·数智医学教育·

静脉血栓栓塞症智能知识库的构建及在医学教育中的应用评价

佘康[1]　郝雨迪[1]　王亚青[1]　李俊梅[1]　马颖林[2]　母光妍[3]　李珂[4]　张晓琳[4]
齐建光[4]　齐心[5]
[1]北京大学第一医院介入血管外科，北京 100034；[2]国家癌症中心/国家肿瘤临床医学研究中心/中国医学科学院北京协和医学院肿瘤医院药剂科，北京 100021；[3]北京大学第一医院药剂科，北京 100034；[4]北京大学第一医院教育处，北京 100034；[5]北京大学第一医院烧伤整形科，北京 100034
通信作者：齐心，Email：randyq@126.com

【摘要】 目的 构建基于大语言模型的检索增强生成技术的静脉血栓栓塞症（venous thromboembolism，VTE）专业知识库，并评价其在医学教育中的应用效果。**方法** 采用“循证证据—专家经验—真实病例”三层语料，在腾讯 ima 平台搭建 VTE 知识库，并实施证据分级、版本控制和人机协同审核。采用混合方法研究：学习者路径通过横断面调查评价医学生及非专科医护人员（n=131）在临床决策支持、知识问答和资源推荐中的使用体验；教师路径采用单盲随机对照设计，评价血管外科教师（n=19）在教案设计、病例讨论和辅助命题中的应用效果。统计分析在 R 4.5.0 软件中完成；计数资料采用频数和构成比描述，Likert 量表评分采用中位数和四分位间距描述，组间比较采用 Kruskal-Wallis H 检验、Friedman 秩和检验及 Wilcoxon 符号秩检验。**结果** 学习者路径显示，知识库在标准化任务中应用效果较好，各维度满意度评分的中位数均为 4.00 分[四分位间距（interquartile range，IQR）为 4.00~5.00]，但循证可信度评分相对较低。教师路径显示，其应用效果具有任务依赖性：在“教案设计”等收敛性任务中，深度使用知识库的适用性略高于入门使用；在“辅助命题”等创造性任务中，深度使用知识库的适用性低于入门使用（P<0.05）。**结论** 基于检索增强生成（retrieval-augmented generation，RAG）技术构建的 VTE 专业知识库在医学教育中具有一定应用价值，但其应用边界尚不明确。未来应通过构建涵盖教学法、亚专科及科研数据的“知识库矩阵”，进一步提升其在真实医学教育场景中的应用效果。

【关键词】 医学教育；　静脉血栓栓塞；　大语言模型；　检索增强生成；　知识库
【中图分类号】 R-4
基金项目：北京大学教学新思路 3.0 立项资助（2026XSL10）；北京大学第一医院院级教学课题（2025YJ09）
DOI: 10.3760/cma.j.cn116021-20260128-02296

Construction of an intelligent knowledge base for venous thromboembolism and evaluation of its application in medical education

She Kang[1], Hao Yudi[1], Wang Yaqing[1], Li Junmei[1], Ma Yinglin[2], Mu Guangyan[3], Li Ke[4], Zhang Xiaolin[4], Qi Jianguang[4], Qi Xin[5]
[1]Department of Interventional Vascular Surgery, Peking University First Hospital, Beijing 100034, China; [2]Department of Pharmacy, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China; [3]Department of Pharmacy, Peking University First Hospital, Beijing 100034, China; [4]Department of Education, Peking University First Hospital, Beijing 100034, China; [5]Department of Burns and Plastic Surgery, Peking University First Hospital, Beijing 100034, China
Corresponding author: Qi Xin, Email: randyq@126.com

【Abstract】 Objective To construct a venous thromboembolism (VTE) knowledge base based on retrieval-augmented generation powered by a large language model, and evaluate its effectiveness in medical

education. **Methods**　A three-layer corpus consisting of evidence-based information, expert experience, and real-world cases was used to build the VTE knowledge base on the Tencent IMA platform. Evidence grading, version control, and human-AI collaborative review were performed. A mixed-methods design was used. For the learner pathway, a cross-sectional survey was used to evaluate the experience of medical students and non-specialist healthcare workers (n=131) in clinical decision support, knowledge question answering, and resource recommendation. For the teacher pathway, a single-blind randomized controlled design was used to evaluate the effectiveness of application among vascular surgery teachers (n=19) in lesson planning, case discussion, and assisted exam item generation. Statistical analyses were performed using R 4.5.0. Categorical data were described using frequencies and proportions; Likert scale scores were described using medians and interquartile ranges. Inter-group comparisons were performed using the Kruskal-Wallis H test, Friedman test, and Wilcoxon signed-rank test. **Results**　The learner pathway showed that the knowledge base performed well in standardized tasks, with median satisfaction scores of 4.00 (interquartile range: 4.00-5.00) across dimensions, although the evidence-based credibility score was relatively low. The teacher pathway showed that the application effectiveness was task-dependent. For convergent tasks such as lesson planning, intensive use of the knowledge base had slightly higher applicability than basic use. For creative tasks such as assisted exam item generation, intensive use had lower applicability than basic use (P<0.05). **Conclusions**　The VTE knowledge base constructed based on retrieval-augmented generation has certain application value in medical education, but its application boundaries remain unclear. Future efforts should focus on constructing a knowledge base matrix that integrates pedagogy, subspecialty knowledge, and research data to further improve its application effectiveness in real medical education settings.

【**Key words**】 Medical education;　Venous thromboembolism;　Large language model;　Retrieval-augmented generation;　Knowledge base

Fund program: The Innovative Teaching and Learning Program 3.0 of Peking University (2026XSL10); The Educational Research Project of Peking University First Hospital (2025YJ09)

DOI: 10.3760/cma.j.cn116021-20260128-02296

静脉血栓栓塞症（venous thromboembolism，VTE）作为院内可预防死亡的首要原因，其医学教育具有显著的“高认知负荷”特征[1-2]。这是因为，一方面，VTE的诊疗是一个十分复杂的临床过程，并非基于简单的线性规则；在肿瘤、妊娠、围手术期等临床情景下，VTE的治疗需要动态平衡出血风险与血栓形成风险，高度依赖临床经验与情境化推理。另一方面，VTE治疗的循证医学依据更新较快，教科书往往落后于临床实践，学习者可能会出现“理论不全、决策受限”的现象。这种任务复杂性与知识时效性的双重压力会明显增加医护人员的认知负荷，进而影响决策质量[3]。传统的课堂讲授式教学难以在有限的时间内完成如此庞大、繁杂的证据资料讲授及复杂的临床情境模拟，迫切需要一种能够实时结合最新证据并帮助医学生更好地开展临床知识学习与思维训练的数字化工具[4-5]。虽然基于规则的临床决策支持系统已被证明能提高指南遵循度[6]，但其缺乏灵活性与解释性。近年来，大语言模型（large language model，LLM）展现出强大的知识整合能力，但在严肃医学教育场景中，通用模型普遍存在的“幻觉”与引用溯源困难成为其应用瓶颈[7-8]。检索增强生成（retrieval-augmented generation，RAG）技术通过将生成过程锚定在外部权威证据上，能够极大程度地降低幻觉率、提高准确度和可信赖性[9-10]。因此，为了将医学信息更精准地提供给学生，给教师提供更好的教学工具，本研究提出一种“AI+专科语料”的设计思路，即利用基于RAG技术的LLM平台搭建专科知识库，融合权威循证底座、专家经验及真实临床案例形成知识底座，并建立知识库内容更新迭代机制，通过混合方法研究知识库在学生和教师2种路径及多种临床教学场景中的应用效果。

1　知识库构建

1.1　研究设计与技术路线

本研究采用“知识库构建—教学应用—量化评价—讨论”的总体设计思路。首先，围绕VTE构建专科智能知识库，形成涵盖循证证据、专家经验和真实病例的结构化知识底座；其次，将该知识库嵌入医学教育场景，分别面向学习者和教师开展应用；再次，通过问卷调查和单盲随机对照试验评价其在不同教学任务中的应用效果。本研究依托腾

讯 ima 平台完成知识库搭建、资料管理与问答应用。本研究方案经北京大学第一医院生物医学伦理委员会审核批准（审批号：申 2026R0125-0002），所有受试者均知情同意。总体研究技术路线如图 1 所示。

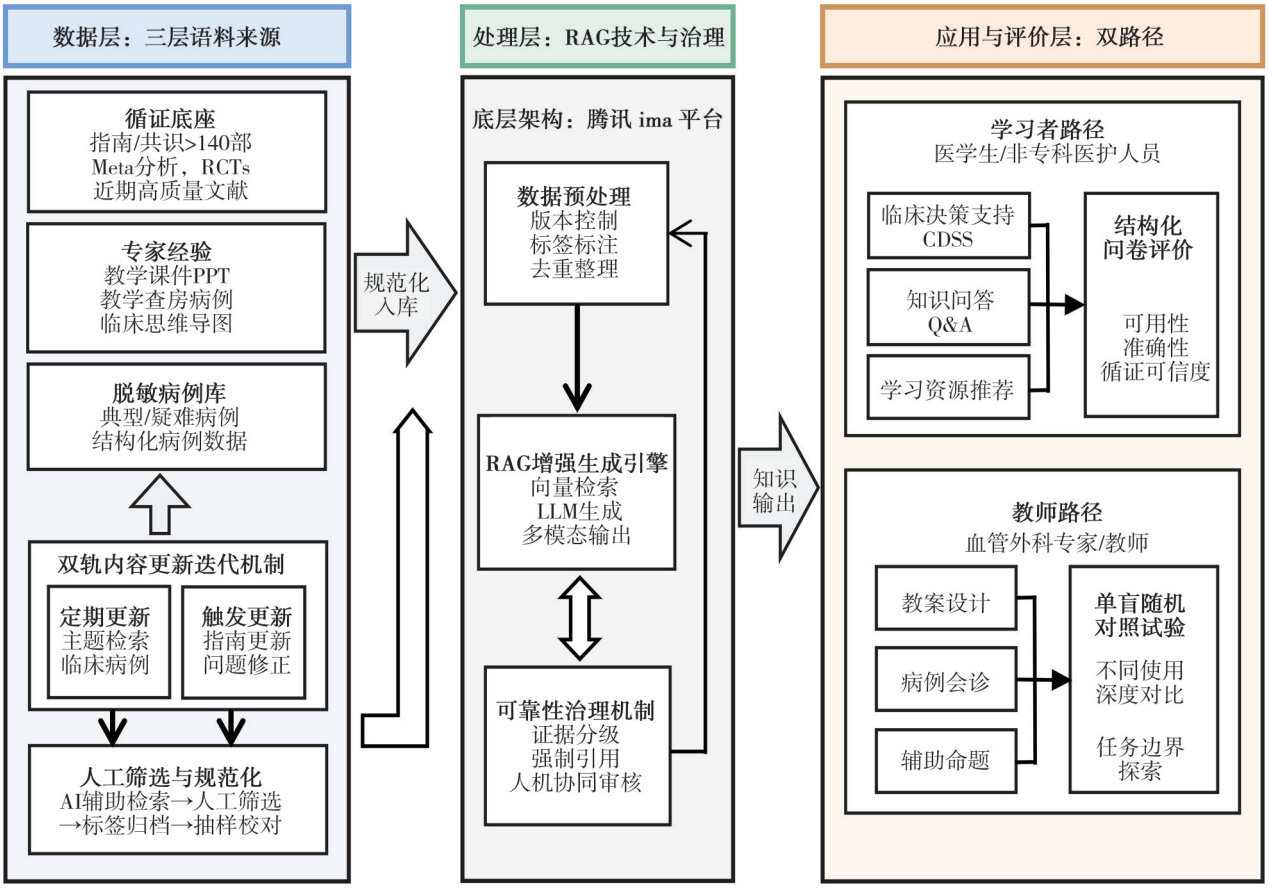
1.2　知识库主题与内容范围

本研究选择 VTE 作为知识库建设主题，主要基于以下考虑：第一，VTE 是院内可预防死亡的重要原因，也是血管外科高频会诊病种，涉及骨科、妇产科、肿瘤科、重症医学科等多个临床场景，具有较高的跨学科教学价值；第二，VTE 诊断、风险评估、抗凝治疗和围手术期管理等内容专业性强，学习者需要在复杂临床情境中综合判断血栓风险与出血风险，认知负荷较高；第三，VTE 相关指南、专家共识和高级别循证证据更新较快，传统教材和课堂讲授难以及时覆盖最新进展。

知识库内容范围主要包括 VTE 风险评估、诊断技术、抗凝治疗、溶栓与介入治疗、特殊人群管理、围手术期预防、肿瘤相关血栓、妊娠相关血栓、基础研究进展及真实临床病例等。通过对上述内容进行结构化组织，知识库旨在为医学学习者提供循证化、情境化的知识支持，并为教师开展教案设计、病例讨论和教学资源准备提供辅助。

1.3　语料来源与知识库架构

本研究按照“循证证据—专家经验—脱敏病例”三层架构组织 VTE 知识库语料。循证证据层主要纳入国内外权威指南、专家共识、近 3 年 JCR Q1/Q2 区高质量 Meta 分析、随机对照研究及中文核心期刊文献，用于构建知识库的循证底座。专家经验层主要纳入本中心血管外科多年积累的经典授课 PPT、典型会诊情境、临床诊治误区总结及教学资料，用于补充指南和文献中较少呈现的临床隐性知识。脱敏病例层纳入具有教学价值的真实临床病例，经规范化脱敏和结构化整理后，以独立文件夹形式管理，用于支持情境化教学、病例讨论和虚拟病例生成。所有文档导入 ima 平台后，由平台完成知识库索引、检索与生成相关处理。由于本研究属



注：本图展示了本研究“知识库构建—RAG 处理与治理—双路径应用评价”的总体流程。研究以循证证据、专家经验和脱敏病例为三层语料来源，经规范化入库、RAG 增强生成和可靠性治理后，分别面向学习者和教师开展教学应用，并评价其在临床决策支持、知识问答、学习资源推荐、教案设计、病例讨论和辅助命题等场景中的应用效果。CDSS，即临床决策支持系统（clinical decision support system）。

图 1　VTE 智能知识库构建及双路径应用评价技术路线

于基于商业化平台用户端开展的应用研究，研究团队无法获取其底层嵌入模型类型、模型版本、参数配置及检索排序策略等核心工程参数。因此，本文方法学描述主要聚焦于语料来源、资料组织、标注管理、质量控制和教学应用流程等可复现环节，不对平台未公开的底层技术实现作进一步推断。

1.4　入库规范与资源管理

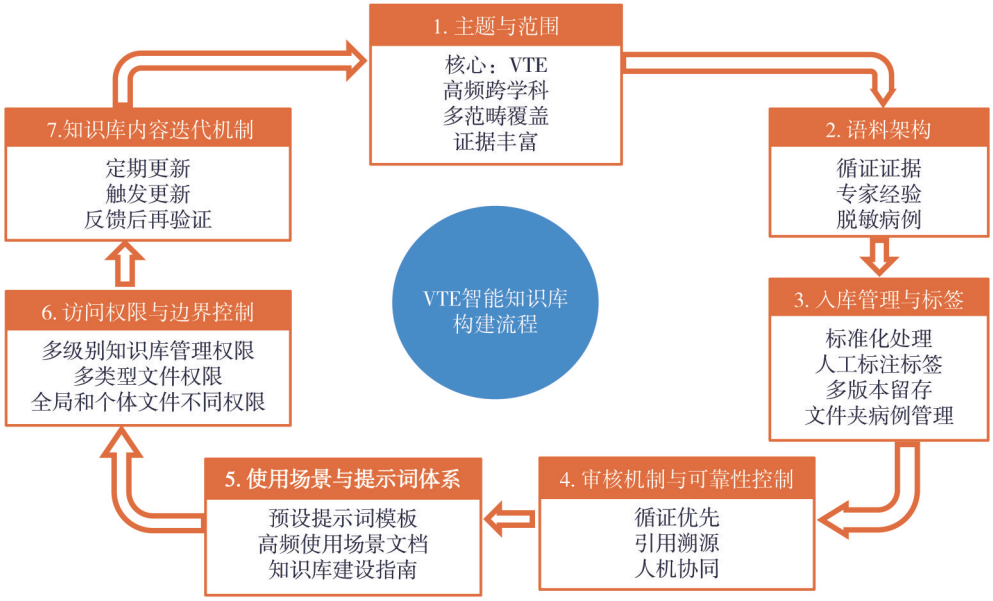
为提高知识库管理规范性和资料检索准确性，所有入库材料在导入前均进行标准化处理，包括文献去重、文件命名统一、来源核对、证据类型标识和版本登记。对于存在更新版本的指南、专家共识或教学资料，以最新版作为主要调用文件，同时保留历史版本信息以便追溯。资料标注以人工完成为主。研究团队在资料审核后统一添加核心主题词、资料类型、年份或版本信息及主题方向等标签。例如，“VTE 预防”“VTE 诊断”“抗凝治疗”“肺栓塞”“深静脉血栓形成”“指南/共识”“Meta 分析”“随机对照研究”“病例”“2024 版”等。对于脱敏病例、教学 PPT、流程图等资料，则根据其内容特点采用标签和文件夹相结合的方式管理，以提高资料调用效率。为增强知识表达的直观性和教学适配性，本研究还将 Caprini 评分表、诊疗流程图、影像示意图及部分关键教学资料以图片、PDF 等形式纳入知识库。此类多模态资料主要用于补充文本资料不易清晰呈现的结构化信息，辅助学习者和教师在知识查询、流程学习及病例讨论中快速定位关键图表、评分工具和诊疗框架。知识库构建流程见图 2。

1.5　质量控制与迭代更新

针对生成式人工智能在医学场景中可能存在的“幻觉”、表述偏差和引用不充分等问题，本研究构建了“循证优先、引用溯源、人机协同”的质量控制策略。所谓“循证优先”，主要通过高等级证据优先入库、三层语料分层管理、统一命名与版本控制、人工标签标注、提示词模板约束及人工复核等方式实现，即在知识库内容组织和使用规范层面引导模型优先调用本地高质量资料，而非指研究团队能够直接设定或验证平台底层检索排序机制。

在知识库运行过程中，研究团队持续收集高频提问、争议性结论和回答不理想的问题，对相关回答的证据来源、准确性、时效性和完整性进行人工复核。根据问题性质，有针对性地补充高质量文献、更新资料版本、优化标签管理、修订帮助文档或调整提示词模板，并使用原问题再次检索和提问，由研究团队复核修订后的输出效果，判断是否改善，从而形成“反馈—修正—再验证”的闭环。

在权限管理方面，VTE 主知识库设定为“公开发布、仅查看、不可导出”，以兼顾知识共享与核心内容保护。提示词方案等工具性文档根据实际需要设置为“可查看、可导出”，以便教学方法传播。人员管理采用“管理员—协同作者—普通用户”三级权限体系，管理员负责整体维护，协同作者参与内容共建，普通用户具有浏览与提问权限。知识库更



注：VTE 智能知识库的构建流程以主题界定和三层语料架构为基础，经规范化入库、RAG 应用处理、质量控制和权限管理后形成可应用知识库，并通过定期更新与触发更新实现持续迭代。

图 2　VTE 知识库的构建流程

新采用“定期更新+触发更新”双轨机制。定期更新是指按固定周期检索新发布指南、共识及高等级证据，经人工筛选和审核后补充入库；触发更新是指当出现重大指南变更、教学中集中出现的新问题或临床实践中的典型新病例时，启动即时更新流程。

2　教学应用

2.1　应用对象与教学场景

为覆盖真实的医学学习和教学工作流，本研究设置学习者路径和教师路径2类应用场景。学习者路径主要面向医学生及非专科医护人员，模拟其在日常学习和临床认知支持中的使用需求，重点考察知识库在临床决策支持、知识问答和学习资源推荐中的应用价值。教师路径主要面向血管外科教师，模拟其在教学准备阶段的工作需求，重点考察知识库在教案设计、疑难病例讨论和辅助命题中的支持能力。2条路径对应不同应用目标。学习者路径更侧重知识获取、复杂概念理解和临床认知支持；教师路径更侧重教学设计、病例组织和教学资源生成。通过学习者和教师双路径应用，本研究旨在从“学”和“教”2个维度评价VTE智能知识库在医学教育场景中的适用性。

2.2　学习者路径应用

学习者路径采用横断面调查设计。研究对象为医学生及非专科医护人员。研究团队围绕VTE医学教育中的常见需求，设置“临床决策支持”“知识问答”和“学习资源推荐”3个标准化应用场景。所有场景均基于入门级ima平台的使用状态生成标准化回答，受访者阅读回答后，从可用性、有效性、循证可信度和持续使用意向等维度进行评价。“临床决策支持”场景主要模拟学习者面对VTE风险评估、诊断路径或抗凝治疗选择时，对知识库进行提问并获得循证化回答的过程；“知识问答”场景主要模拟学习者围绕VTE基础概念、诊疗原则和特殊人群管理等内容进行即时查询；“学习资源推荐”场景主要模拟学习者在希望围绕某一主题开展系统性文献学习时，由知识库推荐指南、共识或高质量研究资料。该路径主要用于评价知识库在标准化学习任务中的实用性和学习支持效果。

2.3　教师路径应用

教师路径采用“虚拟模拟”单盲随机对照设计。研究对象为血管外科教师。研究团队选取“教案设计”“疑难病例讨论”和“辅助命题”3类高频教学准备任务，分别代表不同复杂程度和生成要求的教学场景。针对同一教学任务和同一提示词，在相同时间点、相同软件版本和相同大模型类型的前提下，分别由3种不同使用深度的AI客户端生成回答方案。

3种使用深度分别为入门使用、中度使用和深度使用。入门使用状态来源于研究团队中血管外科规培学生初次下载ima后，仅加入本研究发布知识库的使用状态；生成问卷情境文本前，其个人知识库为5.55 MB，共享知识库为0 B，笔记为15.98 KB。中度使用状态来源于研究团队中血管外科护士连续使用ima约5个月后的使用状态；生成问卷情境文本前，其个人知识库为16.1 MB，共享知识库为322.42 MB，笔记为15.01 KB。深度使用状态来源于研究团队中血管外科教师日常使用ima约9个月后的使用状态；生成问卷情境文本前，其个人知识库为476.3 MB，共享知识库为8.86 GB，笔记为21.72 MB。

针对每项任务，3种使用深度分别生成1份回答方案，共形成A、B、C 3份匿名方案。研究团队将3份方案随机打乱顺序后呈现给受评教师，以降低评分偏倚。教师在不知晓方案来源的情况下，从专业准确性、教学适用性和总体可用性3个维度进行评分，从而评价不同使用深度下知识库对教学准备任务的支持效果。

2.4　知识库矩阵的教学拓展应用

在VTE单体知识库应用基础上，研究团队进一步探索构建“5+Cluster”知识库矩阵，用于支持更复杂的医学教育任务。该矩阵并非多个知识库的简单叠加，而是根据不同教学和科研需求，将专科知识、教学资源、科研资料、工具方法和动态信息进行分工组织，使其能够在复杂任务中协同调用。

“5”是指5类功能定位不同的基础知识库。第1类为教学核心库，以“北京大学第一医院介入血管外科教学知识库”为代表，整合本中心多年积累的教学PPT、精品教案、教学大纲、详细教学计划、关键知识点及教学文献。第2类为科研支撑库，以“疑难血液透析通路MDT课题组”知识库为代表，集中存储课题申报书、原始文献、试验方案和研究进展。第3类为工具与方法库，以“AI提示词库和AI工作日志”为代表，整理医学教学、科研和临床工作中常用的提示词模板及工作流。第4类为专业动态实录库，以“教学会议实录”知识库为代表，将学术会议、教学会议的录音、图片和纪要转化为可检索资料。第5类为专科知识调用库，用于承载不同疾病或亚专业方向的高质量知识资源。

“Cluster”是指围绕不同亚专业主题持续扩展的专科知识库集群，包括VTE、外周血管病、肿瘤介入、血液透析通路及血管外科护理等方向。在实际应用中，单一专科问题可直接在对应知识库中提问；对于教案设计、病例教学整合、试题命制等复合型任务，则可通过“@知识库”方式联合调用多个相关知识库，也可在平台首页发起问题，由平台在用户已发布或已加入的知识库范围内综合组织回答。必要时还可开启互联网检索功能，以补充知识库外部信息。该矩阵化应用方式为解决单体专科知识库在创造性教学任务中的局限提供了可能的路径。

3　量化评价

3.1　评价指标

本研究分别设置了学习者路径和教师路径的评价指标。学习者路径采用Likert 5级量表，从可用性、有效性、循证可信度、理解辅助和持续使用意向等维度对知识库的回答进行评价。评分越高，表示受访者对该维度的认可程度越高。教师路径同样采用Likert 5级量表，从专业准确性、教学适用性和总体可用性3个维度评价匿名回答方案。专业准确性主要反映回答内容是否符合VTE相关的专业知识和临床规范；教学适用性主要反映回答是否适合用于教案设计、病例讨论或命题准备；总体可用性主要反映教师对该方案在实际教学准备中使用价值的综合判断。

3.2　数据处理与统计学方法

所有数据清洗、统计分析和图形绘制均在R语言环境（Version 4.5.0）中完成。首先对问卷数据进行完整性检查和逻辑核查，剔除关键信息缺失或存在明显逻辑错误的记录；对Likert量表条目统一按1~5分编码，并根据研究对象和任务类型完成分组整理。

描述性统计中，计数资料采用频数和构成比表示；Likert评分及其他不符合正态分布的计量资料采用中位数和四分位间距表示。学习者路径中，不同用户特征之间评分差异采用Kruskal-Wallis H检验。教师路径中，由于同一教师需对3种不同使用深度生成的方案进行配对评分，故先采用Friedman秩和检验比较总体差异；若差异有统计学意义，再进行两两比较的Wilcoxon符号秩检验，并对多重比较的P值进行校正。图形展示采用发散式堆叠条形图描述各维度评分分布，并计算Top-box，即4~5分比例，以反映高满意度水平。所有检验均为双侧检验，检验水准α=0.05。

3.3　知识库构建及运行结果

本研究基于“循证证据—专家经验—真实病例”三层架构，构建并上线了“VTE指南和高级别证据”知识库。截至2026年1月18日，该知识库已收录核心文档688份，涵盖VTE风险评估、诊断、治疗、预防、特殊人群管理和研究进展等内容；其中脱敏病例库已沉淀151例结构化真实病例。

平台用户侧可见运行数据显示，知识号累计访问量达5 424次，产生深度问答915次，提示该知识库具有一定的用户活跃度和真实应用需求。基于VTE单体知识库建设经验，研究团队进一步形成“5+Cluster”知识库矩阵体系，使专科知识库建设由单一主题拓展至教学、科研、工具方法、专业动态和亚专业知识集群等多个方向。该体系通过不同知识库之间的功能分工和协同调用，为后续开展复杂医学教育任务提供了结构化知识底座。

3.4　学习者路径评价结果

本研究共纳入有效问卷150份，其中学习者路径问卷131份。学习者路径量化评价结果显示，VTE智能知识库在标准化学习任务中整体表现较好。各评价维度评分中位数均为4.00分，四分位间距多集中于4.00~5.00分，提示学习者对知识库在临床决策支持、知识问答和学习资源推荐等场景中的应用效果总体是认可的。从高满意度比例看，“持续使用价值”和“理解辅助”2个维度的Top-box比例分别为83.97%（110/131）和82.44%（108/131），提示知识库能够在一定程度上降低学习者面对VTE复杂知识体系时的检索和理解负荷，具有较好的持续使用潜力。“循证依据可信度”维度的评分相对离散，Top-box比例为74.05%（97/131），低于其他主要维度，提示即使在检索增强生成技术支持下，学习者对AI生成内容的证据来源和医学可信度仍保持相对审慎态度。学习者路径各评价维度评分分布见图3。

3.5　教师路径评价结果

教师路径共纳入19名血管外科教师。单盲随机对照评价结果显示，VTE智能知识库在教师教学准备任务中的应用效果具有一定任务依赖性。在“教案设计”和“疑难病例讨论”等相对收敛性任务中，深度使用状态下生成方案的评分均值略高于入门使用状态下生成的方案，均分差值为+0.05~+0.26，虽未达到统计学意义，但提示知识库在此类教学准备任务中具有一定辅助价值。

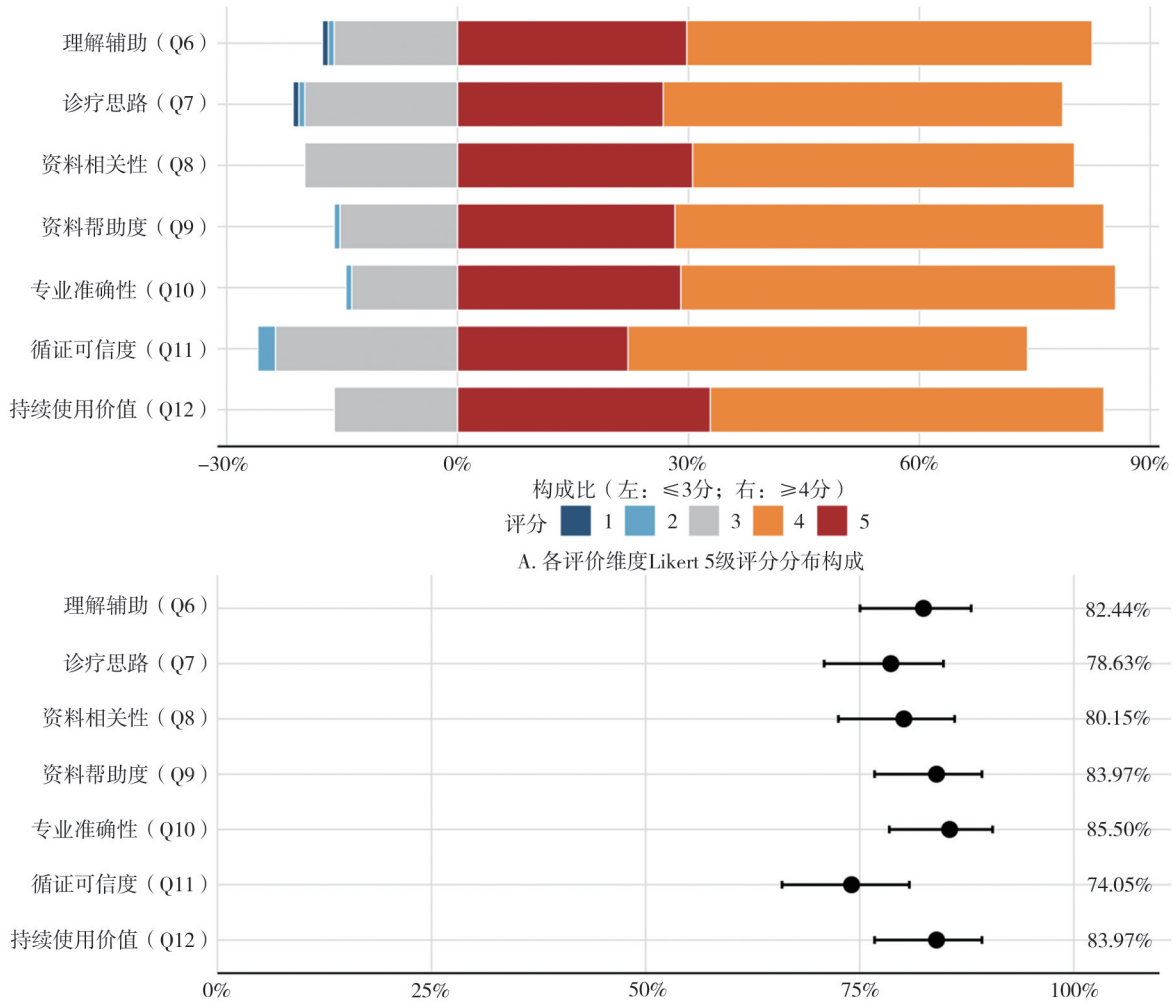


图3　学习者路径评价指标评分分布（n=131）

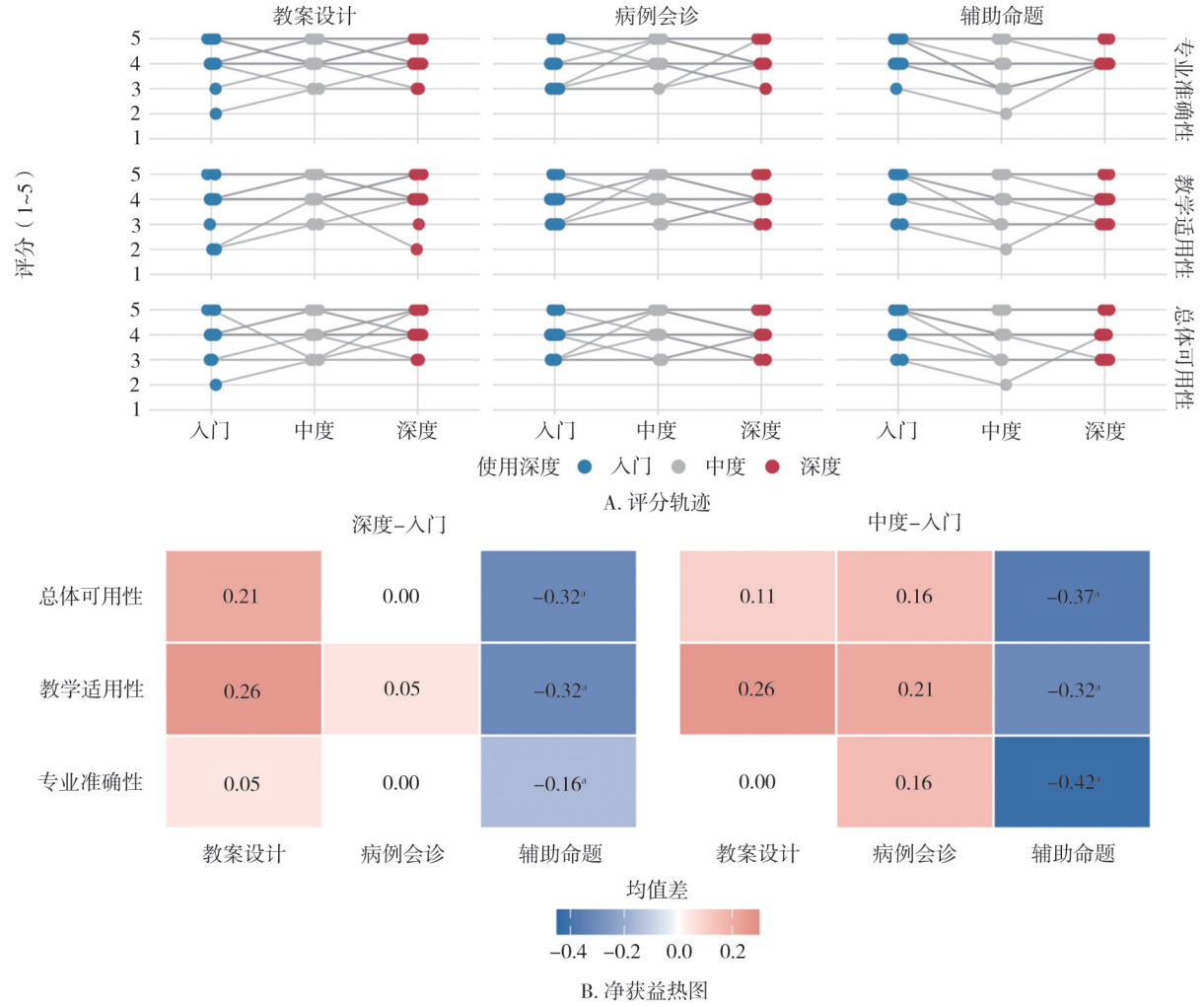
在“辅助命题”任务中，深度使用状态生成方案的表现反而低于入门使用状态，且Friedman检验显示差异具有统计学意义（P<0.05）。其中，深度使用方案在专业准确性和教学适用性维度均呈负向差值，差值为−0.16～−0.32。该结果提示，单体VTE临床知识库虽然能够提供较充分的专业知识支持，但在辅助命题等创造性教学任务中仍存在局限。医学命题不仅需要专科知识，还需要教育测量学、命题规范、难度控制和干扰项设计等教学法相关语料支持。教师路径评价结果见图4。

4　讨论

4.1　RAG+LLM驱动的知识库在降低VTE教学认知负荷中的作用

VTE医学教育难点是知识体系庞杂，循证证据更新频繁，所以其临床风险评估和抗凝决策具有很高的认知负荷。而从本研究的结果来看，RAG技术驱动的LLM构建专业知识库后，“临床决策支持”“知识问答”类标准化任务Top-box满意度都在82%以上。也就是说，将大语言模型的生成过程通过RAG技术“锚定”于经过筛选的质量较高的循证底座上，可以发挥较好的认知辅助工具作用。即通过降低学习者对零散、碎片化知识的记忆与检索压力，使学习者将更多的认知资源聚焦到高层次的临床推理与决策分析当中[3,5]。这也符合已有研究结论：检索增强技术可以显著降低通用大模型在医学场景中出现幻觉的概率，进而提高工具的可用性和透明度[9-10]。

本研究中学习者对知识库“循证可信度”维度的评分相对低于其他维度，这一结果并不一定说明知识库内容质量不足，更可能反映出医学学习者对AI工具在专业场景中的应用仍持较为审慎的态度。



注：A图为评分轨迹，显示同一教师在教案设计、病例会诊及辅助命题3类任务中，随知识库使用深度增加的评分变化趋势；B图为净获益热图，显示中度或深度使用相对于入门使用的均分差值。红色表示评分提升，蓝色表示评分下降；a，$P<0.05$，Friedman检验。

图4　知识库使用深度对不同教学任务评分的影响（n=19）

尤其是在VTE诊断、抗凝治疗及特殊人群管理等高风险问题上，学习者往往更关注回答能否明确对应指南依据、证据来源及适用边界。另一方面，VTE领域本身指南更新较快，不同人群、不同情境下的推荐意见也可能存在差异，这也会提高学习者对“循证可信度”的感知阈值。虽然本研究已通过高等级证据优先入库、版本控制、引用溯源及人机协同审核等方式提升内容可靠性，但平台最终呈现给用户的回答仍以整合生成结果为主，在证据链展示的直观性方面还有进一步优化的空间。后续可通过强化证据分级标识、增加指南版本提示等方式，进一步提升学习者对知识库可信度的感知。

本研究在评估阶段未单独设置基于标准问题集的客观准确率或事实一致性测试，而主要通过学习者与教师对不同任务的评价结果综合判断知识库的应用效果。但这并不意味着评价过程完全脱离对回答准确性的考查。一方面，学习者在实际使用和评分过程中，通常会结合已有知识对AI回答进行基础的事实核查，并将其反映在专业准确性、循证可信度等评分维度中；另一方面，本研究还纳入了19名血管外科专业教师参与评价，其对回答内容专业性、准确性及教学适配性的判断，也在一定程度上构成了对知识准确性的间接检验。尽管如此，这种评价方式仍不同于基于标准答案的客观准确性测试，因此后续研究仍有必要进一步引入标准问题集、事实一致性评估等客观指标，以更全面地刻画知识库加载后模型回答的基础性能。

4.2　单体知识库在创造性教学任务中的局限性分析

教师端单盲随机对照试验结果表明，虽然当前

技术路线适用于教案生成等收敛性任务，但对于更为创造性的工作任务——辅助命题，依然达不到预期目标。评分结果显示，采用单一 VTE 单体知识库的评分结果（$P<0.05$）显著低于入门模型（总体平均得分 69 分），说明目前单体 VTE 知识库的数据集存在一定的局限性。医学命题既需要医学专业知识的支持，也需要基于测量学的知识（难度控制、干扰项设计等）。

本文所搭建的 VTE 知识库虽然包含详细的指南资料和病历资料，但并未对教学法方面的语料加以关注，RAG 技术通过强制模型只能看到临床文档的方式，会使得这种过于严格的语义约束降低通用大模型的逻辑迁移能力及发散思维能力。由此可见，AI 在一些涉及价值判断及情境复原等教育环节仍存在一定的局限性，并不能直接替代教师的创造性工作[11]。与此同时，本研究对教师路径的评价主要聚焦于教学准备阶段的支持效果，因此更适合反映 AI 生成内容在专业性、教学适配性和可用性方面的前期表现，而不宜直接外推为真实教学实施后的最终效果。对于命题难度、区分度等教育测量学指标，仍需结合真实受试者的作答结果进一步验证。后续研究可在真实教学实践中，结合知识点覆盖率、题型结构合理性、难度分布、区分度及评分一致性等客观指标，对教师路径的应用效果作出更为全面的评价。

4.3　“知识库矩阵”体系的构建与应用价值

鉴于单体知识库在应用于创造性任务时存在瓶颈问题，本研究进一步提出并构建了“5+Cluster”知识库矩阵体系。其核心并不在于单纯增加知识库数量，而在于通过不同类型知识库的分工互补，将专科知识、教学规范、科研资料、工具方法和动态信息组织为可协同调用的知识底座。该体系针对不同教学维度进行定向补充：其一，针对命题与教案设计等教学场景需求，构建专科教学知识库，纳入教学大纲、经典试题与教育学方法论，补充教学逻辑语料；其二，针对不同疾病的专业知识需求，构建亚专业知识库集群，覆盖血管通路、肿瘤介入等亚专业领域；其三，针对科研转化需求，构建课题组知识库，收集课题组前期资料，协同完成科研设计、执行和产出。知识库矩阵建立后，可在同一任务中联合调用不同知识库的内容资源，实现专科知识、教学规范及外部信息的协同支持。与单体知识库相比，这种方式更适合命题、教案设计和病例教学等复合型任务，因为这类任务不仅需要专业知识，还需要教学方法、测量学规范及情境组织能力的共同参与。例如在 VTE 试题命制场景中，教师既需要 VTE 专业知识支持，也需要教学知识库中的命题规范作为补充，矩阵体系因此比单一专科知识库更能满足实际教学任务需求。因此，从本研究结果来看，知识库的应用效果并不只取决于是否接入 RAG 技术，还与知识底座的组织方式和协同调用能力密切相关。对于命题、教案设计和病例教学等复合型任务，分工互补的知识库矩阵较单体知识库更有可能提升 AI 在真实教学场景中的应用效果，也提示教师在 AI 时代的角色将更多转向知识系统的架构者与协同者[4,12]。

4.4　知识库的可靠性治理与数据安全

尽管矩阵体系拓展了应用场景，但可靠性治理仍是其应用的基础。VTE 领域存在证据冲突（如不同版本评分表差异[13]）及真实应用场景差异，要求 AI 输出兼顾准确性与适宜性。因此，需形成“证据分级—版本控制—人机协同审核”三位一体治理闭环，以保障 AI 输出的规范化与科学化[14]。未来如果知识库进入多中心协作阶段，需要解决的问题不仅是如何将知识库输出内容输入各中心，也需要考虑如何将不同医院的输入信息进行统一的结构化，在此过程中数据的安全与隐私问题同样重要。通用数据模型+联邦学习的数据治理框架可能会成为保障数据安全与隐私以及跨机构合作的技术手段[15]。从应用效果角度看，知识库的可靠性治理不仅关系到输出内容的规范性和科学性，也直接影响学习者和教师对其持续使用的意愿及推广的可行性。只有当知识来源可追溯、版本可核查、内容可复核时，知识库在医学教育中的应用效果才具有较为稳定的基础。

4.5　局限性与展望

本研究仍存在一定局限性。首先，本研究为单中心初步应用研究，样本数量相对较少，尤其是教师样本有限，可能在一定程度上影响结果的稳健性和外推性。其次，本研究的应用效果评价仍以教师和学习者主观满意度及自评获益为主，客观学习结局指标相对不足，因此研究结论的可信度仍有待进一步提高。再次，本研究主要完成了横断面评价和短期对照试验，尚未开展长期随访评价，因此目前仍缺乏知识库在医学教育中持续使用效果、知识保持情况及真实教学转化效果的长期证据。后续研究拟在现有基础上进一步扩大样本量，逐步开展多中心验证，并增加基于病例推理的知识型学习结局

指标、模型基础性能指标及教师路径相关客观任务指标，如学习前后测试成绩、病例分析能力、知识检索效率、基于标准问题集的回答准确率、事实一致性、教学准备耗时，以及命题知识点覆盖率、题型结构合理性、难度分布、区分度和一致性等，从知识掌握、临床应用和教学支持等多个维度综合评价知识库的实际效果。同时，拟将知识库矩阵进一步融入教学查房、MDT 讨论等真实临床教学实践活动中，推动其形成常态化应用机制，并结合长期使用情况与阶段性随访，进一步检验该方法的真实教育应用效果。

5　结语

本研究结果表明，基于 RAG+LLM 构建的 VTE 专业知识库在医学教育中具有一定应用价值。对于临床决策支持、知识问答和学习资源推荐等标准化任务，其能够在一定程度上降低学习者面对高认知负荷内容时的检索与整合压力，并提供较好的认知支持；对于教案生成等教学准备任务，也显示出一定辅助作用。但与此同时，单体临床知识库在辅助命题等更具创造性的教学任务中仍存在明显能力边界。进一步研究提示，知识库的应用效果不仅取决于模型本身，还与知识底座的质量、组织方式和协同调用能力密切相关。因此，医学教育知识库的建设不应停留在单一专题语料库层面，而应逐步走向覆盖“临床—教学—科研”多维场景的知识库矩阵，并在可靠性治理、长期随访和真实教学实践中持续验证其应用效果与边界。

AI 声明：图 1 来源于 AI（Gemini 3 pro）的辅助设计。在本次研究过程中，R 语言代码由 AI（ChatGPT 5.2 Extended Think）进行了辅助修改。所有用到的图片均经过了人工校准与重绘，并与原文档的原始数据进行了比对，确认无误。

利益冲突　所有作者声明不存在利益冲突

作者贡献声明　佘康：研究设计、基金获取、统计分析、论文撰写与修改；郝雨迪、王亚青、李俊梅、马颖林、母光妍、李珂、张晓琳：资料整理、项目实施；齐建光、齐心：论文指导、审核与统稿

参考文献

[1] Wendelboe A, Weitz JI. Global health burden of venous thromboembolism [J]. Arterioscler Thromb Vasc Biol, 2024, 44(5): 1007-1011. DOI: 10.1161/atvbaha.124.320151.

[2] Henke PK, Kahn SR, Pannucci CJ, et al. Call to action to prevent venous thromboembolism in hospitalized patients: a policy statement from the American heart association [J]. Circulation, 2020, 141(24): e914-e931. DOI: 10.1161/cir.0000000000000769.

[3] Zhang S, Liu L, Ding S, et al. Mediator effects of cognitive load on the relationship between task complexity and guideline adherence among clinical nurses: a cross-sectional survey of nurses in China [J]. J Res Nurs, 2025, 30(5/6): 479-490. DOI: 10.1177/17449871251329197.

[4] 田蕾，孙宝志. 第四代医学教育考核评价：精准教育评价[J]. 中华医学教育探索杂志，2025，24(11)：1441-1446. DOI: 10.3760/cma.j.cn116021-20250401-02127.

[5] 徐赫，焦洋. 人工智能在临床推理教学新模式中的研究及展望[J]. 中华医学教育探索杂志，2025，24(10)：1372-1378. DOI: 10.3760/cma.j.cn116021-20250310-02066.

[6] Karajizadeh M, Hassanipour S, Sharifian R, et al. The effect of information technology intervention on using appropriate VTE prophylaxis in non-surgical patients: a systematic review and meta-analysis [J]. Digit Health, 2022, 8: 205520762211188. DOI: 10.1177/20552076221118828.

[7] Rao A, Pang M, Kim J, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow: development and usability study [J]. J Med Internet Res, 2023, 25: e48659. DOI: 10.2196/48659.

[8] Janumpally R, Nanua S, Ngo A, et al. Generative artificial intelligence in graduate medical education [J]. Front Med, 2025, 11: 1525604. DOI: 10.3389/fmed.2024.1525604.

[9] Neha F, Bhati D, Shukla DK. Retrieval-augmented generation (RAG) in healthcare: a comprehensive review [J]. AI, 2025, 6(9): 226. DOI: 10.3390/ai6090226.

[10] Liu S, McCoy AB, Wright A. Improving large language model applications in biomedicine with retrieval-augmented generation: a systematic review, meta-analysis, and clinical development guidelines [J]. J Am Med Inform Assoc, 2025, 32(4): 605-615. DOI: 10.1093/jamia/ocaf008.

[11] 唐江桥，戴珑珑. 人工智能辅助教学的边界：高校教师的专业判断与价值坚守[J]. 黑龙江高教研究，2026，44(1)：13-18. DOI: 10.19903/j.cnki.CN23-1074/G.2026.01.003.

[12] 关旭，刘骞，王锡山. 人工智能在肿瘤外科医学专科教育中的现状与应用研究[J]. 中华医学教育探索杂志，2024，23(10)：1337-1342. DOI: 10.3760/cma.j.cn116021-20240511-01869.

[13] Gharaibeh L, Albsoul-Younes A, Younes N. Evaluation of VTE prophylaxis in an educational hospital: comparison between the institutional guideline (caprini 2006) and the ACCP guideline (ninth edition) [J]. Clin Appl Thromb Hemost, 2016, 22(7): 627-632. DOI: 10.1177/1076029615575344.

[14] 徐锐，郑兰斌，张燕蕾，等. 生成式 AI 在医学教育中的应用：影响机制与风险应对[J]. 中国卫生事业管理，2025，42(12)：1398-1401，1411. DOI: 10.3969/j.issn.1004-4663.2025.12.014.

[15] von Gerich H, Chomutare T, Kytö V, et al. A rapid review on the application of common data models in healthcare: recommendations for data governance and federated learning in artificial intelligence development [J]. Digit Health, 2025, 11: 20552076251395536. DOI: 10.1177/20552076251395536.

（收稿日期：2026-01-28）
（本文编辑：邬小梅）