人工智能时代医学诊断基本功的退化隐忧与教学应对

于浩鹏，李 雨，王雨琪，刘冰洁，刘曦娇，李峥艳*

四川大学华西医院放射科　四川成都

【摘要】AI 进入诊室和阅片室后，年轻医师常常先看到结论，再回头找依据。问诊、查体、阅片、写病历这几步被压缩，诊断基本功的形成顺序也跟着改变。本文讨论的正是这种变化。文章先区分去技能化、从未习得、错误技能化和自动化偏差等概念，再把医学诊断基本功拆成感知、知识、推理三层，分析 AI 介入时机、任务层级和监督方式如何影响能力形成，并梳理国际实验、各专科教学研究和中文文献的证据。现有研究多属短期、单中心，结局以准确率和满意度为主，只能提示风险，尚不足以证明因果关系。据此，文章提出基础期 AI-free、整合期 Guided、协同期 Supervised 的三阶段防线，以及无 AI 基线与 AI 辅助表现并行的双轨评价。放射科是文章的主要示例，讨论范围则覆盖内科、外科、急诊等以诊断为核心的临床学科。底线很清楚。离开 AI，医师仍要能独立、安全地完成诊断。

【关键词】人工智能；医学诊断；诊断基本功；技能退化；认知卸载；自动化偏差；住院医师规范化培训；医学教育

【基金项目】华西临床医学院 2024 年研究生教育教学改革研究项目（HXYJS202416）

【收稿日期】2026 年 10 月 12 日　【出刊日期】2026 年 11 月 25 日　【DOI】10.12208/j.ije.20260157

The hidden cost of artificial intelligence in medical diagnosis education: Concerns about the erosion of diagnostic foundational skills and educational safeguards

*Haopeng Yu, Yu Li, Yuqi Wang, Bingjie Liu, Xijiao Liu, Zhengyan Li**

Department of Radiology, West China Hospital, Sichuan University, Chengdu, Sichuan

【**Abstract**】AI now reaches young physicians before they have formed their own diagnostic reading. In the clinic and the reading room, answers often arrive before history-taking, physical examination, image interpretation, or note-writing are complete, and that order may reshape how diagnostic foundations develop. This narrative review distinguishes deskilling, never-skilling, mis-skilling, automation bias, and cognitive offloading; divides diagnostic competence into perceptual, knowledge, and reasoning layers; and examines how timing, task level, and supervision influence learning. Evidence from international experiments, specialty-specific teaching studies, and Chinese literature is reviewed. Most studies are short-term, single-center, and use surrogate endpoints, so they signal risk rather than prove causation. A three-stage safeguard—AI-free baseline, guided integration, supervised AI-augmented practice—and dual-track assessment are proposed. Radiology serves as the main example, while the discussion extends to internal medicine, surgery, emergency medicine, and other diagnostic specialties. The bottom line: physicians should be able to diagnose independently and safely when AI is unavailable.

【**Keywords**】Artificial intelligence; Medical diagnosis; Diagnostic foundational skills; Deskilling; Cognitive offloading; Automation bias; Standardized residency training; Medical education

1 引言

诊室里，一位低年资住院医师先打开 AI 辅助诊断界面，看到“胰腺癌可能”的提示，随后照着这个思路写病历。上级医师问起疼痛的放射方式、体重变化、既往胰腺炎史，又让他复述腹部查体的发现和影像上的关键征象，他却答不完整。AI 给出的方向未必错，被

*通讯作者：李峥艳

跳过的是问诊、查体、阅片、建立假设和验证证据这一整套过程。答案提前出现，学习者还需不需要经过这些步骤？

大语言模型、多模态模型和医学 AI 智能体已经进入医学教育。它们能生成病例、解释检查结果、修改病历、模拟沟通，也能按学习者的薄弱点出题。2026 年《Nature Medicine》提出“AI 诱导的从未习得”：学习者在临床推理尚未成形时就依赖 AI，可能无法建立独立执业所需的基础能力[1]。系统综述和范围综述也提到自动化偏差、认知去技能化和解释自主性下降[2,3]。国内血液内科、全科医学和医学教育领域的综述已经梳理了这些应用及其风险[4-6]。把诊断基本功作为核心、跨专科讨论这一问题的中文梳理，仍然少见。

放射科是本文的主要示例，原因在于这里的链条完整：看片、描述征象、做鉴别、写报告、和临床沟通，每一环都能被 AI 介入；影像数字化程度高，AI 工具进入工作流早；报告结构相对规范，便于评价；住培对读片量和报告质量有明确要求。同样的顺序问题也出现在内科问诊、外科术前评估、急诊快速决策、病理和内镜判读中。作者团队此前的综述讨论过生成式 AI 在医学实践教育中的应用与思政融入[7]。这一篇换一个方向：不列应用清单，而是追问 AI 可能削弱什么、证据有多硬、教学上怎么防。放射科贯穿全文，结论面向整个医学诊断教育。

2 概念辨析：退化、未形成与错误技能化

讨论“AI 是否让学习者变差”之前，先要区分几种不同的失败形态。中文文献常把它们统称为“能力退化”，但不同概念对应的机制、对象和干预方式并不相同。概念不清，容易把“学生本来就不会”误判为“AI 让他退化”，也容易把“AI 用得好”误判为“能力增强”。

2.1 去技能化（deskilling）

去技能化指的是“曾经会，后来生疏”。航空和工业控制领域早有观察：自动化系统越可靠，操作者手动操作的次数越少；一旦系统失效或遇到罕见情境，表现便明显下降[8,9]。临床诊断也有类似路径。资深内科医师长期只复核 AI 提示，问诊的条理和查体的手法会生疏；病理医师只看 AI 初筛结果，镜下搜索和形态判断会变慢；放射科医师长期只做 AI 初筛后的复核，独立搜索、描述征象和鉴别推理的机会减少。遇到 AI 没有提示的细微病变、少见病或伪影，漏诊和误判就容易出现。这种退化往往悄无声息，甚至以“效率提高”的形式出现。等到系统升级、发生故障，或遇到超出训练数据的病例，能力缺口才会显现。

2.2 从未习得（never-skilling）

从未习得则是“从来没有真正学会”。关键能力还没形成，学习者就先依赖 AI，诊断图式和推理框架没有搭起来[1]。它和去技能化的区别，可以概括为“生疏”与“没学会”。医学生和低年资住院医师更容易落入这一状态：第一次接触某个症状、体征或征象就看到 AI 的结论，独立观察、描述、假设、验证的过程被跳过；以后即使能复述答案，也未必能在没有 AI 时完成同样的诊断。Ke 等[1]把这种现象与“错误技能化”分开，强调它发生在能力形成的早期窗口。

2.3 错误技能化（mis-skilling）

错误技能化是指把 AI 的错误当成知识，写进自己的诊断路径[1,3]。生成式 AI 的幻觉具有语言流畅、语气自信的特点，初学者仅凭文本很难判断真伪。前述随机对照试验中，大语言模型单独答题的表现优于医师组，但医师使用大语言模型后，诊断推理评分并未显著提高[10]。会用 AI，不等于能识别 AI 的错误；人的推理能力也未必随之提高。错误一旦被纳入疾病脚本，其纠正难度高于单纯遗忘。

2.4 自动化偏差（automation bias）

自动化偏差指的是，人倾向于接受自动化系统的建议，即使它与自己的判断不一致[8,9]。放射科、内镜和病理的读片实验都观察到类似现象：错误的 AI 建议会把读片者引向错误的分级、诊断或随访判断，低年资医师受到的影响更大[3,11-13]。可解释 AI 能够部分缓解这种偏差，但并不能将其消除[13]。将其归因于“态度不端正”无法解释这一现象；它属于人机协作中的系统性偏差，仅靠提醒“要批判”并不够。

2.5 认知卸载（cognitive offloading）

认知卸载是把本该自己承担的认知任务交给外部工具[14]。认知卸载本身并非负面。检索、排版、记忆这些低阶工作交给 AI，可以释放工作记忆，让学习者把精力放在高阶推理上。关键在于卸载的对象与时机：如果连鉴别诊断、证据评价、不确定性判断都一起交出去，认知卸载就变成了认知外包，学习所需的生成性加工被绕过[15,16]。Wang 等[17]范围综述了 123 项研究，把生成式 AI 的使用分成“支持能动性”和“侵蚀能动性”两类：前者伴随自我调节学习、反馈素养和反思性投入，后者伴随过度依赖、不加批判地接受和判断力弱化。工具本身不是关键，教学如何嵌入才是关键。

2.6 伪胜任（pseudo-competence）

伪胜任是一种“看起来会、实际上不会”的错觉。

AI 输出流畅、易读[18-20]，学习者容易高估自己的理解程度。生成式 AI 把信息处理的表面难度降下来，读懂一段解释，很容易被当成掌握知识。初学者尤其危险：既缺少识别错误的知识基础，也缺少判断“自己还不会什么”的元认知能力[20,21]。这也是为什么“让学生自己讲一遍”仍然管用：讲的时候，理解缺口藏不住；读 AI 的解释，缺口可以被流畅的文字盖过去。表 1 把几组概念放在一起对照。

表 1　人工智能时代医学诊断能力退化的相关概念辨析

概念	核心含义	临床学习中的典型表现	教学启示
去技能化（deskilling）	长期依赖自动化，原本会的能力慢慢生疏	长期只做 AI 初筛后的复核，独立问诊、查体和鉴别推理都生疏	保留无 AI 诊疗量，定期安排撤除 AI 的复测
从未习得（never-skilling）	能力还没成形就被 AI 接管，独立诊断图式没搭起来	先看 AI 结论，再说不出病史、体征或征象依据	先把 AI-free 基线建起来，AI 只作课后对照
错误技能化（mis-skilling）	把 AI 的错误当成知识，写进自己的诊断路径	把 AI 的错误解释、错误随访建议写进病历或报告	做对抗性错误识别训练，逐条核验证据
自动化偏差（automation bias）	倾向于跟着自动化建议走，哪怕和自己的判断不一致	错误的 AI 建议改变了分级、诊断或随访判断	训练信任校准，记录置信度，必要时用可解释 AI 辅助
认知卸载（cognitive offloading）	把本该自己做的认知任务交给外部工具	让大语言模型直接生成鉴别诊断，自主推理被跳过	分清低阶卸载和核心外包，保留过程记录
伪胜任（pseudo-competence）	因为 AI 说得流畅，高估了自己的掌握程度	读懂了 AI 的解释，就以为掌握了临床要点	安排无 AI 考核，配合反思性写作和元认知训练

3 医学诊断基本功的三层结构

把“诊断基本功”当作一个整体讨论，容易得出“AI 有害”或“AI 无害”的笼统结论。更可行的做法是把它拆成三层：感知、知识、推理。三层能力对 AI 的敏感度不同，教学防线也应不同。放射科是下文的主要示例，但每一层在其他诊断学科里都能找到对应物。

3.1 感知层：看得见、听得见、摸得到

感知层解决“看得见、听得见、摸得到”的问题：问诊是否系统，查体手法是否规范，影像和病理切片能否读出关键信息，正常与异常、伪影与病变能否区分。放射科的视觉搜索有相对固定的顺序和覆盖面，漏诊常与搜索不完整、满足于首个发现有关[22]。胸部 CT 只关注肺内结节，纵隔淋巴结、胸膜增厚或骨破坏就可能被遗漏。更隐蔽的问题是“搜索终止”：AI 标出一个明显病灶后，学习者可能停止继续搜索，遗漏第二、第三个发现。内科问诊遗漏关键病史、查体遗漏阳性体征，机制有相似之处。判别式 AI 与计算机辅助检测（computer-aided detection, CAD）能直接标注病灶，短期内可提高检出率；但如果学习者从第一天起就依赖 AI 标注，独立搜索的练习量减少，其搜索范围可能被 AI 标注所限制。感知层训练不能依靠背诵完成，需要重复、即时反馈和纠错。AI 可以提示，但不能替代学习者建立搜索和查体的图式。

3.2 知识层：认得出

知识层解决“认得出”的问题：病史、体征、检验和影像能不能对上，疾病脚本熟不熟，典型和非典型表现能不能串起来，鉴别诊断框架有没有。AI 可以把机制讲清楚，也能按年级生成分层材料，这是它的长处。但知识牢不牢，取决于学习者有没有主动提取、有没有做生成性加工[15]。只看 AI 的解释，不用自己的话复述，不把临床表现和病理机制挂钩，不在相似病例里迁移，知识就容易停在“熟悉感”层面，真到病人面前调不出来。认知负荷理论讲得明白：材料设计要减少无关负荷、促进图式建构；AI 把答案直接端上来，图式建构所需的认知投入就少了，“必要难度”也被削掉[15]。

3.3 推理层：想得到、说得出

推理层解决“想得到、说得出”的问题：把病史、查体、检验和影像放在一起，形成假设，比较证据，识别不确定性，写出能用的病历或报告，再和患者、家属及同事沟通。临床诊断不能停在“看到什么说什么”，得把发现放回病人的背景里检验。大语言模型能列鉴别诊断，但一项随机对照试验发现，医师用了大语言模型，诊断推理评分没有显著提高（76% vs 74%），大语言模型单独使用的得分反而高于医师组[10]。答案质量和学习质量，不能画等号。教学只盯最终答案，推理过程就可能被跳过。病历和报告书写也是推理的一部分：发现、诊断、鉴别、下一步建议，得由学习者自己组织成临床能用的文本，而不是复制 AI 的草稿[23,24]。同样

是“胰腺占位”，病历要写清疼痛和体重变化、查体发现、实验室指标、影像上胰胆管扩张的程度和血管受累范围；这些判断，一句 AI 结论替代不了。图 1 把三层能力和 AI 的影响路径画在一起。

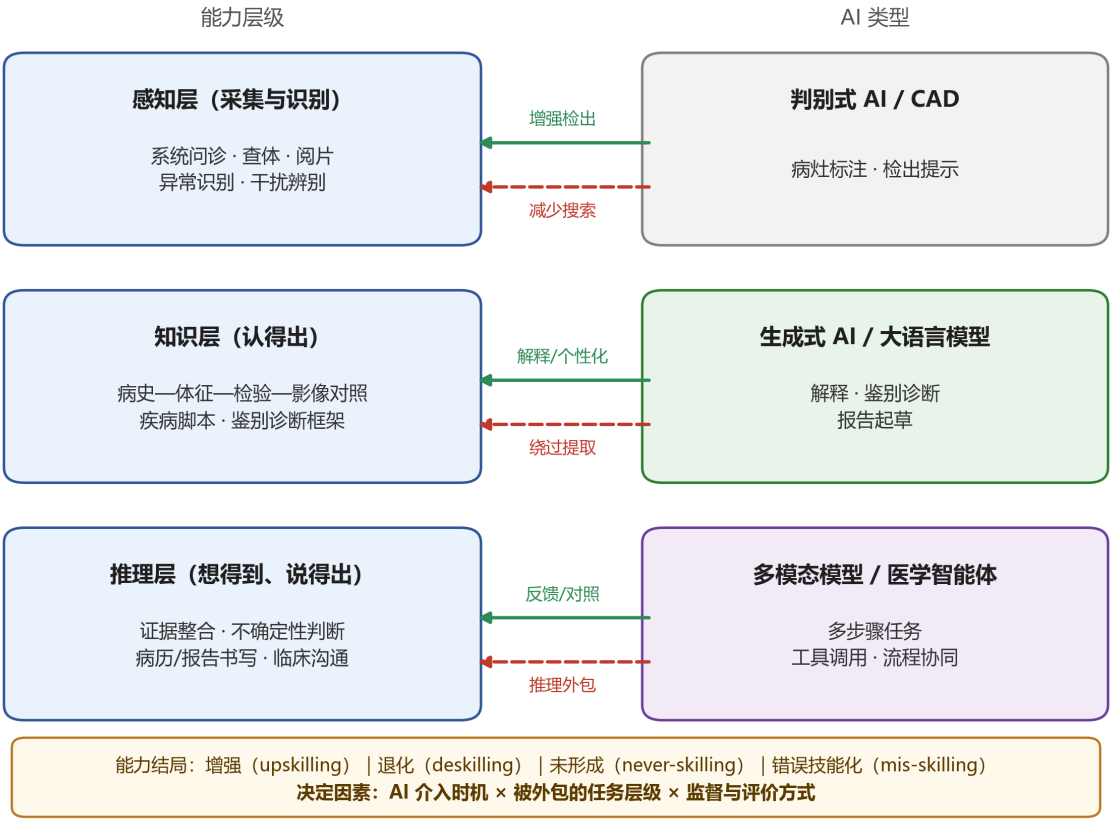


图 1　医学诊断基本功的三层结构与人工智能影响路径

4 AI 影响诊断能力的机制

AI 对三层能力的影响并不均匀。可以把机制归纳为五条路径，并用“时机-任务-监督”三个维度理解其边界。

4.1 认知卸载与生成性加工缺失

生成式 AI 把“检索-组织-表达”这一串认知劳动压缩了。学习者拿到的是已经组织好的答案，自己组织答案的过程被省略[14,15]。认知心理学的基本判断是：学习靠生成性加工——用自己的话重组，和已有知识建立联系，在提取和解释中把记忆结构扎牢[15]。这一步被 AI 替代，流畅感有了，可迁移的理解未必有[16,19]。所谓“effortless excellence”，说的就是这种表面效率提高、深层学习减少的状态。临床诊断最典型的例子，是让大语言模型直接生成鉴别诊断：列表看着完整，但学习者没有经历“为什么把这个病放前面、哪条证据支持、哪条证据反对”的推理。

4.2 自动化偏差与信任校准

自动化偏差的关键在另一处：AI 出错时，人能不能发现并纠正。低年资学习者对 AI 建议更敏感：知识基础不足，独立判断难；AI 的输出又往往比自己的判断更完整、更自信[11,12]。可解释 AI 能缓解一部分，替不了独立判断[13]。信任校准因此应当成为明确的教学目标。学习者要知道 AI 在哪些任务上可靠、在哪些情境下容易出错、自己的判断什么时候该优先。靠一门“AI 原理课”解决不了，得在具体病例里反复练。

4.3 反馈依赖与元认知错位

AI 给即时反馈，本来是好事。可如果反馈总是 AI 给的，学习者就少了自我监控的机会。元认知包括知道自己会什么、不会什么、什么时候该求助。伪胜任正是元认知错位的表现：输出越流畅，学习者越容易以为自己已经掌握[18-20]。一项针对中国大学生的研究发现，AI 依赖程度与批判性思维得分呈负相关，认知疲劳起部分中介作用；信息素养能缓冲这种负面影响，但在高依赖条件下也可能放大认知疲劳[25]。另一项研究编制了中国大学生的 AI 认知卸载量表，给后续实证研究提供了测量工具[26]。这些研究的场景不在医学院，但提示

了“依赖-疲劳-思维弱化”这条可能的路径。

4.4 练习机会减少与隐性课程

AI 进入工作流后，独立诊断的机会可能减少。AI 先筛一遍，学习者只复核阳性或疑难病例；病历先由 AI 起草，学习者只做修改。效率看似提高，但学习者接触“从零开始”的诊断任务的机会减少了。隐性课程也在变化。带教老师习惯先看 AI 结论，学生学到的就是“答案优先”的工作方式[27,28]。这种隐性课程比正式课程更具渗透力：学生模仿的是教师的工作方式；教师口头强调的规范，往往难以抵消这种示范。一项针对放射科住院医师的 AI 辅助训练研究显示，自由选择时多数病例会请求 AI 帮助；作者提醒，独立诊断能力需要持续监测[29]。AI 可以是有效的训练辅助，但“使用频率高”本身并不能作为教学成功的证据。

4.5 错误技能化

AI 输出错误，学习者又识别不了，错误就可能写进病历、带进下一次诊断，慢慢固化成“个人经验”。与单纯遗忘不同，错误技能化会形成错误的疾病脚本，纠正成本更高[1,3]。在临床诊断中，这类错误可能表现为把伪影当病变、把良性征象过度诊断、漏掉次要但重要的发现、在病历里给出不恰当的随访建议。AI 的错误往往“看起来合理”，这让错误技能化比传统的“不会”更隐蔽。

4.6 “时机-任务-监督”三维分析

把上述机制放在一起，可以得到一个简化的分析框架。先看 AI 介入的时机：是在学习者建立独立基线之前，还是之后。再看被外包的任务层级：是检索、排版这类低阶任务，还是问诊、查体、阅片、推理、决策这类核心任务。最后看监督程度：无监督使用、引导下使用，还是教师监督下的协同实践。三个变量共同决定学习结果是增强（upskilling）、退化（deskilling）、未形成（never-skilling），还是错误技能化（mis-skilling）。同一项 AI 工具在不同研究中结论不同，原因往往不在工具，而在教学嵌入方式。

5 证据现状与证据边界

围绕“AI 是否导致诊断能力退化”的证据可以分为三类：读片者或诊断者实验、各专科教学研究、认知与教育心理学的间接证据。三类证据的说服力不同，需要分开评估。放射科的研究最多，其他专科的证据相对分散。

5.1 读片者与诊断者实验

最直接的证据来自读片者实验。乳腺钼靶研究纳入 27 名不同年资的放射科医师，错误的 AI 乳腺影像报告和数据系统（Breast Imaging Reporting and Data System, BI-RADS）建议使读片正确率由约 80%下降至 20%~45%，低年资医师偏差更大[11]。颅内动脉瘤时间飞跃法磁共振血管成像（time-of-flight magnetic resonance angiography, TOF-MRA）研究显示，AI 假阳性会明显提高读片者的动脉瘤怀疑度和随访推荐强度，同时缩短阅读时间[12]。可解释 AI 研究进一步发现，AI 辅助乳腺读片时，自动化偏差和锚定偏差分别达到 36.1%和 33.9%；显著性热图将这两个数字降至 17.8%和 17.2%，但并未归零[13]。这些实验说明，AI 建议会系统性地改变人的判断，低年资者更容易受到影响。内镜和病理的证据也指向同一方向：错误的 AI 建议会使部分读片者推翻原本正确的判断[3]。

5.2 大语言模型与诊断推理

一项单盲随机对照试验将 50 名医师随机分为“大语言模型+常规资源”和“仅常规资源”两组。两组诊断推理评分无显著差异（76% vs 74%）；大语言模型单独使用时，得分反而比常规资源组高 16 个百分点[10]。这一结果提示了另一层问题：模型能够给出高质量答案，但拥有 AI 工具并不等于人的推理能力提高。如果教学只是让学生用 AI 寻找答案，得到的多半是答案，能力未必跟上。另一项针对中国放射科住培结业考试的研究显示，不同 AI 模型在不同条件下的答题表现存在差异，提示“AI 会做题”和“AI 能教学”是两件事[30]。

5.3 各专科教学研究

教学研究更接近课堂。放射科一项纳入 8 名住院医师、150 张胸片的研究比较了无 AI、按需 AI 和整合 AI 三种模式：AI 支持降低了严重程度评分的错误率，将评分者一致性提高 22%；住院医师对 AI 错误具有一定抵抗力，但自由选择时 70%的病例会请求 AI 帮助[29]。国内一项头颈 CT 血管成像（CT angiography, CTA）研究显示，AI 辅助阅片能够缩短住院医师的阅片时间、提高诊断准确率；作者同时指出，完全依赖 AI 的住院医师在脱离 AI 后难以独立作出准确诊断，AI 对伪影和血管外病变也存在盲区[31]。另一项本科生医学影像实习研究显示，AI 组的理论成绩和教学满意度优于对照组，但出科考核成绩无显著差异；作者建议先完成基础征象识别和系统阅片训练，再引入 AI 对照学习[32]。这些研究集中在放射科，内科、外科、急诊的同类研究还不多。它们多以短期准确率和满意度为结局，很少设置“训练结束后撤除 AI”的复测。这些研究能够证明 AI 作为教学工具的有效性，但不能证明学习者已经形成独立诊断能力。AI 的短期收益是真实的，但

短期收益并不能等同于长期能力。

5.4 认知与教育心理学的间接证据

认知卸载研究指出，外部工具能降低认知负荷，也可能减少学习者对材料的深度加工[14,15]。生成式 AI 与学习者能动性的范围综述发现，AI 的使用模式可以分成“支持能动性”和“侵蚀能动性”两类，后者与过度依赖、不加批判地接受和判断力弱化相关[17]。中国大学生样本的研究显示，AI 依赖与批判性思维呈负相关，认知疲劳起部分中介作用[25]。这些证据不能直接证明临床诊断能力退化，但为“机制存在”提供了旁证。

5.5 证据边界

现有证据还不足以支持“AI 必然导致医学诊断能力退化”这一结论。多数研究为短期、单中心、小样本，随访时间不足以观察能力形成或退化的长期过程；结局指标多为诊断准确率、阅读时间、满意度等替代终点，缺少“撤除 AI 后的独立表现”这一关键指标。AI 工具异质性较大：判别式 CAD、生成式大语言模型、多模态智能体的作用机制并不相同，不能一概而论。专科差异同样明显：放射科的证据最多，内科、外科、急诊、病理和内镜的研究相对零散。“never-skilling”目前更多是理论模型和间接证据，针对医学培训的纵向直接证据仍然缺乏[1]。因此，本文的立场可以概括为“风险真实、机制合理、证据有限、需要设防”。将 AI 描述为“有害”或主张“禁用”，均非本文本意。表 2 汇总了主要证据。

表 2　人工智能影响医学诊断能力的主要证据

研究（年份，对象）	设计/AI 类型	主要发现	教学启示
Dratsch 等（2023，27 名放射科医师）	前瞻性读片者实验；AI BI-RADS 建议	错误的 AI 建议使读片正确率由约 80%下降至 20%~45%，低年资医师偏差较大	自动化偏差是系统性问题，需要训练信任校准
Kim 等（2025，9 名医师）	自身对照读片实验；TOF-MRA 动脉瘤 AI	AI 假阳性抬高了动脉瘤怀疑度和随访推荐强度，阅读时间反而变短	错误 AI 建议会改变临床决策
Pesapane 等（2026，6 名乳腺放射科医师）	全交叉模拟读片；AI±可解释 AI	自动化偏差 36.1%、锚定偏差 33.9%；加显著性热图后降到 17.8%和 17.2%	可解释 AI 有帮助，但替代不了独立判断
Goh 等（2024，50 名医师）	单盲随机对照试验；大语言模型辅助	两组诊断推理评分无显著差异（76% vs 74%），大语言模型单独使用反而高 16 个百分点	“有 AI”不等于“人变强”
Savardi 等（2025，8 名住院医师）	三场景对照；胸片 AI	AI 减少了评分错误，把一致性提高 22%；自由选择时 70%的病例会叫 AI 帮忙	短期增强；独立能力仍需监测
陈素月等（2025，头颈 CTA 住培）	AI 辅助阅片教学	诊断准确率提高，阅片时间缩短；脱离 AI 后独立诊断能力不足	先独立后 AI，保留无 AI 训练
张烁琪等（2026，本科生影像实习）	AI 组 vs 对照组	理论成绩和满意度更好，出科考核没有显著差异	分阶段引入，重点评价独立能力
Tian 等（2025，580 名中国大学生）	横断面研究；AI 依赖与批判性思维	AI 依赖和批判性思维呈负相关，认知疲劳起中介作用，信息素养有调节	AI 素养与认知负荷管理
Wang 等（2026，123 项研究）	范围综述；生成式 AI 与学习者能动性	一类用法支持能动性，另一类在侵蚀能动性	教学嵌入方式决定结果
Heudel 等（2026，跨专科范围综述）	证据汇总；内镜、病理、影像	内镜撤除 AI 后腺瘤检出率从 28.4%降到 22.4%；病理读片中超过 30%的参与者在错误 AI 建议下推翻正确诊断	退化与自动化偏差并非放射科独有，需要跨专科监测
Ke 等（2026）	观点/理论框架	提出 never-skilling，并给出三阶段能力保护框架	本文框架的重要来源

6 教学防线：三阶段-双轨评价

如果风险取决于时机、任务和监督，教学干预也得围着这三个变量设计。本文把教学过程分成三段：基础期先不碰 AI，整合期在引导下用 AI，协同期在监督下用 AI。三段之外，再用“无 AI 基线+AI 辅助表现”的双轨评价守住底线。这套框架要给 AI 划定位置，禁止 AI 并不是它的出发点。

6.1 基础期：建立 AI-free 基线

基础期并非完全隔离 AI，但要求学习者先完成独立分析，AI 仅作课后对照。这样安排，是为了让学习者在接触 AI 之前，先建立独立诊断能力。系统采集训练应遵循固定顺序：内科从主诉、现病史、既往史到查

体，放射科从肺窗到纵隔窗，再到胸膜、骨骼和上腹部，逐例训练，要求口头或书面记录每个区域的发现，减少因满足于首个发现而造成的漏诊[22]。发现描述和疾病脚本也需要训练：用自己的话描述症状、体征或征象，解释病理机制，列出鉴别诊断，进行间隔提取，而非阅读 AI 生成的解释[15,23,24]。刻意练习离不开针对性反馈，病例难度应递进，独立诊断量应得到保证[33]。基础期还可设置“无 AI 诊室”或“无 AI 病例讨论”，AI 仅在学习者完成独立分析后用于对照，不在开始前给出答案。训练时长可以灵活安排，但顺序不能颠倒：先独立，后 AI。顺序本身即是能力形成的一部分。操作层面，可为不同年级设定“最低独立接诊量”和“必读病例清单”，将无 AI 训练纳入轮转计划，而不宜完全留给学生自发安排。

6.2 整合期：引导下的 AI 使用

整合期的目标，是让学习者学会与 AI 协同工作，而不是把诊断任务交给 AI。可采用六步流程：先在无 AI 条件下完成初步问诊、查体和阅片；记录自己的诊断假设、置信度和不确定点；再打开 AI，比较 AI 输出与自己的判断；对 AI 的每条建议进行证据核验，回到病史、体征、检验、影像和文献；说明采纳或拒绝的理由，完成病历修订；最后由教师点评推理过程和修订质量。由此，AI 从“答案提供者”转变为“对照工具”和“错误探测器”。还可设计对抗性 AI 错误识别训练：在 AI 输出中故意植入错误，如错误定位、错误解释或错误随访建议，要求学习者识别并纠正。研究显示，低年资者更容易受到错误 AI 建议影响[11,12]，这类训练应尽早开展并反复实施。可解释 AI 能够提供辅助，但不能替代独立判断[13]。

6.3 协同期：监督下的 AI 增强实践

协同期置于真实工作流中，目标是形成“AI 监督能力”。AI 可参与病例分诊、病历初稿、指南检索和随访建议，但需满足三个条件：学习者已经具备独立基线；AI 的使用有明确的任务边界和权限；教师或上级医师对最终病历或报告负责。评价重点也随之改变：能否得到正确答案退居其次，能否发现 AI 的错误、解释自己的判断、对 AI 输出承担监督责任，成为新的核心。这与住培的胜任力要求相一致[34,35]。具体做法包括：在病历中记录 AI 是否参与、参与了哪个环节、自己完成了哪些核验；对急腹症、肿瘤分期、危急值等高风险病例，复核流程应更严格；在多学科团队(Multidisciplinary Team, MDT)讨论中，也应说明 AI 建议的采纳依据。

6.4 双轨评价：测量“没有 AI 时”的能力

如果评价仅依据 AI 辅助后的最终病历，AI 的能力可能被误认为学习者的能力。双轨评价应同时考察两个方面。无 AI 基线考核可置于客观结构化临床考试（objective structured clinical examination, OSCE）或出科考核中，设置“无 AI 接诊站”“无 AI 阅片站”和“口述推理站”，评价学习者的独立问诊、查体、阅片、鉴别诊断和病历质量。AI 辅助表现考核则评价学习者对 AI 建议的采纳或拒绝是否合理、错误识别率、修订质量、证据核验过程，以及对不确定性的表达。AI 撤除测试同样重要：在 AI 辅助训练一段时间后安排一次无 AI 复测，观察独立表现是否下降；这是区分“增强”与“退化”的关键指标。纵向过程档案记录独立接诊量、AI 使用频率、错误类型变化、反思记录和教师反馈，形成能力轨迹。评价工具可借鉴脚本一致性测试评价临床推理，并结合结构化病历评分、口试和反思性写作[36]。基本原则是：凡涉及学习者能力评价，必须以独立表现为基础；AI 辅助表现作为补充，而非替代。

6.5 教师与制度保障

教学防线能否落地，取决于教师与制度。教师应具备基本的 AI 素养：了解 AI 的能力边界、常见错误、数据安全和伦理要求，能够设计“AI 参与但不替代学习”的教学活动，并能够评价学习者的 AI 监督能力[37,38]。制度层面需要明确：哪些考核必须无 AI，哪些课程允许 AI 辅助，学生使用 AI 时如何声明，患者数据如何脱敏，AI 生成内容能否直接进入病历。缺乏这些制度，课堂上的“先独立后 AI”容易被工作流的效率压力抵消。Khashoggi 等[39]的系统综述与国内研究[40,41]均指出，AI 教学面临模型可解释性不足、数据偏倚、内容可靠性、隐私风险和评价体系缺乏等挑战，这些问题同样会落到本文提出的防线上。图 2 总结了“三阶段-双轨评价”框架，表 3 列出了各阶段的教学策略与评价指标。

7 讨论

7.1 先谈能力，再谈工具

本文不反对 AI。病例生成、即时反馈、个性化练习、教学资源建设，AI 都能帮上忙[42,43]。真正的问题落在用法上。怎么用、什么时候用、用来做什么。把 AI 放在“第二读者”的位置，它是学习者的对照工具和错误探测器；放在“第一读者”的位置，它替代的恰好是学习者最需要练的那部分认知过程。

7.2 边界条件与学科差异

AI 对诊断能力的影响，受很多条件调节：AI 本身

靠不靠谱，任务属于感知、推理还是文书，学习者在本科、住培还是继续教育阶段，监督强度多大，评价导向是什么，学科的工作流又是什么样。放射科常被比作“煤矿里的金丝雀”。影像数据数字化程度高、AI 渗透早、报告结构清晰、反馈相对客观，所以问题先在这里暴露。但“时机-任务-监督”这套框架并不只适用于放射科。内科的鉴别诊断、外科的术前评估、急诊的快速决策、病理和内镜的判读，都可以沿着同一逻辑走，只是任务分层和评价指标要重新设计。内科更依赖病史和病程演变，AI 对“信息收集”的影响可能大于对“图像判读”的影响；外科的术前评估更依赖多模态影像和临床决策，AI 监督能力的评价指标也应不同；急诊更看重速度和风险分层，AI 建议的时效性和危险错误识别需要单独设计。

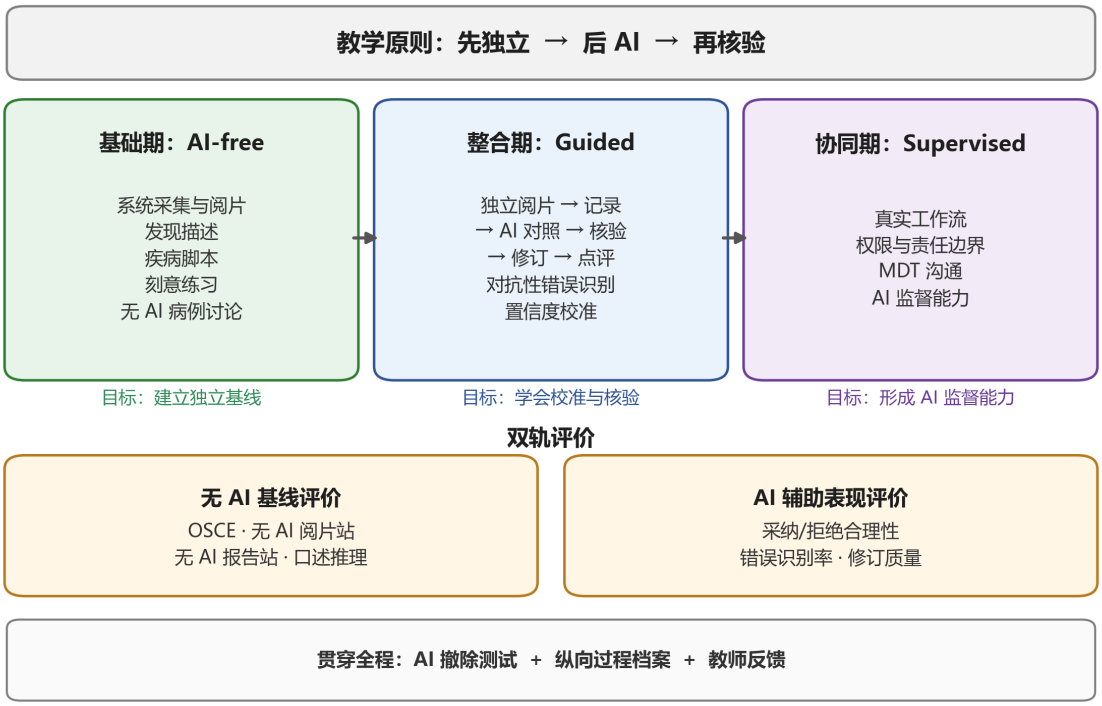


图 2　基础-整合-协同三阶段教学防线与双轨评价

表 3　三阶段教学防线与评价指标

阶段	教学目标	关键策略	评价指标
基础期（AI-free）	建立独立诊断基线	按固定顺序问诊、查体和阅片，练发现描述和疾病脚本，做刻意练习，开无 AI 病例讨论	无 AI 条件下的诊断准确率、发现描述完整性、鉴别诊断质量、口述推理
整合期（Guided）	学会校准与核验	独立阅片、记录判断、AI 对照、证据核验、修订、教师点评，再加对抗性错误识别	AI 建议采纳或拒绝是否合理、错误识别率、修订质量、置信度校准
协同期（Supervised）	形成 AI 监督能力	放进真实工作流，划清权限和责任边界，练 MDT 沟通，做 AI 使用声明	独立表现和 AI 辅助表现一起看，重点看危险错误识别、报告质量和沟通能力
贯穿全程	纵向监测能力轨迹	安排 AI 撤除测试，建过程档案，保留教师反馈	撤除 AI 后的独立表现、能力变化轨迹、反思质量

7.3 从“防退化”到“建能力”

换个说法，AI 时代的教育目标，应当从“防止退化”转向建设新的能力结构。独立诊断的基线能力要有，对 AI 输出的监督能力要有，证据核验、不确定性沟通、AI 失效时仍能安全工作，这些也都要有。AI 素养不止于会写提示词。它要求理解 AI 的能力边界、识别错误、校准信任、保护数据安全，并能把 AI 的输出放回临床证据链里检验。未来医师的核心竞争力，已经从“记住多少”挪到了别处。AI 在场时能否保持独立判断，AI 缺席时能否安全诊断。

7.4 研究议程

接下来需要的研究，至少有几类。纵向队列研究要追踪学习者从医学生到住培的诊断能力轨迹，看 AI 使用模式和能力变化之间是什么关系。AI 撤除随机对照试验要比较“先独立后 AI”和“先 AI 后独立”两种教学顺序的长期效果。过程性评价工具需要开发和效度验证，比如 AI 建议采纳或拒绝合理性评分、错误识别测验、脚本一致性测试。多中心、多学科验证也不能少，要看“三阶段-双轨”框架在放射科、内科、外科、急诊等不同专科是否站得住。中国语境下的研究同样重要：国产大模型、中文病例库、住培制度、医疗数据安全政策，都会影响教学。

7.5 局限性

本文是叙述性综述，没有做系统检索和定量合并，文献选择有主观性。证据以英文文献为主，中文文献多为短期教学研究。放射科作为主要示例，证据最集中；内科、外科、急诊、病理和内镜的文献相对零散，跨专科的外推需要谨慎。“never-skilling”等概念还在发展，本文的框架属于理论整合，尚需实证检验。

8 结语

AI 可以当医师的“第二读者”，但不该当“第一读者”。AI 进入诊室和阅片室的时间越来越早、程度越来越深，医学诊断教育得守住一条底线：离开 AI，医师仍然能独立、安全地完成诊断。这条底线不靠禁止 AI 实现，靠的是“先独立、后 AI、再核验”的教学顺序，靠基础期先把 AI 放一边、整合期带着 AI 练、协同期在监督下用 AI 的三段安排，靠无 AI 基线与 AI 辅助表现并行的双轨评价。技术会继续进步；诊断基本功、判断力和对患者负责的意识，始终是医师不能交出去的东西。

参考文献

[1] KE Y, JIN L, ONG J C L, et al. AI-induced never-skilling in medical education[J]. Nat Med, 2026, 32(6): 1997-2006.

[2] AL-ANEZI F M. Generative artificial intelligence in healthcare: automation bias, deskilling, and cognitive implications—a systematic review[J]. J Healthc Leadersh, 2026, 18: 590498.

[3] HEUDEL P E, CROCHET H, FILORI Q, et al. Artificial intelligence in medicine: a scoping review of the risk of deskilling and loss of expertise among physicians[J]. ESMO Real World Data Digit Oncol, 2026, 12: 100693.

[4] 付淳曦. 血液内科人工智能应用:临床价值与医学教育中的“技能退化”困境[J/OL]. 四川生理科学杂志, 2026 [2026-11-17]. http://scslkxzz.aiencm.com/scslkxzz/Article/abstract/D2026062001.

[5] 王欣宜, 朱俊勇, 吴彦瑞, 等. 人工智能在全科医学教育中的应用现状和挑战及其应对策略[J]. 中华医学教育杂志, 2026, 46(10): 761-764.

[6] 程伟. 人工智能推动医学教育变革的现实与可能[J]. 医学与哲学, 2026, 47(15): 12-16.

[7] 于浩鹏, 王颖, 李雨, 等. 生成式人工智能在医学实践教育中的思政融入与实践应用[J]. 国际教育学, 2026, 8(5):1-9.

[8] PARASURAMAN R, RILEY V. Humans and automation: use, misuse, disuse, abuse[J]. Hum Factors, 1997, 39(2): 230-253.

[9] MOSIER K L, SKITKA L J, HEERS S, et al. Automation bias: decision making and performance in high-tech cockpits[J]. Int J Aviat Psychol, 1998, 8(1): 47-63.

[10] GOH E, GALLO R, HOM J, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial[J]. JAMA Netw Open, 2024, 7(10): e2440969.

[11] DRATSCH T, CHEN X, REZAZADE MEHRIZI M, et al. Automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance[J]. Radiology, 2023, 307(4): e222176.

[12] KIM S H, SCHRAMM S, RIEDEL E O, et al. Automation bias in AI-assisted detection of cerebral aneurysms on time-of-flight MR angiography[J]. Radiol Med, 2025, 130(4): 555-566.

[13] PESAPANE F, LATRONICO A, ABBATE F, et al. Evaluating cognitive biases in AI-assisted mammography interpretation: a simulation reader study of explainable AI across radiologist experience levels[J]. Eur Radiol, 2026, 36(10): 7639-7650.

[14] RISKO E F, GILBERT S J. Cognitive offloading[J]. Trends Cogn Sci, 2016, 20(9): 676-688.

[15] SWELLER J, VAN MERRIËNBOER J J G, PAAS F. Cognitive architecture and instructional design: 20 years later[J]. Educ Psychol Rev, 2019, 31(2): 261-292.

[16] HADIDCHI R, PATEL V J, CHUNG C, et al. The perils of “effortless excellence”: generative AI and the erosion of

effort in medical education[J]. Med Sci Educ, 2026, 36(6): 2905-2909.

[17] WANG G, WANG W, YANG D, et al. Generative AI, cognitive offloading, and learner agency in higher education: a scoping review[J]. Behav Sci (Basel), 2026, 16(7): 1150.

[18] DING C, WENG S. Generative artificial intelligence-assisted learning and the rise of pseudo-competence in medical trainees[J]. Acad Med, 2026, 101(9): 1058-1059.

[19] NOEL Z R. Wielding magic without mastery: the illusion of competence in the age of AI[J]. Am J Pharm Educ, 2026, 90(8): 102030.

[20] IZQUIERDO-CONDOY J S, ARIAS-INTRIAGO M, TELLO-DE-LA-TORRE A, et al. Generative artificial intelligence in medical education: enhancing critical thinking or undermining cognitive autonomy?[J]. J Med Internet Res, 2025, 27: e76340.

[21] KRUGER J, DUNNING D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments[J]. J Pers Soc Psychol, 1999, 77(6): 1121-1134.

[22] BERBAUM K S, FRANKEN E A JR, DORFMAN D D, et al. Satisfaction of search in diagnostic radiology[J]. Invest Radiol, 1990, 25(2): 133-139.

[23] CROSKERRY P. The importance of cognitive errors in diagnosis and strategies to minimize them[J]. Acad Med, 2003, 78(8): 775-780.

[24] NORMAN G R, MONTEIRO S D, SHERBINO J, et al. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking[J]. Acad Med, 2017, 92(1): 23-30.

[25] TIAN J, ZHANG R. Learners' AI dependence and critical thinking: the psychological mechanism of fatigue and the social buffering role of AI literacy[J]. Acta Psychol (Amst), 2025, 260: 105725.

[26] CHANG H. Development and validation of the AI cognitive offloading scale for Chinese college students[J]. Front Psychol, 2026, 17: 1884676.

[27] BARTLETT E, SCHMIDT H, ALCAIDE LEON P, et al. Preserving the cognitive process: rethinking radiology education in the age of artificial intelligence[J]. Can Assoc Radiol J, 2026, 77(1): 29.

[28] WU K, GAUBE S, COLAK E, et al. Training radiologists for an AI-augmented future[J/OL]. Can Assoc Radiol J, 2026[2026-11-17]. DOI:10.1177/08465371261449694.

[29] SAVARDI M, SIGNORONI A, BENINI S, et al. Upskilling or deskilling? Measurable role of an AI-supported training for radiology residents: a lesson from the pandemic[J]. Insights Imaging, 2025, 16(1): 23.

[30] ZHU Z, ZHAO Y, LI L, et al. Artificial intelligence performance under different conditions in answering China's standardized training examination for resident physician in radiology: a comparative analysis[J]. Health Care Sci, 2026, 5(5): 436-447.

[31] 陈素月, 郭荣, 吉玉刚, 等. 头颈 CT 血管成像人工智能辅助阅片在住院医师规范化培训中的应用[J]. 江苏卫生事业管理, 2025, 36(1): 125-129.

[32] 张烁琪, 张顺, 刘城霞, 等. AI 辅助诊断工具在本科生医学影像学实习教学中的应用[J]. 医学新知, 2026, 36(10): 1190-1195.

[33] ERICSSON K A, KRAMPE R T, TESCH-RÖMER C. The role of deliberate practice in the acquisition of expert performance[J]. Psychol Rev, 1993, 100(3): 363-406.

[34] AGARWAL G, MAROO K, ZOMORODIAN P, et al. Radiology AI training and assessment—challenges, innovations, and a path forward[J]. BJR Open, 2025, 7(1): tzaf026.

[35] LOURENCO A P, SLANETZ P J, BAIRD G L. Rise of ChatGPT: it may be time to reassess how we teach and test radiology residents[J]. Radiology, 2023, 307(5): e231053.

[36] CHARLIN B, ROY L, BRAILOVSKY C, et al. The script concordance test: a tool to assess the reflective clinician[J]. Teach Learn Med, 2000, 12(4): 189-195.

[37] MELLO-THOMS C. Teaching artificial intelligence literacy: a challenge in the education of radiology residents[J]. Acad Radiol, 2023, 30(7): 1488-1490.

[38] HOUGH J, CULLEY N, ERGANIAN C, et al. Potential risks of GenAI on medical education[J]. BMJ Evid Based Med, 2025, 30(6): 406-408.

[39] KHASHOGGI K G, KHIZINDAR H T. Large language models in radiology education: a systematic review of applications and learning outcomes[J/OL]. Curr Probl Diagn Radiol, 2026 [2026-11-17].
DOI :10.1067/j.cpradiol.2026.10.016.

[40] 蔡秋钰, 容逍, 马朗, 等. 人工智能在医学影像学教学中

的应用与发展路径探讨[J]. 高校医学教学研究(电子版), 2026, 16(2): 8-13.

[41] 宋雨欣, 李子怡, 丁宁. 生成式人工智能对医学教育的影响探析[J]. 中国医学教育技术, 2024, 38(3): 281-286.

[42] PRABHU V, YOUNG M G, VERDONE A, et al. Artificial intelligence-enabled precision education: a novel tool to augment radiology residency training[J/OL]. Acad Radiol, 2026[2026-11-17]. DOI:10.1016/j.acra.2026.08.068.

[43] LYO S K, COOK T S. Integrating large language models into radiology education: an interpretation-centric framework for enhanced learning while supporting workflow[J]. J Am Coll Radiol, 2025, 22(11): 1271-1280.

版权声明：©2026 作者与开放获取期刊研究中心（OAJRC）所有。本文章按照知识共享署名许可条款发表。
http://creativecommons.org/licenses/by/4.0/

OPEN ACCESS