生成式人工智能融入医学教育的教学设计、风险防控与治理路径

赵劲歌
四川大学华西医院，泌尿外科 , 四川 成都 610041
DOI: 10.61369/SDME.2026020026

摘　　要： 生成式人工智能（generative AI，GenAI）尤其是以大语言模型为代表的对话式工具，正在快速改变医学教育的知识获取方式、教学组织形态与评价反馈流程。已有研究与教学实践提示，GenAI 可在课程内容生产、病例与题库生成、形成性评价反馈、临床推理训练与学习支持等方面显著提升效率，并可能促进以胜任力为导向的课程与评价改革；但医学教育对信息准确性、伦理合规与患者安全的高要求，使得幻觉与错误信息、学术诚信与认知外包、偏倚与公平性、隐私与数据安全、责任边界不清等风险在该场景中被放大。本文在综合国内外文献的基础上，提出面向教学一线与管理者的可执行路径：以“分级使用、人在回路、透明披露、可追溯审计、数据最小化与 AI 素养建构”为原则，从机构制度、课程设计、考核重构、师资发展与技术护栏五个方面形成闭环治理，为 GenAI 在医学教育中的规范应用提供教学论文式的实施框架。

关 键 词： 生成式人工智能；大语言模型；医学教育；教学设计；学术诚信；教学评价；治理

Instructional Design, Risk Prevention and Control, and Governance Paths of Integrating Generative AI into Medical Education

Zhao Jin'ge
Department of urology, West China hospital, Sichuan university, Chengdu, Sichuan 610041

Abstract: Generative artificial intelligence (GenAI), particularly large language model–based conversational tools, is rapidly reshaping medical education across teaching, learning support, assessment, and educational governance. Emerging evidence indicates that GenAI can improve efficiency in content generation, case and item development, and formative feedback, and may facilitate competency-based curricular and assessment reforms. However, the medical education context amplifies critical risks, including hallucinations and misinformation, academic integrity threats and cognitive offloading, algorithmic bias and inequity, privacy and data security concerns, and unclear accountability. Drawing on Chinese and international literature, this teaching-oriented paper proposes an implementable pathway grounded in risk-tiered use, human-in-the-loop oversight, transparent disclosure, auditability, data minimization, and structured AI literacy training. A practical governance package is presented from institutional policies to curriculum design, assessment redesign, faculty development, and technical safeguards.

Keywords: generative AI; large language models; medical education; instructional design; academic integrity; assessment; governance

前言

医学教育长期面临“知识迭代快、课程容量大、临床情境复杂、反馈周期长”的现实压力。传统课堂在知识传递方面仍有效，但对临床推理、沟通协作、循证决策与专业精神等能力目标的支持不足，教师在课程组织、病例脚本撰写、题库建设与形成性评价反馈上的时间成本持续上升。在这一背景下，GenAI 以“低门槛、可对话、可生成、可迭代”的方式提供了新的教学生产力。然而，医学教育的迁移终点指向临床实践，任何“看似合理但不准确”的输出都可能在学习者心中固化为错误规则，从而形成患者安全隐患。因此，GenAI 在医学教育中的合理定位应当是“教学脚手架与效率工具”，而不是“知识权威与评分主体”。这一定位与国内关于 ChatGPT 介入医学教育伦理风险的讨论相一致：GenAI 可能带来师生信息风险、交往异化、算法裹挟与学术公平失信等问题，必须通过多层面协同治理来确保其正向效能 [1]。

国际医学教育界同样强调机构应尽快建立政策、治理与课程体系，明确学生在教育情境中使用 GenAI 的边界并保护机构与患者数

据，从而在技术变革中维持教育质量与公众信任[2]。由此，本文将“教学论文”的关注点放在“如何教、如何用、如何评、如何管”的可操作方案上：既承认 GenAI 带来的效率红利，也坚持医学教育的底线要求，即可核验、可追溯、可辩护、可合规。进一步而言，“是否允许使用”不宜作为唯一的治理工具。更可行的路径是：根据任务风险与学习目标对使用场景进行分级，把 GenAI 从“结果导向的答案提供者”转化为“过程导向的思维教练”。在这一转化中，教师的关键职责并未被削弱，而是发生了迁移：从“内容讲授者”更多转向“学习设计者、质量把关者与评价建构者”。这一角色转型也与国内对 ChatGPT 推动医学教育模式变化的反思相呼应，即需要前瞻性思考 AI 时代医学生培养重点，并强化对临床思维、自主学习与人文素养的引导[3]。

一、面向课堂的教学设计：把GenAI从“答案工具”转为“思维脚手架”

课堂教学中，GenAI 最容易产生立竿见影效果的环节是教学材料的生成与课堂互动的组织。国内针对生成式 AI 大语言模型在医学教育实践中的探讨指出，GenAI 可用于生成教学内容、临床案例、讨论提纲与教学辅助文本，但同时可能引发错误与偏见、学术诚信与教育公平、隐私与数据安全、过度依赖等问题，因此需要明确使用边界并强化人工核验[4]。

基于上述证据与教学现实，本文建议将课堂应用拆解为三类典型任务，并分别配置不同的核验与评价要求。第一类是“低风险教学生产任务”，如课程大纲初稿、术语解释、知识点框架、病例讨论问题清单、OSCE 站点情境草案等。这类任务的核心是提升效率，但必须要求教师对关键医学事实逐条核验，尤其是检查指征、药物用法、禁忌证、诊疗路径与指南表述等内容。第二类是“中风险教学互动任务”，如在 PBL/TBL 课堂中用 GenAI 生成追问链条、模拟患者提问、对学生答案提出反例挑战。此时 GenAI 更适合作为“挑战者”而非“裁判”，教师应把控课堂节奏并对模型输出进行即时纠偏，避免学生把模型的流畅表达误认为权威结论。第三类是“高风险临床决策相关任务”，如以真实病例细节让模型给出诊疗建议。该类任务应原则上避免在开放式公共模型中进行，若确需用于教学演示，必须使用脱敏案例、明确“非指南、非医嘱”的教学声明，并以权威指南与本院路径进行对照核验，强调“证据链优先于结论”。

在具体教学法上，更推荐将 GenAI 嵌入“过程可见”的任务结构中，而不是让学生直接提交“看起来完美”的产出文本。例如在临床推理教学中，可以将作业设计为三段式：学生先独立完成鉴别诊断与证据链；随后用 GenAI 生成“反例与遗漏点清单”；最后学生提交修订后的推理路径与反思，说明哪些建议被采纳、哪些被拒绝以及拒绝的依据。这样做的目的，是把 GenAI 的价值锁定在“促使学生暴露并修正推理过程”，而非替代推理本身。国内关于医学实践课程改革趋势的研究也提示，AI 背景下的教学应更强调过程性学习、反馈与能力培养，而非简单的知识堆叠[5]。

此外，课堂层面的“提示词规范”本身也可以成为教学内容。教师可示范三类提示词：一类用于“生成多样化情境”，例如同一疾病的不同主诉表达与并发症组合；一类用于“要求给出来源与不确定性”，例如要求模型标注关键判断点并提示可能的错误来源；一类用于“引导反思与纠错”，例如让模型针对学生推理提出质疑而不是给结论。通过把提示词当作“思维工具”教授给学生，可将 GenAI 使用从隐性行为转为可讨论、可评价的学习技能。

二、学习支持与能力培养：AI 素养纳入胜任力的教学实施

GenAI 对学习端的影响往往更深远，因为它可能改变学习者的知识获取习惯与思维负荷分配。一方面，GenAI 能够提供即时答疑、概念解释、学习计划建议与写作反馈，降低学习门槛并提高学习连续性；另一方面，若缺乏规范与训练，学习者很容易将检索、整合、判断与表述外包给模型，从而出现“认知外包”与“临床推理能力空心化”。围绕这一矛盾，国内有关医学生对 GenAI 融入医学教育认知的研究指出，学生对 GenAI 的接受度与担忧并存，反映出“可用性”与“可信度、规范性”之间的张力[6]。

因此，把 AI 素养纳入胜任力培养目标，不是附加模块，而应成为课程的基础能力要求。本文建议在教学实施中至少覆盖五个维度。第一，信息核验能力，即学习者能将模型输出与教材、指南、系统综述等权威证据进行对照，识别不一致与潜在幻觉；第二，证据表达能力，即能清晰说明结论来自何种证据、证据强度如何以及存在何种不确定性；第三，偏倚识别能力，即能意识到模型可能对疾病谱、语言表达、特殊人群存在系统性偏差，从而主动补足视角；第四，数据安全与隐私意识，即能遵循“数据最小化与脱敏输入”原则，不在不受控环境中输入可识别信息；第五，专业精神与诚信意识，即能在使用 GenAI 时进行透明披露，承担对产出内容的责任。

护理教育领域对 ChatGPT 应用的讨论尤其强调学术诚信与专业判断问题，认为其在带来效率提升的同时，可能引发依赖与失真，需要通过规范与培训加以引导[7]。医学遗传学等课程的教学实践也显示，GenAI 可用于概念解释与学习支持，但必须避免学生以其替代原文阅读与推导过程，关键在于明确“允许何种辅助、禁止何种代办”[8]。

在教学组织上，AI 素养训练宜采用“案例化、反例化、任务化”的方式。所谓案例化，是把真实课堂中的错误输出作为教学材料，让学生练习核验与纠错；所谓反例化，是故意让模型在模糊提示词下生成可能的误导性建议，从而训练学生识别“流畅语言并不等于正确医学”；所谓任务化，是将提示词记录、证据链提交与反思报告纳入作业评分标准，使“正确使用”成为可评价的学习成果。国际观点研究同样提醒，GenAI 对批判性思维既可

能产生促进，也可能损害学习者认知自主性，关键在于教育设计是否将“思维过程”置于评价中心[9]。

三、教学评价与考核重构：形成性评价、过程证据与可辩护性

GenAI 进入教学评价环节时，风险与争议最集中，但也是推动评价改革的关键窗口。实践中常见的矛盾是：如果仍以“产出文本质量”作为核心评分依据，学习者使用 GenAI 提升文本外观将变得几乎不可避免；而若简单禁用，则会切断教师利用 GenAI 提升反馈效率的机会，并将使用行为推向隐蔽化。解决这一矛盾的路径，不是“全面放开或全面禁止”，而是以任务类型为导向重构评价证据结构，把“过程证据”纳入成绩构成，并对不同风险等级考核施加不同约束。

国内对类 ChatGPT 大语言模型在护理课程考核中的研究具有直接启示：模型在基础知识题上准确率可超过90%，但在需要知识应用的题目上准确率较低且存在差异，提示其在“知识性回答”与“情境应用”之间存在能力断裂[10]。这意味着，若将模型输出直接引入高风险评价（例如关键胜任力结业考核、准入性考试或临床决策能力评价），将难以保证评价的稳定性与可辩护性。相较之下，形成性评价更适合“在规范前提下使用”：教师可以用 GenAI 生成反馈模板、指出结构性问题、提出改进建议，从而提升反馈频率与一致性，但必须要求学习者对事实性内容自行核验并承担责任。

评价框架方面，国内基于 CIPP 理论的 AI 赋能医学教学评价研究提出，可从背景、投入、过程与结果四个维度系统推进 AI 与评价融合[11]。将该思路落地到 GenAI 场景，可形成三条可执行规则。其一，允许使用的作业必须提交“使用声明与提示词记录”，并在文末用简短段落说明使用范围与人工核验方式；其二，课程成绩应提高“过程性证据”权重，例如要求提交草稿迭代、证据链、反思报告或口头答辩，以降低“最终文本”对成绩的决定性；其三，高风险考核应限制 GenAI 介入，或将其用途限定为后台管理性工作（如试题排版、评分表格整理）而不进入评分决策。

围绕临床实践教学场景，亦有研究讨论 ChatGPT 对医学临床实践教学的影响与思考，提示应在教学中强调正确使用场景与风险意识，避免学生滥用与依赖[12]。进一步的实践证据还来自住院医师培训考核：协和医学杂志报道生成式 AI 在重症医学住培考核中的应用探索，提出其在试卷生成与批阅评分方面与人工水平相当，但题目难度优化仍是问题，并强调未来需探索与现有教育体系的深度融合[13]。这些研究共同指向同一结论：评价体系必须从“文本产出”转向“能力证据”，否则 GenAI 将以不可控方式重塑评价生态。

四、风险防控与治理路径：分级使用、人在回路、可追溯审计

医学教育中的 GenAI 治理，应当以“患者安全与教育公平”为价值底座，以“流程可执行”为核心要求。国内关于 ChatGPT 介入医学教育伦理风险的研究提出协同治理思路，强调需从科学技术、政府部门、教学单位与师生个体等多层面共同应对[1]。与之相呼应，国际医学教育治理建议强调制定适当使用政策、保护机构与患者数据并为学生提供清晰边界[2]。 在此基础上，本文提出一套“分级使用 + 人在回路 + 透明披露 + 可追溯审计 + 数据最小化”的闭环路径，便于医学院或附属医院以制度与流程方式落地。

第一，建立分级使用清单，把教学任务按风险分为允许、限制与禁止三类。允许类通常为低风险生产任务与学习辅学任务，如生成教学大纲初稿、病例讨论问题清单、学习计划建议、写作结构优化等；限制类为中风险互动与反馈任务，如课堂追问链、形成性评价反馈、OSCE 脚本初稿与题目草案等，要求教师复核与来源核验；禁止类通常为高风险任务，如输入真实可识别患者信息、让模型在不受控环境中给出真实病例诊疗建议并作为学习者决策依据等。该清单应与课程目标、学生年级与教学场景绑定，并定期更新。

第二，坚持“人在回路”，明确责任边界与审核节点。对于进入课堂与考核的材料，至少应设置两类审核：事实核验（与教材、指南、路径一致性）与伦理合规核验（是否含隐私、是否存在歧视与偏倚表述）。尤其在题库建设与 OSCE 脚本中，审核要覆盖题干逻辑、答案唯一性、误导性表述与难度适配。对形成性评价，教师可使用 GenAI 提高反馈效率，但必须明确反馈的“建议性质”，并引导学生完成自证与反思。

第三，落实透明披露与学术诚信规则。建议学校出台统一的“GenAI 使用声明”要求：凡在作业、报告、教学材料中使用 GenAI，应说明使用工具类型、使用环节（如结构优化、语言润色、生成问题清单等）、人工核验方式与未使用的环节。此举的目的不是增加负担，而是让使用行为从隐性走向可讨论、可管理。护理与医学教育相关文献普遍提示学术诚信与过度依赖风险，透明披露是治理的基础设施之一[7]。

第四，建立可追溯审计与最小化数据原则。审计并不等于监控个体，而是为高风险事件提供可复盘的证据链。建议机构在允许场景中要求保留关键提示词、输出版本与使用时间，特别是进入考核环节的材料应保留生成与审核记录。同时，对数据输入坚持最小化：不输入姓名、住院号、精确时间线与可组合识别信息，教学病例采用统一脱敏模板，必要时使用机构内部受控环境或本地知识库方案。

第五，把“风险识别与批判性使用”纳入师资发展与课程评估。GenAI 治理的长期效果，取决于教师队伍能否形成稳定的核验与设计能力。建议将教师培训聚焦在三点：提示词驱动的教学活动设计、事实核验与来源追溯能力、以过程证据为核心的评价重构能力。与此同时，课程质量评价应增加与 GenAI 相关的指标，例如学生对核验流程的掌握程度、使用声明规范度、反思质量与临床推理证据链的完整性。国内对 ChatGPT 医学研究热点的文献计量分析指出，ChatGPT 相关研究快速增长，但也存在技术盲区、数字依赖与伦理困境，需要审慎开展伦理反思与风险防范策略探索[14]。这提示治理不应一次性定稿，而应伴随技术与教学

实践演进持续迭代。

在风险议题上，国外对生成式 AI 伦理与实践挑战的综述进一步强调系统性偏倚、法律责任不清、黑箱性与数据隐私风险，并提出技术、程序与监管层面的综合应对方案[15]。这些观点可为国内医学教育治理提供参照：对于高风险场景，要优先采用制度与流程约束，而非寄希望于个体自律；对于中低风险场景，则应鼓励在规范内探索，以教学创新反哺治理完善。

五、结语

GenAI 进入医学教育并非简单的“新工具替代旧工具”，而是对教学设计、学习方式与评价逻辑的系统性重构。教学层面，最关键的转向是将 GenAI 从“答案提供者”转化为“思维脚手架”，以过程性证据、证据链表达与反思性任务促进临床推理能力成长；管理层面，应以分级使用、人在回路、透明披露、可追溯审计与数据最小化为底线，构建可执行、可评估、可迭代的治理闭环。未来研究需要在真实世界教学场景中进一步检验：GenAI 对临床推理与专业精神的长期影响、不同课程类型下的最佳分级政策组合，以及本土化合规工具包（脱敏模板、使用声明、核验清单与审计规范）的有效性。只有当制度、课程、考核、师资与技术护栏同步到位，GenAI 才能在医学教育中实现“增效而不降质、创新而不越线”。

参考文献

[1] 王茹俊，王丹 . ChatGPT 介入医学教育的伦理风险及应对策略 . 医学与哲学 2024；45(02)：76-81.

[2]Triola MM, Rodman A. Integrating Generative Artificial Intelligence Into Medical Education: Curriculum, Policy, and Governance Strategies. Acad Med 2025; 100(4): 413-8.

[3] 瞿星，杨金铭，陈滔，张伟 . ChatGPT 对医学教育模式改变的思考 . 四川大学学报 (医学版) 2023；54(05)：937-40.

[4] 陈湘，邓然，吴川清 . 生成式人工智能大型语言模型在医学教育实践的探讨 . 临床急诊杂志 2024；25(06)：310-4.

[5] 程珊，丛林，胡文东，熊凯文，马进 . "AI+ 教育 " 时代背景下医学实践课程教学模式现状与改革趋势 . 医学新知 2024；34(08)：950-6.

[6] 袁志怡，裴志怡，林佳艺，贾烨如，康晓凤 . 医学生对生成式人工智能融入医学教育的认知——一项 Meta 整合研究 . 医学与哲学 2025；46(14)：52-7.

[7] 罗华宇，许敏，曾朝蓉，田晓宇，李鑫 . ChatGPT 在护理领域中应用的前景与挑战 . 中华护理教育 2023；20(12)：1520-3.

[8] 杨玲，刘雯，左伋 . ChatGPT 在医学遗传学教学中的运用 . 中国优生与遗传杂志 2023；31(06)：1283-5.

[9]Izquierdo-Condoy JS, Arias-Intriago M, Tello-De-la-Torre A, Busch F, Ortiz-Prado E. Generative Artificial Intelligence in Medical Education: Enhancing Critical Thinking or Undermining Cognitive Autonomy? J Med Internet Res 2025; 27: e76340.

[10] 徐文博，周晓平 . 类 ChatGPT 大语言模型在护理课程考核中的应用探索——基于 ChatGPT、文心一言、讯飞星火测试 . 中国医学教育技术 2024；38(05)：567-71.

[11] 刘晖，何欣悦，寇丽圆，任曼 . 基于 CIPP 理论的 AI 赋能医学教学评价研究 . 中国医学教育技术 2024；38(05)：600-5.

[12] 刘晖，任曼，寇丽圆，夏小川，何欣悦 . ChatGPT 对医学临床实践教学的影响与思考 . 中国医学教育技术 2024；38(04)：389-93+400.

[13] 张博，何云涓，顾新龙，等 .DeepSeek 赋能住院医师规范化培训实践探讨 [J]. 现代医院 ,2025,25(5):764-766.DOI:10.3969/j.issn.1671-332X.2025.05.027.

[14] 吕静，何平，王永芬，et al. ChatGPT 在医学领域研究态势的文献计量学分析 . 医学与哲学 2024；45(07)：30-5.

[15]Tung T, Hasnaeen SMN, Zhao X. Ethical and practical challenges of generative AI in healthcare and proposed solutions: a survey. Front Digit Health 2025; 7: 1692517.

