Advances in Education 教育进展, 2026, 16(11), 1327-1338
Published Online November 2026 in Hans. https://www.hanspub.org/journal/ae
https://doi.org/10.12677/ae.2026.16112507

Hans汉斯

人工智能在提升医学研究生教育质量中的批判性综述

毛宇昂[1]，于　夫[2]，杨再学[3]，黄代政[3]，黎柏伶[4]，黄　超[4*]

[1]广西医科大学研究生院，广西　南宁
[2]广西医科大学校办，广西　南宁
[3]广西医科大学生命科学与人工智能学院，广西　南宁
[4]广西医科大学信息中心，广西　南宁

收稿日期：2026年10月10日；录用日期：2026年11月10日；发布日期：2026年11月18日

摘　要

人工智能(Artificial Intelligence, AI)正深刻改变医学研究生教育的培养模式。文章采用文献综述法，从个性化与自适应学习、临床思维与技能训练、考核评估体系以及医学科研创新四个维度，系统梳理AI在医学研究生教育中的主要应用进展。在此基础上，重点分析AI应用过程中面临的数据安全、算法偏见、评估异化、学术伦理及教育公平等现实挑战。研究表明，AI能够有效提升医学研究生培养的精准化、智能化和个体化水平，但其教育价值的充分实现仍受到技术局限、师资能力、资源配置和制度环境等多重因素制约。未来医学研究生教育的发展应坚持人机协同理念，在推进技术创新的同时，加强伦理治理、师资建设和教育体系改革，以实现AI技术与医学教育高质量发展的深度融合。此外，研究设立独立章节集中论述AI应用的四类批判性议题(数据与算法维度、技术保真维度、评估维度与伦理维度)，并与四大应用场景进行交叉分析；结合我国临床医学硕士专业学位研究生教育与住院医师规范化培训并轨(“四证合一”)的制度背景，剖析评估异化与师资转型的特殊表现与治理难点；最后提出分阶段实施路径，包括临床导师“AI + 医学教育”能力发展工作坊课程框架与基于联邦学习的区域性医学教育数据协作体构想。

关键词

AI，医学研究生教育，批判性综述，住院医师规范化培训，教育治理

A Critical Review of AI Technology in Enhancing the Quality of Medical Postgraduate Education

Yu’ang Mao[1], Fu Yu[2], Zaixue Yang[3], Daizheng Huang[3], Bailing Li[4], Chao Huang[4*]

[*]通讯作者。

文章引用: 毛宇昂, 于夫, 杨再学, 黄代政, 黎柏伶, 黄超. 人工智能在提升医学研究生教育质量中的批判性综述[J]. 教育进展, 2026, 16(11): 1327-1338. DOI: 10.12677/ae.2026.16112507

[1]Graduate School, Guangxi Medical University, Nanning Guangxi
[2]Administrative Office, Guangxi Medical University, Nanning Guangxi
[3]School of Life Sciences and Medical Engineering, Guangxi Medical University, Nanning Guangxi
[4]Information Center, Guangxi Medical University, Nanning Guangxi

Received: October 10, 2026; accepted: November 10, 2026; published: November 18, 2026

Abstract

Artificial intelligence (AI) is profoundly reshaping the training of medical postgraduates. Adopting a literature review approach, this article systematically examines the major application advances of AI in medical postgraduate education across four dimensions: personalized and adaptive learning, clinical reasoning and skills training, assessment and evaluation systems, and medical research innovation. On this basis, it focuses on analyzing the practical challenges encountered in AI applications, including data security, algorithmic bias, assessment alienation, academic ethics, and educational equity. The study indicates that AI can effectively enhance the precision, intelligence, and individualization of medical postgraduate training; however, the full realization of its educational value remains constrained by multiple factors such as technological limitations, faculty capacity, resource allocation, and institutional environments. The future development of medical postgraduate education should adhere to the concept of human-machine collaboration, strengthening ethical governance, faculty development, and educational system reform while advancing technological innovation, so as to achieve deep integration of AI technology with high-quality development in medical education. Furthermore, the study establishes a dedicated section to concentrate on four categories of critical issues in AI applications (data and algorithmic dimension, technological fidelity dimension, assessment dimension, and ethical dimension), conducting a cross-analysis with the four major application scenarios; in light of China's institutional context of integrating the clinical medicine professional master's degree postgraduate education with standardized residency training ("four certificates in one"), it analyzes the distinctive manifestations and governance difficulties of assessment alienation and faculty transformation; and finally, it proposes a phased implementation pathway, including a workshop curriculum framework for developing clinical supervisors' "AI + Medical Education" competencies and a conception of a regional medical education data consortium based on federated learning.

Keywords

Artificial Intelligence, Medical Postgraduate Education, Critical Review, Standardized Residency Training, Educational Governance

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

医学研究生教育是培养高层次医学精英人才的主要途径，其质量直接关系国家医疗卫生事业的未来。从弗莱克斯纳报告奠定的“医学科学 + 临床实践”两阶段模式，到以师傅带徒弟为核心的住院医师规范化培训，其本质始终是一种基于经验主义与精英传承的模式，高度依赖导师个人水平、有限的临床病例(病源)以及传统的主观评估手段，在标准化、同质化、个性化和效率方面的瓶颈日益显现，已难以应对现

代医学知识半衰期急剧缩短、临床需求日趋复杂和健康中国对卓越人才的迫切需求[1] [2]。

人工智能(AI)的崛起，特别是以机器学习(ML)、深度学习(DL)、自然语言处理(NLP)和生成式 AI (Generative AI)为代表的核心技术集群，为破解上述难题提供了历史性机遇[3] [4]。AI 不再是简单的信息化工具，而是作为一种类智能的认知代理，深度介入学习、训练、评估和科研的全链条。然而，当前学术界和产业界对 AI 应用的讨论普遍存在技术赋能的乐观主义倾向，多集中于罗列其功能与潜力，对其在真实、复杂、高风险的医学教育情境中的局限性、潜在风险、实施挑战以及教育伦理的批判性分析尚显不足。本文的核心观点是：AI 技术应用是一把精巧的双刃剑，必须从技术决定论的迷思转向审慎融合的现实主义，技术乌托邦式的畅想应让位于对其社会技术属性的冷静审视[5]。

在系统检索相关文献的基础上，本文结合近年的代表性研究成果，对 AI 在医学研究生教育中的四大应用领域进行批判性综合，不仅回答能做什么，更深入分析应做什么、不能做什么以及潜在的危害是什么，并构建 AI 赋能医学研究生教育的应用与挑战分析框架(见图 1)。全文结构如下：第 2 章梳理 AI 在医学研究生教育中的四大应用场景及其技术实现路径；第 3 章设立独立的批判性分析章节，将散见于各类应用的批判性议题归并为数据与算法、技术保真、评估异化与伦理侵蚀四个维度，并通过交叉分析矩阵与第 2 章的应用场景系统对应；第 4 章结合我国临床医学硕士专业学位研究生教育与住院医师规范化培训并轨的制度背景，剖析评估异化与师资转型的特殊形态与治理难点；第 5 章在识别师资与系统瓶颈的基础上，提出分阶段实施路径、临床导师能力发展工作坊课程框架与区域性医学教育数据协作体构想；第 6 章为结论与展望。

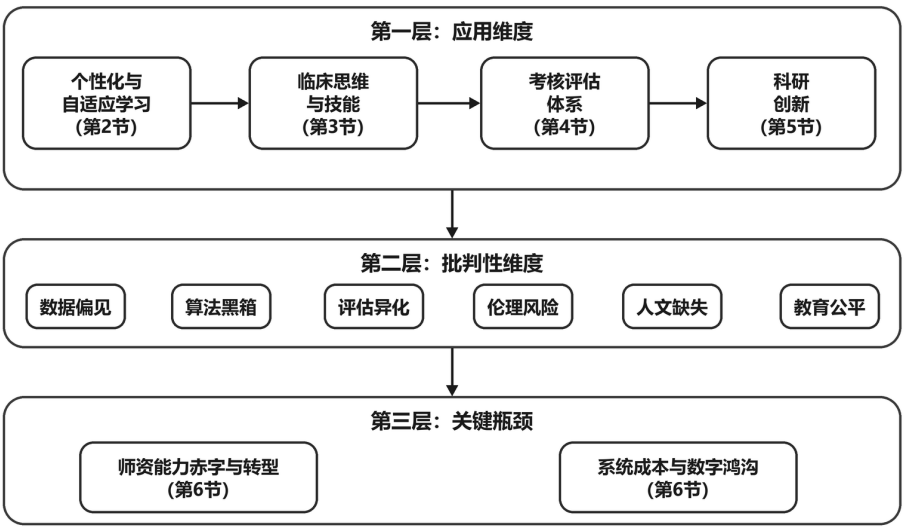


Figure 1. Theoretical framework of AI applications and challenges in medical postgraduate education
图 1. AI 赋能医学研究生教育的应用与挑战理论分析框架图

2. AI 在医学研究生教育中的四大应用场景

在明确分析框架的基础上，本章梳理 AI 在医学研究生教育中的四类核心应用场景——个性化与自适应学习、临床思维与技能训练、考核评估体系及医学科研创新，重点说明其技术实现路径与已报告的教学效果，为第 3 章的批判性分析提供事实基础。

2.1. 个性化与自适应学习：从智能导师到生成式 AI 辅导

传统一刀切的教学大纲和轮转计划，往往忽视研究生在知识背景、学习曲线和专业兴趣上的巨大差

异，AI 的核心优势在于通过数据驱动实现千人千面的精准教育[3] [4]。智能导师系统(ITS)是 AI 教育应用的经典范式，其核心是医学知识图谱：通过深度学习算法，将海量的医学教科书、临床指南、最新文献与病例数据库等结构化与非结构化知识，编织成庞大的概念–关联网络。学习过程中，ITS 通过课程测试、病例报告、操作视频等多模态数据诊断知识短板，据此从图谱中动态规划最优路径，例如为心内科研究生推荐最新 ESC 指南，为尚存欠缺者推送基础病理生理学视频[5]。自适应学习平台(ALP)则更侧重学习内容的动态调整，实时追踪页面停留时间、答题反应时间、重复观看次数等学习行为，并即时调整后续内容的难度与呈现方式[3]。

以 ChatGPT、Gemini 等为代表的大型语言模型(LLMs)正扮演全天候辅导员的角色，研究生可随时就复杂的临床问题开展深度、多轮的探究式提问。LLMs 以启发式、对话式的方式引导思考而非简单给出答案，极大弥补了临床带教老师时间不足的短板[6]-[8]。2022 年以来，GPT-4V、Gemini、Med-PaLM 2 等多模态大模型进入医学教育场景，可同时处理文本、影像、病理切片与生理信号，为“图文对照”的病例教学、影像报告撰写训练与跨模态临床推理训练提供了新的技术可能[9]。一项覆盖 2023~2025 年医学教育文献的文献计量与范围综述显示，该领域已从单一关注 ChatGPT 拓展至多模型并存格局，研究类型由观点性文章转向实证研究与基准测试，但证据仍集中于知识考核通过率与学习者满意度，对临床推理、沟通技能与长期职业能力的影响尚未验证[10]。

从功能定位看，ITS 负责知识诊断与路径规划，ALP 侧重学习过程中的动态内容适配，LLMs 强化即时交互与自主学习支持。三者共同推动医学研究生教育由标准化教学转向个性化、智能化与持续反馈的模式，但其有效性仍高度依赖数据质量、算法性能与教师引导。

2.2. 临床思维与技能训练：从 VR 手术到 AI 驱动的虚拟病人

临床能力的培养是医学研究生教育的核心，传统“在岗学习”模式受病源、风险、伦理等多重限制。而 AI 结合虚拟现实(VR)、增强现实(AR)与混合现实(MR)技术，为临床技能训练提供了高保真、可重复、零风险的环境。在 VR 环境中，学生佩戴头盔进入完全沉浸的虚拟手术室，骨科研究生可反复练习“全膝关节置换术”，AI 算法实时评估其精确度、速度与潜在错误[11] [12]。蒲军等[11]将 68 名外科学专业学位研究生分为 VR 教学组(36 人)与传统教学组(32 人)，在肾切除术教学中比较两组效果：VR 组考试成绩为(92.17±0.38)分，显著高于传统组的(87.94±0.43)分($P < 0.001$)；超过 82%的 VR 组学生对 VR 教学的逼真度、交互性及学习兴趣激发作用持积极评价，且所有调查对象均认为 VR 教学可替代传统手术教学方法。该研究验证了 VR 的教学可行性与短期效果，但仅评价理论知识掌握，样本量有限，未衡量操作技能的迁移效果。

AR 使外科研究生能将患者 CT/MRI 三维重建影像(如肿瘤、血管)精准叠加到模拟人或真实患者相应部位，实现透视指导，提升手术规划与穿刺精准性[13]；MR 进一步允许手术器械等真实物体与虚拟影像实时交互。AI 驱动的高保真模拟器则能模拟复杂的生理反应，例如在气管插管操作失误时即时呈现喉头水肿或氧饱和度下降的生理指标变化，迫使学生作出应对。交互式虚拟病人(VP)借助生成对抗网络(GANs)，基于真实临床数据生成大量具有不同病史、症状、体征和社会心理背景的病例，供研究生开展全流程的交互式诊疗[4] [14]。

2.3. 考核评估体系：从量化评估到形成性画像

传统出科考核、三基考试与基于导师主观印象的评估长期存在主观性强、反馈滞后、维度单一的问题，AI 正推动考核体系向客观、量化、即时、形成性方向发展。在客观结构化临床技能考试(OSCE)或模拟手术考核中，AI 通过计算机视觉(CV)与传感器技术，可精确捕捉传统人工评估无法量化的指标，如腹腔镜与显微手术中的器械运动轨迹(路径经济性)、速度、震颤幅度与频率、完成时间、缝合力度、对虚拟

组织的损伤程度等数十项参数，并与专家模型数据库比对，给出客观、可重复的量化评分[15]。

NLP 技术可自动分析研究生书写的电子病历，评估其完整性、逻辑性、诊断依据是否充分、治疗方案是否合规，甚至能通过情感分析初步判断其“医患沟通”记录的人文关怀水平；对与标准化病人(SP)沟通的录音录像进行分析，则可评估其问诊技巧、信息告知和共情表达能力。AI 评估最大的优势在于即时性：无论在 VP 训练后还是模拟手术后，学生几乎可以立刻收到系统生成的详细反馈报告，这种贯穿全程的“形成性评估”能够动态追踪成长轨迹、及时干预，是传统滞后的终结性考核无法比拟的。

2.4. 医学科研创新：大数据挖掘、科研选题与全流程提效

AI 能帮助研究生从海量临床数据、基因组学、蛋白质组学与影像组学中高效发现相关性、识别生物标志物、构建预后预测模型；智能文献工具可加速系统性综述的完成；自然语言生成(NLG)技术可辅助撰写论文的材料与方法部分或润色英文文稿；AI 驱动的数据分析平台则简化了复杂的生物信息学与统计分析过程[4] [15]。科研环节的深度嵌入在提升效率的同时，也使学术诚信与科研能力的养成面临新的考验，这一问题将在 3.4 节展开分析。

2.5. 四类应用场景的特征比较与证据评估

综合上述四类场景可见，AI 的应用已由单一工具辅助走向全流程嵌入，但证据基础并不均衡：技能训练与知识辅导环节的短期效果报告较多，涉及临床推理迁移、职业素养养成与长期随访的高阶证据明显不足。四类场景的典型技术、教学效果与主要局限见表 1。

Table 1. Typical applications and limitations of AI technologies in medical postgraduate education
表 1. AI 技术在医学研究生教育中的典型应用及局限

AI 技术	代表工具	应用场景	主要教学效果	主要局限	参考文献
ITS	知识图谱、智能导师系统	个性化学习	精准识别知识薄弱点	数据依赖强	[1] [5] [12]
LLMs	ChatGPT、Gemini	学习辅导、科研写作	提升学习效率与信息获取	幻觉与伦理风险	[3] [6]-[8]
VR	虚拟手术平台	外科技能训练	提升理论成绩与兴趣	触觉反馈不足	[11]
AR	AR 导航系统	解剖教学、穿刺训练	增强空间认知	设备成本高	[13]
VP	虚拟病人	临床思维训练	提高诊疗决策能力	难模拟真实情感互动	[4] [14]
AI 评估	CV、NLP	OSCE、病历评价	实现客观量化评价	黑箱效应	[15]-[17]

3. AI 应用的批判性维度：四类议题的系统分析

批判性审视是本文区别于一般性技术综述的核心。四类场景在提升效率的同时，各自衍生出共性风险与场景性张力。为避免论点散落，本章将其归并为数据与算法、技术保真、评估与伦理四类议题集中分析，3.5 节再与第 2 章的场景系统对应。

3.1. 数据与算法维度的三重隐忧

3.1.1. 数据壁垒与隐私红线：高质量长周期数据的供给瓶颈

AI 的燃料是数据，个性化学习的实现高度依赖数据质量。所需的高质量、大规模、长周期教学数据(完整学习画像)与构建病例库的临床数据极度稀缺：各医学院、附属医院、教务与图书馆系统之间数据壁垒森严、标准难以统一，模型只能在小范围、低质量数据上训练[1] [5]；记录学生犯错的学习数据与患者临床数据又高度敏感，如何在隐私安全与可用不可见之间取得平衡，是亟待解决的技术与法规

难题[5] [18]。联邦学习、安全多方计算等隐私计算技术提供了可行路径：Kuo 等[19]在五家医院数据上的研究表明，分布式协同学习在不共享原始记录的前提下，预测性能接近集中式训练并明显优于单中心模型；国内基于隐私计算平台的多中心模型同样证实，参与机构增加可带来稳定的性能增益[20]。但此类技术在我国医学教育场景的落地仍受制于数据标准不统一、机构间算力不均衡与合规责任不清，相关构想见 5.5 节。

3.1.2. 算法偏见与教育公平：统计学黑箱的价值预设与解释鸿沟

算法并非价值中立，AI 算法是一个统计学上的黑箱。若训练 ITS 的“金标准”数据主要来自特定优势院校或导师的经验，其推荐的最优路径可能并不适用于所有学生，甚至固化并放大既有教育不公，形成资源配置的马太效应[18] [21]。可解释 AI (XAI)的缺位使问题更棘手：当 AI 判定某学生掌握不牢却无法清晰解释依据时，这种算法权威难以被师生信服[5] [18]。

3.1.3. 知识茧房与认知窄化：过度适配对创新思维的潜在抑制

知识茧房是个性化推荐系统在教育领域最值得警惕的风险。AI 过度迎合学生的既有兴趣与知识舒适区，不断推送其可能喜欢或亟需弥补的内容，可能导致视野的认知窄化[6] [18]；医学的进步往往源于跨学科的意外发现和对困难问题的主动挑战，过度优化的学习路径可能剥夺学生自主探索、遭遇失败与“慢思考”的机会，培养出高效的解题者而非具有批判性思维与创新能力的学者。

上述担忧并非纯理论推演。Li 等[22]对 11 项随机对照试验、786 名医学生的 Meta 分析显示，生成式 AI 教学与传统教学在知识获取得分上差异无统计学意义(SMD = 0.27, 95% CI −0.31~0.85, P = 0.36)，仅在实践技能(SMD = 0.63, P = 0.02)与学习满意度方面占优。Bastani 等[23]在近千名中学生中的实验发现，无约束使用生成式 AI 的一组练习阶段成绩提升 48%，但撤除访问权限后的测试成绩反而比从未使用 AI 的对照组低 17%；设置只给提示、不给答案护栏的 AI 导师则基本消除了这一效应。Yang 等[24]在泌尿外科学教学中的随机对照试验亦发现，使用 ChatGPT 的学生成绩数值更高但未达统计学意义。AI 带来的即时表现提升并不等同于真实能力的习得。

3.2. 技术保真维度的落差：模拟与真实之间

3.2.1. 技术保真度的天花板

模拟训练解决了练习机会问题，但与真实临床之间仍存在鸿沟。触觉反馈是当前 VR/AR 训练最大的短板：尽管视听觉保真度已大幅提高，力反馈手套与超声波触觉技术仍无法模拟真实组织(皮肤、筋膜、肝脏)的复杂阻力、弹性与手术手感，学生在模拟器上切得动不等于在真人身上切得好[11] [12]。还须注意模拟保真度悖论：过于逼真的模拟(大量出血、音效嘈杂)可能导致初学者认知超载，妨碍核心技能学习[4]；过于简化又无法训练其应对真实压力。

3.2.2. 临床情境与压力的失真

模拟器无法复现真实手术室或抢救室的情境压力——器械护士递错工具、麻醉医生突然报告血压下降、上级医师斥责、团队沟通混乱，而应对这些非技术因素恰是高级医师能力的核心。风险的去真实化亦值得警惕：犯错没有严重后果固然是优点，却可能使学生对医疗风险与患者安全产生钝感[11]。

3.2.3. 临床思维与人文关怀的片面性

VP 擅长训练规则化、标准化的诊疗决策，即遵循指南的循证医学；真实临床的核心却是对“非标准化”信息的处理——患者的表情、语气、家庭支持系统与社会经济状况，即叙事医学。AI 难以模拟患者的恐惧、绝望或依从性不佳，过度依赖 VP 可能培养出精通看病却不懂看病人的技术主义者[13] [21]。

3.3. 评估维度的异化：客观性追求下的三重风险

3.3.1. 算法黑箱与评估标准的僵化

AI 的绝对客观是一把双刃剑，在消除主观偏见的同时也可能带来新的算法暴政与评估异化[18]。AI 的客观基于预设算法与金标准数据，若给出低分却不能解释扣分依据，这种算法裁决难以服众，并打击学生积极性。临床实践尤其是疑难杂症往往没有唯一的金标准答案，过度依赖 AI 评估可能导致防御性学习——学生不敢尝试创新性疗法，只选择最安全、最符合模型的方案，从而压抑批判性思维与临床直觉。

3.3.2. 为 AI 而学的应试化倾向

当 AI 评分成为标准，学生与教师的注意力可能从提升真实能力转向获取 AI 高分：花大量时间练习算法看重的器械平稳度而忽视临床决策，或在书写病历时堆砌模型偏好的关键词而非真正改善病历质量；AI 评估的可重复性还可能诱发刷分，替代真正的反思与提高。

3.3.3. 评估者(导师)的去技能化

一个常被忽视的风险是评估者的去技能化：当导师过度依赖 AI 评分，其评估职业素养、团队协作等软技能的敏锐度将逐渐退化，由指导者沦为 AI 评分的传达者[16] [25]。更需警惕自动化偏见的叠加效应——当 AI 评估结果被赋予过高制度权威，导师与学生均倾向于接受而非质疑其结论。一项系统综述显示，该现象在临床决策支持场景中普遍存在且难以通过简单培训消除[26]。二者相互强化，将使 AI 从辅助判断的工具异化为定义能力的标准，这也是 4.2 节所述制度性风险的技术前因。

3.4. 伦理维度的侵蚀：学术诚信与科研能力的双重挑战

3.4.1. 学术伦理与诚信的新灰色地带

AI 在大幅降低科研门槛的同时，也带来严峻的学术伦理挑战[15]。以 ChatGPT、DeepSeek 为代表的生成式 AI 造成空前的诚信危机：用 AI 润色语法是否允许？用 AI 生成引言初稿或图表数据是否构成抄袭与学术不端？各期刊与学术机构虽已出台初步指南，但对如何界定与甄别 AI 贡献度仍缺乏成熟的技术手段与伦理共识[6] [18]，亟需建立技术甄别与伦理引导机制。

3.4.2. 科研能力的空心化与 AI 依赖

当 AI 可一键完成数据分析与论文初稿，研究生是否会丧失最核心的科研能力？AI 擅长发现相关性却无法区分因果性，研究生可能满足于 $P < 0.05$ 的统计显著性，而丧失对临床意义与生物学合理性的批判性思考。数据挖掘异化为数据疏浚：强大算力可能加剧 P 值操纵——不断更换模型挖掘数据直至得到阳性结果，违背科学研究的初衷。

3.4.3. 科研的数字鸿沟

科研的数字鸿沟同样不容忽视：能够接触高质量专有数据集并拥有强大 GPU 算力集群的顶尖院校研究生，将对资源匮乏机构的学生形成降维打击，AI 的普及若缺乏公共资源平衡，反而可能加剧科研产出与人才培养的贫富差距[5] [12]。

3.5. 批判性维度与应用场景的交叉分析

上述四类议题并非彼此孤立，而是以不同组合嵌入第 2 章的四类场景，并在不同场景下呈现不同严重程度，其对应关系见表 2。数据与算法风险主要集中于个性化学习与考核评估，根源是二者对长周期行为数据的依赖；技术保真维度的风险几乎专属于临床技能训练；评估异化同时作用于考核评估与师资队伍，并经由评价制度向其他场景外溢；伦理维度的风险集中于科研创新，其影响最终传导至人才培养的整体质量。这提示 AI 教育应用的治理不宜“一刀切”，而应按“场景－议题”组合分别设计干预重点。

Table 2. Cross-analysis matrix of critical dimensions and application scenarios
表 2. 批判性维度与应用场景的交叉分析矩阵

批判性议题	主要表现	关联应用场景	主要影响对象	可观测指标
数据与算法维度	数据孤岛、隐私红线、算法偏见与知识茧房	2.1 个性化学习；2.3 考核评估	学生、患者、院校	数据可用率、脱敏合规率、群体间推荐差异度
技术保真维度	触觉反馈瓶颈、情境压力失真、人文关怀缺失	2.2 临床技能训练	学生、教师	技能迁移率、人文沟通评分
评估维度	算法黑箱、应试化倾向、导师去技能化、自动化偏见	2.3 考核评估；5.1 师资队伍	学生、导师	复测成绩差值、评分申诉率、导师独立评估一致性
伦理维度	学术诚信灰色地带、科研能力空心化、数字鸿沟	2.4 科研创新	学生、导师、机构	AI 使用声明率、结果独立复现率

4. 中国情境下的特殊表现与治理难点：基于“规培合一”制度背景的分析

四类议题在不同国家的表现形式并不相同，脱离制度情境的讨论容易沦为通用的技术批判，这也是现有文献本土解释力不足的原因之一。我国自 2015 年起全面推行临床医学硕士专业学位研究生教育与住院医师规范化培训并轨的培养模式，研究生同时具有学生与住院医师双重身份，其培养过程受学位教育与行业准入两套标准的双重约束[27] [28]。这一安排提升了训练规范性，也使 AI 应用的部分挑战呈现出显著不同的形态。评估异化与师资转型是决定 AI 应用成败的两个核心变量，本章围绕二者展开剖析。

4.1. “规培合一”制度对 AI 应用的三重约束

其一，时间约束。并轨培养将住院医师规范化培训压缩进研究生学制，临床轮转、课程学习与科研训练高度叠加，可用于 AI 辅助自主学习的弹性时间极为有限，个性化学习系统所依赖的长周期、连续性行为数据天然难以采集[27]。其二，标准约束。研究生须同时应对学位课程、年度与结业考核，不同体系指标口径不一，AI 评估模型难以在多套标准间建立统一映射。其三，身份约束。双重身份使学习数据同时具有教育数据与医疗数据属性，数据归属、使用授权与责任划分在现行制度中缺乏明确规定，而《中华人民共和国个人信息保护法》[1]对敏感个人信息的处理提出了更严格的要求。三重约束叠加，使第 3 章所述的通用性风险在我国被进一步放大。

4.2. 评估异化的制度性放大：由“为 AI 而学”演变为“为考核而学”

在规培合一的背景下，评估异化呈现出更隐蔽也更顽固的形态。研究生面临出科考核、年度考核、结业考核与学位论文答辩等多重节点，考核结果直接与执业资格和就业挂钩，其应试动机本已较为强烈。当 AI 评估工具被引入后，学生的适应策略未必是提升临床能力，而可能是同时针对 AI 评分逻辑与考核指标进行双重投机：在病历书写中堆砌算法偏好的关键词，在技能训练中反复打磨被算法赋权的动作指标，以换取两套体系下的双重高分。更值得警惕的是，由于结业考核通过率是基地与导师的重要绩效指标，基地存在将 AI 评分作为训练强度调节依据的激励，可能进一步强化指标即能力的误认。此时，AI 并未缓解原有的应试倾向，反而以“客观量化”之名提供了合法性外衣。已有调查显示，并轨培养研究生对考核与薪酬的满意度明显低于对培训内容与方式的满意度，且对带教教师安排的合理性评分偏低[27]，评价环节本就是制度痛点；若缺乏配套治理，AI 的引入极易由技术赋能滑向评价加压。

4.3. 师资转型的双重角色困境：带教医师的第三重负荷

承担并轨培养指导任务的临床带教医师同时具有“临床医师－带教教师－研究生导师”三重角色，

[1]https://www.cac.gov.cn/2021-08/20/c_1631050028355286.htm

其工作时间已被临床诊疗与规培带教高度占用。调查显示，半数以上的并轨培养学员认为自身处于超负荷工作状态，带教教师积极性不高、缺乏有效激励机制是并轨培养质量的主要短板[27]。在此情境下要求带教医师再承担“AI 教学”的第三重负荷，若无制度性激励，几乎必然流于形式。更深层的困难在于角色认知冲突：规培带教强调规范、标准与可考核性，而 AI 时代导师的核心价值在于引导学生批判性审视 AI 输出、处理不确定性并传递人文精神，二者存在目标张力。当 AI 评分与规培考核指标直接挂钩，带教医师更倾向于成为 AI 评分的传达者而非临床判断的示范者，导师的去技能化由此被制度性放大。

4.4. 治理难点与成因分析

上述特殊表现可归结为三方面的治理难点：一是权责分散。学位教育由教育行政部门与高校管理，住院医师规范化培训由卫生健康行政部门与培训基地管理，而 AI 应用涉及的数据治理与评价改革横跨两个系统，跨部门协同的制度成本较高。二是标准缺位。尚缺乏针对医学教育 AI 应用的伦理审查规范、算法审计制度与效果评价标准，院校采购与部署缺少可依据的技术与伦理门槛。三是激励错位。培训基地与带教医师的绩效考核仍以临床工作量与结业通过率为主，AI 教学改革既不计入教学工作量，也缺少稳定的经费与课时支持。三者叠加，使通用性的技术治理建议难以直接落地，这也正是 5.3~5.5 节提出分阶段实施路径的现实依据。

5. 关键瓶颈的破解路径：分阶段实施方案

如果说前述应用场景与批判性议题指向技术问题，那么师资和系统则是决定 AI 教育变革成败的社会问题和组织问题。

5.1. 师资队伍：从抵制者到引导者的艰难转型

AI 时代对医学导师(临床带教老师)提出了前所未有的挑战[15] [25]。当前绝大多数的临床导师自身是数字移民，他们是卓越的临床专家，但对 AI/ML/DL 的基本原理、操作技能和(尤其是)其教学法意义普遍缺乏了解——自己都不会用、不理解、不相信，自然难以指导研究生正确地使用 AI 工具。部分导师则担心 AI (如 AI 影像诊断)会威胁其专业权威，甚至在未来取代其部分工作，因而在心理上对新技术持抵制或漠视态度；长期依赖师傅带徒弟模式的导师，对 AI 评估的客观数据持怀疑态度，更相信自己的临床直觉和主观判断。

最现实的障碍是激励机制的缺失。临床导师本已承担临床与科研的巨大压力，AI 变革又要求其承担起 AI 教学的第三重负荷，而在现有的职称晋升和绩效考核体系中，对教学创新的激励远远不足，导师普遍缺乏动力和时间去学习和实践 AI 教学。与此相应，导师的角色必须从知识的传授者转变为学习的引导者和 AI 的批判性使用者[17]：核心任务不再是灌输知识，而是引导学生批判性地审视 AI 的输出结果，教授 AI 无法替代的能力，如临床决策的智慧与处理不确定性的能力，并通过言传身教传递医学的人文精神。

5.2. 系统与经济：基础设施的重资产壁垒

系统层面的障碍首先表现为高昂的准入成本：高保真的 VR/AR 实验室、企业级的 ITS 平台、私有化部署的大语言模型、高性能 GPU 集群，动辄数百万至数千万的初期投入，是许多中小型医学院校和医院难以承受的，这进一步加剧了教育资源的数字鸿沟。其次是缺乏互操作性的系统孤岛：即使购买了系统，来自不同厂商的 AI 教学系统、电子病历系统(EMR)、教务系统(SIS)与影像系统(PACS)之间往往互不兼容，数据无法流畅整合，AI 赋能的效果大打折扣。第三是持续的维护与更新成本：医学知识不断更新，

AI 模型必须持续用新数据再训练和维护，这需要一支昂贵的、由数据科学家与 IT 工程师组成的团队长期支持，是一笔巨大的隐性成本。

5.3. 分阶段实施路径图

针对前述师资与系统瓶颈，本文提出为期五年的四阶段实施路径，见表 3。四个阶段并非线性递进，而是在第二阶段即启动师资能力建设、在第三阶段同步开展效果评估，形成建设–评估–修正的闭环，以避免技术方案先行、制度与能力建设滞后所导致的建而不用。

Table 3. Phased implementation roadmap for AI-enabled medical postgraduate education
表 3. AI 赋能医学研究生教育的分阶段实施路径

实施阶段	时间跨度	主要任务	责任主体	阶段性标志
第一阶段 奠基与试点	第 1 年	完成 AI 教学基线调研，组建跨学科工作组，遴选 1~2 个试点科室	研究生院、培训基地	形成基线数据集与试点方案
第二阶段 能力建设	第 1~2 年	实施临床导师 AI 能力发展工作坊，制定 AI 使用伦理与学术诚信规范	教师发展中心、伦理委员会	完成首轮导师培训与认证
第三阶段 场景嵌入	第 2~3 年	在技能训练、病历评价等场景部署 AI 工具，同步开展前后对照评估	教研室、信息中心	获得可比的效果证据
第四阶段 制度固化	第 3~5 年	将数据协作体、算法审计与评价改革纳入制度，形成常态化治理机制	院校与行政主管部门	出台校级 AI 教育应用治理办法

5.4. 临床导师“AI + 医学教育”能力发展工作坊课程框架

回应 5.1 节所述的能力赤字与角色重塑需求，本文设计了面向临床导师的 AI + 医学教育能力发展工作坊课程框架，见表 4。该框架以认知–实操–批判–伦理–迁移–追踪为逻辑主线，总学时 32 学时，其中情境模拟与行动学习占比超过 50%，强调在真实带教任务中完成能力迁移；考核方式摒弃纸笔测验，以作品、情境测验与行动方案为主，避免培训本身沦为新的形式主义。工作坊由教育技术专家、临床带教骨干与伦理学教师联合授课，并将培训学时与教学工作量认定挂钩，以回应 4.3 节所述的激励错位问题。

Table 4. Curriculum framework of the “AI + medical education” competency development workshop for clinical supervisors
表 4. 临床导师“AI + 医学教育”能力发展工作坊课程框架

6	主题	学时	教学方式	考核方式
M1 认知启蒙	AI 基本原理与医学教育应用场景辨析	4	讲授 + 案例研讨	案例分析报告
M2 工具实操	生成式 AI 辅助教案设计、虚拟病人构建与智能评阅	8	工作坊实操	教案作品
M3 批判性使用	幻觉识别、自动化偏见防范与 AI 输出的临床核验	6	情境模拟 (含错误输出辨析)	情境测验
M4 伦理与治理	数据合规、学术诚信边界与学生 AI 使用指导	4	伦理困境讨论	伦理决策陈述
M5 教学迁移	将 AI 工具嵌入规培带教与出科考核的行动设计	6	行动学习与同伴互评	微课题方案
M6 追踪复训	教学实践反馈、效果数据与年度知识更新	4/年	线上社群 + 年度复训	年度教学反思报告

5.5. 区域性医学教育数据协作体的初步构想

针对 3.1.1 节指出的数据供给瓶颈，本文提出在省级层面构建区域性医学教育数据协作体的初步构想，包含五项要素。1) 治理架构：由省级教育行政部门与卫生健康行政部门联合发起，依托高校与住培

基地设立协作体理事会，下设数据治理委员会与伦理审查快速通道，明确数据不出域、用途限定、责任可追溯三条底线[1]。2) 技术架构：采用“联邦学习 + 隐私计算”的双层结构，各成员单位在本地完成数据治理与模型训练，仅交换加密的模型参数或中间结果，不共享原始记录；多中心研究已表明，该路径在不迁移原始数据的前提下可获得接近集中式训练的模型性能，并显著优于单中心模型[19] [20]。3) 数据标准：统一学习者标识、轮转记录、技能考核项与病例标签的最小数据集，优先解决 5.2 节所述的互操作性问题。4) 激励机制：以数据贡献度与模型增益为依据分配协作体成果使用权，并将数据治理工作量纳入带教医师与技术人员的工作量认定。5) 风险控制：对学生“犯错”的过程性数据实行单独的去标识化与限期销毁策略，避免其被误用于终结性评价。需要说明，该构想目前仍处于设计阶段，其技术可行性与治理成本有待在试点中检验。

6. 结论

AI 技术为破解传统医学研究生教育的瓶颈提供了历史性机遇，通过重构学习、训练、评估与科研范式，深刻改变着高层次医学人才的成长路径。然而，本文的批判性分析表明，技术应用绝非坦途：AI 在带来效率和客观性的同时，也伴随数据偏见、现实差距、评估异化、伦理失范和教育不公的潜在风险。这不仅是一场技术革命，更是一场教育思想、组织结构与利益格局的深刻重塑——最大的瓶颈往往不在于技术本身，而在于人(师资)的转型和系统(体制)的变革。

未来的医学教育不应是 AI 替代人类，而必须是人机协同、AI 辅助人类。为此，必须将技术研发与伦理规范、师资培训和系统改革置于同等重要的位置，加强顶层设计，打破数据壁垒，投入人力物力推动师资队伍的能力转型，并加快完善 AI 教育应用的伦理与法规框架；最终目标不是以 AI 替代教师，而是在坚持医学人文精神与教育本质的基础上，实现人机协同驱动的医学研究生教育高质量发展。

本文亦存在明显局限：作为批判性综述，本文证据以已发表综述与实证研究为主，未开展原始数据收集，所引研究多为小样本、短期结局，结论外推需谨慎。其次，关于规培合一情境下评估异化与师资转型的分析，主要依据制度文献与区域性调查，尚缺乏跨省、跨基地的比较证据。同时，5.3~5.5 节提出的实施路径、工作坊课程框架与数据协作体构想均为设计性方案，其可行性与实际效果有待在实践中检验。后续研究将围绕工作坊课程框架的试点评估与数据协作体的技术验证展开。

基金项目

广西学位与研究生教育改革课题(GJY2024084, GJY2024093)；广西高等教育本科教学改革工程项目(2025JGA151, 2025JGB156)；2026 年度广西重点研发计划(桂科 FN2600640239)。

参考文献

[1] 余梦雪, 蔡昕昱, 王刚, 等. 人工智能在医学研究生教学中的应用现状与分析[J]. 中国继续医学教育, 2025, 17(5): 170-174.

[2] 刘睿, 刘莹. 大数据与人工智能在医学教育中的应用及发展趋势[J]. 中国高等医学教育, 2026(1): 40-42, 65.

[3] Crompton, H. and Burke, D. (2023) Artificial Intelligence in Higher Education: The State of the Field. *International Journal of Educational Technology in Higher Education*, **20**, Article No. 22. https://doi.org/10.1186/s41239-023-00392-8

[4] 袁靳闲, 李金政. 人工智能应用于医学教育的现状、挑战与前景[J]. 中华医学教育探索杂志, 2025, 24(6): 777-783.

[5] 徐梓硕, 曲巍, 徐晶格. 人工智能赋能医学教育的伦理困境与疏解[J]. 医学与社会, 2026, 39(3): 320-328.

[6] 黄明芳, 侯青涵, 张韦. 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.

[7] 陈露露, 张偲, 涂宁, 等. ChatGPT 在医学教育中的应用分析[J]. 中国高等医学教育, 2026(2): 15-17.

[8] Verghese, B.G., Iyer, C., Borse, T., Cooper, S., White, J. and Sheehy, R. (2025) Modern Artificial Intelligence and Large

Language Models in Graduate Medical Education: A Scoping Review of Attitudes, Applications & Practice. *BMC Medical Education*, **25**, Article No. 730. https://doi.org/10.1186/s12909-025-07321-5

[9] Acosta, J.N., Falcone, G.J., Rajpurkar, P. and Topol, E.J. (2022) Multimodal Biomedical AI. *Nature Medicine*, **28**, 1773-1784. https://doi.org/10.1038/s41591-022-01981-2

[10] Wang, H., Shan, W., Liu, R. and Wang, Z. (2025) Can Large Language Models Serve as Digital Assistants for Medical Undergraduates? A Bibliometric Mapping and Scoping Analysis of the Medical-Education Literature. *Digital Health*, **11**. https://doi.org/10.1177/20552076251390280

[11] 蒲军, 唐华, 全真, 等. 虚拟现实技术应用于外科学研究生手术技能培训效果与评价[J]. 中华医学教育探索杂志, 2020, 19(8): 941-944.

[12] 张俊祥, 李传富, 吕维富. 人工智能在医学教育、科研和临床实践中的应用前景与挑战[J]. 中华全科医学, 2024, 22(7): 1085-1089.

[13] 樊宁, 申晓丽, 司姣, 等. 增强现实教学在眼科专业型硕士生显微手术培训中的应用[J]. 科教导刊, 2021(25): 58-60, 68.

[14] 陈元, 孙书珍, 成海燕, 等. 虚拟标准化病人用于儿科专业住院医师临床思维能力培养的探索[J]. 中国毕业后医学教育, 2023, 7(9): 749-752.

[15] 陈松, 黄雷, 杨洁, 等. 人工智能背景下医学研究生教育的变革、问题与对策[J]. 实用医院临床杂志, 2026, 23(3): 230-234.

[16] Triola, M.M. and Rodman, A. (2024) Integrating Generative Artificial Intelligence into Medical Education: Curriculum, Policy, and Governance Strategies. *Academic Medicine*, **100**, 413-418. https://doi.org/10.1097/acm.0000000000005963

[17] Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., *et al*. (2024) A Meta Systematic Review of Artificial Intelligence in Higher Education: A Call for Increased Ethics, Collaboration, and Rigour. *International Journal of Educational Technology in Higher Education*, **21**, Article No. 4. https://doi.org/10.1186/s41239-023-00436-z

[18] 吴佳蕊, 李昭莹, 夏悦箐, 等. 生成式人工智能赋能医学教育发展的脉络因子、风险表征与纾困路径[J]. 中国卫生事业管理, 2026, 43(2): 195-199.

[19] Kuo, T., Gabriel, R.A., Koola, J., Schooley, R.T. and Ohno-Machado, L. (2025) Distributed Cross-Learning for Equitable Federated Models—Privacy-Preserving Prediction on Data from Five California Hospitals. *Nature Communications*, **16**, Article No. 1371. https://doi.org/10.1038/s41467-025-56510-9

[20] Wang, D., Qi, G., Li, J., Wang, Y., Dong, K., Ding, J., *et al*. (2025) The Development and Validation of a Privacy-Preserving Model Based on Federated Learning for Diagnosing Severe Pediatric Pneumonia. *Translational Pediatrics*, **14**, 1287-1295. https://doi.org/10.21037/tp-2025-349

[21] Singhal, K., Azizi, S., Tu, T., Mahdavi, S.S., Wei, J., Chung, H.W., *et al*. (2023) Large Language Models Encode Clinical Knowledge. *Nature*, **620**, 172-180. https://doi.org/10.1038/s41586-023-06291-2

[22] Li, J., Yin, K., Wang, Y., Jiang, X. and Chen, D. (2025) Effectiveness of Generative Artificial Intelligence-Based Teaching versus Traditional Teaching Methods in Medical Education: A Meta-Analysis of Randomized Controlled Trials. *BMC Medical Education*, **25**, Article No. 1175. https://doi.org/10.1186/s12909-025-07750-2

[23] Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö. and Mariman, R. (2025) Generative AI without Guardrails Can Harm Learning: Evidence from High School Mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, **122**, e2422633122. https://doi.org/10.1073/pnas.2422633122

[24] Yang, W., Xu, T., Wei, J., Zheng, W., Yan, W., Wang, J., *et al*. (2026) Effectiveness of ChatGPT and DeepSeek in Urology Medical Education: Randomized Controlled Trial. *Journal of Medical Internet Research*, **28**, e89315. https://doi.org/10.2196/89315

[25] 王振宇, 郑晓洁, 张靖. 医学院校教师人工智能伦理素养研究[J]. 中华医学教育探索杂志, 2026, 25(2): 134-138.

[26] Goddard, K., Roudsari, A. and Wyatt, J.C. (2012) Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators. *Journal of the American Medical Informatics Association*, **19**, 121-127. https://doi.org/10.1136/amiajnl-2011-000089

[27] 谭梅美, 李慧娟, 甘继瑞, 等. 关于临床医学硕士专业学位研究生培养与住院医师规范化培训并轨模式的探讨[J]. 医学教育研究与实践, 2020, 28(4): 611-614.

[28] 关于印发临床医学、口腔医学和中医硕士专业学位研究生指导性培养方案的通知[EB/OL]. http://www.moe.gov.cn/srcsite/A22/moe_826/201506/t20150618_190613.html, 2015-05-29.