 实用性维度得分高。二是某些问题存

在多种解读方式, DeepSeek 在生成答案时有时偏向理论层面的多角度解读, 缺乏针对临床场景的具体决策, 导致准确性得分低于答案解析<sup>[20]</sup>。此外, DeepSeek 的答案有时过于理论化, 未能充分考虑临床情境、患者实际需求及操作的可行性, 进而影响了实用性得分<sup>[21]</sup>。三是 DeepSeek 为确保全面性, 倾向提供更详细的信息, 将多个相关知识点和背景信息融入生成答案, 未能有效简化内容, 导致简洁性得分较低<sup>[22]</sup>。相比之下, 答案解析通常经严格审核和精练, 更加准确、实用且简洁<sup>[23]</sup>。

### 4.3 DeepSeek 得分稳定性分析

DeepSeek 在准确性、全面性、实用性、简洁性和总分维度标准差大于答案解析, 表明其得分波动大, 稳定性较差。DeepSeek 基于大规模数据训练, 数据多样性可能导致模型处理不同类型问题时表现不一致<sup>[24]</sup>。尽管其具备强大推理能力, 但其推理依赖于大规模数据训练, 而非对问题的深入理解, 处理有歧义或多种解读方式的问题时, DeepSeek 的理解可能偏离题目核心, 影响答案准确性。其输出常包含多维度详细信息, 易使回答简洁性差, 得分波动大。相比之下, 答案解析通常经精练, 答案简洁且符合实际护理需求, 得分波动小。因此, 尽管 DeepSeek 在信息处理和推理能力方面具有优势, 但面对复杂和歧义性问题时, 得分波动性较答案解析更大<sup>[25]</sup>。

### 4.4 DeepSeek 模型在护理领域应用展望

DeepSeek 在准确性、全面性和条理性维度平均分均超过 4 分, 总分平均 21.883 分, 高于答案解析, 表明其能有效处理护理个案问题且在多个维度表现突出。DeepSeek 具备数据检索、信息提取和分析功能, 能帮助护理人员提升工作效率, 推动护理服务智能化发展。护理教育者可利用 DeepSeek 辅助教学, 临床护士可将其用于辅助决策, 科研人员则可借助其进行数据分析与资料整理。此外, DeepSeek 能整合医学、护理学、心理学及中医学等多学科知识, 解决跨学科问题<sup>[26]</sup>。然而, DeepSeek 应作为辅助工具, 非决策主导者, 确保护理人员保持独立判断, 避免过

度依赖<sup>[27]</sup>。同时, DeepSeek 可能面临信息泄露风险, 护理人员使用时应加强隐私保护, 防止泄露医疗信息和患者隐私。未来应为护理人员提供专业的 DeepSeek 应用培训, 确保使用安全<sup>[28]</sup>。

## 5 结语

DeepSeek 在护理个案理解与作答中表现良好, 具备较强的护理学知识理解能力, 可作为护理教育和临床决策辅助工具。但其应用存在一定挑战, 如过度依赖、隐私泄露等问题。本研究仍存在一定局限: 一是主要依靠两名专家参考教材“标准答案”进行人工评分, 可能存在一定主观偏倚; 二是仅从 5 个维度评分, 尚无法完全捕捉模型在处理复杂临床决策时的细微逻辑差异。未来研究可进一步优化评价体系, 邀请更多元化的临床专家参与, 并引入基于真实循证指南的多维度指标, 从而更客观、更深入地评估大语言模型在临床一线环境下的适用性与实际效果。

**作者贡献:** 王东丽负责研究设计、论文撰写; 邵帅负责提供指导、论文修订; 丁磊负责数据收集与分析; 陆玉梅、严颖、丁菊红参与数据分析、论文审核。

**利益声明:** 所有作者均声明不存在利益冲突。

## 参考文献

- 邱锡鹏. 解剖大型语言模型: 原理、应用及影响 [J]. 探索与争鸣, 2023 (5): 10-12.
- 龚中楠, 赵俊强, 任文杰, 等. 大语言模型在护理教育领域应用的范围综述 [J]. 中国医学教育技术, 2025, 39 (5): 588-594.
- 闫海荣, 江瑞, 张学工, 等. DeepSeek 与医学大语言模型: 技术创新与医疗服务模式重构 [J]. 医学信息学杂志, 2025, 46 (2): 1-7.
- 王绍源, 杨东航, 任宇东. 大语言模型在护理领域的应用场景与伦理探讨 [J]. 护理学杂志, 2025, 40 (5): 108-113.
- SHOOL S, ADIMI S, SABOORI AMLESHI R, et al. A systematic review of large language model (LLM) evaluations in clinical medicine [J]. BMC medical informatics and decision making, 2025, 25 (1): 117.
- GOODMAN K E, ROBINSON M L, SHAMS S M, et al. Identification of long-term care facility residence from admission notes using large language models [J]. JAMA network open, 2025, 8 (5): e2512032.
- 王涛, 张华, 蒙莉萍, 等. 护理综合案例分析 [M]. 北京: 科学出版社, 2019.
- 徐文博, 陈凤敏, 王超, 等. GPT-4 大语言模型对护理知识理解的测试研究 [J]. 护理学杂志, 2024, 39 (19): 93-96.
- 袁建文, 李科研. 关于样本量计算方法的比较研究 [J]. 统计与决策, 2013 (1): 22-25.
- DAI Z, WANG F, SHEN C, et al. Accuracy of large language models for literature screening in thoracic surgery: diagnostic study [J]. Journal of medical internet research, 2025, 27 (3): e67488.
- 庄世正, 刘澎涛, 李璇, 等. DeepSeek 本地部署方案研究与探索 [J]. 网络新媒体技术, 2025, 14 (3): 1-9.
- 陈行. 基于 Deepseek-R1 大模型本地部署知识问答方法比较与分析——以建筑业数据分析为例 [J]. 住宅与房地产, 2025 (15): 13-17.
- PUGH S L, CHANDLER C, COHEN A S, et al. Assessing dimensions of thought disorder with large language models: the tradeoff of accuracy and consistency [J]. Psychiatry research, 2024, 341 (11): 116119.
- DUMIT J, ROEPSTORFF A. AI hallucinations are a feature of LLM design, not a bug [J]. Nature, 2025, 639 (8053): 38.
- LEE J H, SHIN J. How to optimize prompting for large language models in clinical research [J]. Korean journal of radiology, 2024, 25 (10): 869-873.
- 李康, 贺佳. 医学统计学 [M]. 北京: 人民卫生出版社, 2024.
- WILLIAMS C Y K, SUBRAMANIAN C R, ALI S S, et al. Physician- and large language model-generated hospital discharge summaries [J]. JAMA network open, 2025, 185 (7): 818-825.
- 刘漫, 沈鹏熠, 张茹梦. 人工智能技术在医疗护理中的应用研究 [J]. 中国农村卫生事业管理, 2025, 45 (3): 177-182.
- 洪学海, 史峰. AI 大模型驱动智库高质量发展的思考——基于开源 DeepSeek R1 大模型应用的部分实证 [J]. 智库理论与实践, 2025, 10 (4): 1-8.

(下转第 109 页)

- 8 JOSEPH C, RAYMOND R, SUMANTH D, et al. Transforming medical education: assessing the integration of ChatGPT into faculty workflows at a caribbean medical school[J]. *Cureus*, 2023, 15(7): e41399.
- 9 MACALHAES A, CORREIA C R. Incorporating ChatGPT in medical informatics education: mixed methods study on student perceptions and experiential integration proposals [J]. *JMIR medical education*, 2024(10): e51151.
- 10 MALIK S. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns[J]. *Healthcare*, 2023, 11(6): 887.
- 11 SRIDHARAN K, SEQUEIRA R P. Artificial intelligence and medical education: application in classroom instruction and student assessment using a pharmacology & therapeutics case study[J]. *JMIR medical education*, 2024, 24(1): 431.
- 12 袁欢欢, 全世悦, 林祎, 等. 当医学课程考核遇上生成式人工智能: 价值、挑战与未来展望[J]. *中国医学教育技术*, 2023, 37(3): 254-259.
- 13 廖宁, 陈怡然, 杨倩, 等. 以“三平台、三结合”行业实践为引领的教学创新研究——以机器学习课程为例[J]. *科教导刊*, 2023(14): 22-24.
- 14 瞿星, 杨金铭, 陈滔, 等. ChatGPT 对医学教育模式改变的思考[J]. *四川大学学报(医学版)*, 2023, 54(5): 937-940.
- 15 王小梨, 冯影, 吴润杰. 电子信息专业硕士研究生机器学习课程交叉应用案例库建设研究[J]. *现代信息科技*, 2024, 8(12): 193-198.
- 16 黄兰.“专创融合”视域下生物医学工程专业科研育人模式探索[J]. *佳木斯职业学院学报*, 2024, 40(7): 198-200.
- 17 陈依桐, 陈建, 朱鲁闯, 等. 项目制驱动的医工融合工程实践课程体系研究与实践[J]. *实验科学与技术*, 2024, 22(4): 41-46.
- 18 曹建荣, 李成栋, 孙雪梅, 等. 深度学习课程实践教学案例库建设[J]. *计算机教育*, 2024(7): 124-128.
- 19 赵卫东. 面向课程群的递进式教学案例库一体化设计[J]. *计算机教育*, 2023(5): 160-164.
- 20 方赵嵩, 郑志敏, 冯锡文, 等. 基于医工交叉的人工环境方向人才培养探究[J]. *制冷*, 2024, 43(4): 72-76.
- 21 康红艳, 周钢, 刘美丽, 等. 面向医工交叉人才培养的基础医学实验教学教学改革[J]. *实验室研究与探索*, 2024, 43(7): 183-186, 205.
- 22 黄明芳, 侯青涵, 张伟. 生成式人工智能在医学教育领域的应用现状与未来趋势[J]. *医学与社会*, 2025, 38(1): 29-34, 47.
- 23 江哲涵, 奉世聪, 王维民. 人工智能生成内容在医学教育中的应用、挑战与展望[J]. *中国教育信息化*, 2024, 30(8): 29-40.
- 24 赵欣, 柳云. 生成式人工智能赋能医学人文教育的伦理审视[J]. *临床荟萃*, 2024, 39(1): 65-69.
- 25 齐凤林, 沈佳杰, 王茂异, 等. 人工智能在高校信息化中的应用研究综述[J]. *计算机工程*, 2025, 51(4): 1-14.

(上接第 102 页)

- 20 孙中海, 赖小琴, 颜怿炜. DeepSeek 医院私有化部署的实现与应用探索 [J]. *中国卫生信息管理杂志*, 2025, 22(3): 456-463.
- 21 马红燕, 周晓, 张碧瑶, 等. DeepSeek 大语言模型的处方合理性评估中的实用性 [J/OL]. *医药导报*, 1-13 [2025-08-12]. <https://link.cnki.net/urlid/42.1293.R.20250528.1116.002>.
- 22 黄慧瑛, 欧阳汉栋, 林胜钊. 探讨 DeepSeek 在护理工作中的应用场景及面临的挑战和应对策略 [J]. *广州医药*, 2025, 56(5): 591-598.
- 23 PATIL A, SERRATO P, CHISVO N, et al. Large language models in neurosurgery: a systematic review and meta-analysis [J]. *Acta neurochirurgica*, 2024, 166(1): 475.
- 24 HARTMAN V, ZHANG X, PODDAR R, et al. Developing and evaluating large language model-generated emergency medicine handoff notes [J]. *JAMA network open*, 2024, 7(12): e2448723.
- 25 陈昌茂, 张瑶, 谭韦池, 等. 基于 DeepSeek 大语言模型的医院智能数据分类分级探索及应用 [J]. *中国数字医学*, 2025, 20(6): 30-36.
- 26 闫温馨, 刘珏, 梁万年. DeepSeek 赋能全科医学: 潜在应用与展望 [J]. *中国全科医学*, 2025, 28(17): 2065-2069.
- 27 王绍源, 李梦. 从 ChatGPT4.0 到 DeepSeek-R1: 人工智能在医疗卫生领域应用的革新场景和伦理治理范式转变 [J/OL]. *中国医学伦理学*, 1-9 [2025-08-12]. <https://link.cnki.net/urlid/61.1203.R.20250704.1632.014>.
- 28 刘金枝, 刘万鹏, 高跃. DeepSeek 类生成式人工智能助推医患命运共同体的构建研究 [J]. *中国医学教育技术*, 2025, 39(5): 583-587.