认知与评价:生成式 AI 教育应用的理性边界

简思宇

(湖北大学 历史文化学院,湖北 武汉 430062)

[摘要] 当前生成式 AI 的教育应用存在认知外包风险、理性边界模糊、伦理风险管控缺失等问题,与之相关的教育评价也有一定缺陷,因而需要厘清生成式 AI 教育应用的合理边界,应对技术滥用、评价异化等现实挑战。具体应从学生认知、教学评价双重视角分析生成式 AI 融入教育的深层风险与伦理矛盾,由于学生侧方面的认知水平、任务类型、使用方式与主导权透明度是区分良性人机协同与思维替代的核心边界,评价侧方面的技术僭越易于导致掩盖教育本质、引发评价不公与权责模糊等困境,亟待从认知建构、指标优化、算法迭代、人机互信、伦理规约等五个方面建构实践路径。

[摘要] 生成式 AI;教育应用;认知外包;教育评价;人机互信;伦理规约

Cognition and Evaluation: The Rational Boundaries of Generative AI Application in Education

JIAN Si-yu

(School of History and Culture, Hubei University, Hubei Wuhan 430062)

[Abstract] The current educational application of generative AI is confronted with prevalent problems including cognitive outsourcing risks, ambiguous rational boundaries and insufficient ethical risk governance. Meanwhile, relevant educational evaluation mechanisms also present obvious deficiencies. Therefore, it is necessary to clarify the reasonable application boundaries of generative AI in education to address practical challenges such as technology abuse and evaluation alienation. This paper analyzes the in-depth risks and ethical contradictions of generative AI integration in education from the dual perspectives of student cognition and teaching evaluation. On the student side, cognitive level, task type, usage mode and subject dominance transparency serve as the core boundaries to distinguish benign human-machine collaboration from thinking substitution. On the evaluation side, technological transcendence tends to obscure the essence of education and trigger evaluation
inequity and blurred rights and responsibilities. It is urgent to construct a practical optimization path from five dimensions, namely cognitive construction, index optimization, algorithm iteration, human-machine mutual trust and ethical regulation.

[Keywords] Generative AI; educational application;cognitive outsourcing; educational evaluation; human-machine mutual trust; ethical regulation

[作者简介]

简思宇(2001-),女,湖北大学历史文化学院 2025 级图书情报专业硕士研究生。

一、引言

伴随人工智能技术的快速迭代，生成式人工智能（Generative AI,GAI）实现突破性发展。在教育从数字化阶段迈入智能化的新阶段，“AI+教育”已从个别探索进入规模化应用阶段，生成式 AI 也普遍应用于教育领域，并从外部辅助工具逐步深度介入学习认知与教学评价。然而，这种深度融合在提升教育效率的同时，也暴露出诸多深层次问题。针对生成式 AI 融入教育场景的理性边界和伦理挑战，目前学界主要探讨了 AI 在教育应用中的学术诚信、认知依赖等风险[1]；过度外包的风险、发生机制[2]；AI 赋能教育评价产生的弊端，包括技术至上遮蔽教育评价本质、标准缺失误导评价方向、算法偏见限制评价效度、反馈缺失削弱评价价值、人机互信危机阻碍评价实施、技术僭越加剧评价风险等方面[3]；生成式 AI 的治理研究[4]。但既有成果对学生侧认知外包与评价侧技术僭越的分析相对独立，未能将两者有机结合，一定程度上割裂了相互之间的依存关系，并且缺乏规范性的边界界定。有鉴于此，本文拟从学生侧和评价侧两个维度展开探讨，客观审视生成式 AI 融入教育场景的理性边界与伦理约束，以驱动 AI 与教育的良性融合。

二、生成式 AI 教育应用的认知外包及其风险

认知外包本质上是人类将部分思维活动和认知任务，外包给外部智能设备或其他组织，以完成某个任务或达成某个目标的一种认知分工方式。这种方式旨在通过利用外部的智力和资源，弥补人类自身认知能力的不足，从而提高任务完成的效率和质量。

人工智能在教育中的应用认知外包，则是学习者将本应由自身完成的认知活动，转交给外部智能系统代为执行的过程。生成式人工智能所驱动的认知外包不再局限于“信息检索”或“简单计算”等低阶操作，而是具备了生成完整论证、构建逻辑结构、输出创造性方案的“高阶替代”能力[2]。这意味着，原本需要学生投入大量努力方能完成的理解、推演、判断与创作等核心思维活动，可以仅仅通过简单的自然语言指令即可交由人工智能代为产出。

（一）认知外包与过度认知外包的区分边界

认知外包是智能时代人类认知方式变革的必然趋势，并非必然导向负面后果。学生使用 AI 工具引发的过度认知外包问题，并非“用与不用”的二元对立问题，而是存在明确的条件与边界：合理的认知外包在突破边界条件后会滑向过度认知外包，进而对学生认知发展产生截然相反的影响。这种边界可以拆分为四个维度：

其一是学生的认知水平。认知力强的学生将 AI 视作助手，会对 AI 输出进行核查、质疑与修正，能进一步激发深度思考；认知力弱的学生将 AI 视为权威，直接采纳输出，导致思维被替代。现有研究表明，认知外包教育成效取决于内部认知网络与外部认知网络之间能否保持平衡：当学生具备足够完备的内部认知网络和相应的元认知监控能力时，外部认知工具可以成为认知能力的有效延伸；反之，当内部认知网络尚未建立、学生对外部工具缺乏监控与判断能力时，认知外包便可能滑向过度外包，触发思维幼稚化、认知地位边缘化和发展主体性丧失等教育陷阱[5]。

其二是任务类型的边界。将 AI 用于资料检索、语言润色等低阶机械性任务，此类外包属于低阶认知操作，对思维损害较小甚至能提升效率；若用于问题拆解、论证建构、逻辑推演等高阶认知任务的外包思维核心环节，势必引发高阶能力退化等风险。因此，人工智能介入课堂教学的认知边界时，在人机协同的学习链条中应当明确区分可交由人工智能辅助的低阶处理与必须保留的高阶思维—信息的检索、数据的整理以及初级文本的生成，可以适度交由人工智能承担；但问题重构、价值判断、逻辑推演、情感体验等涉及意义建构的深层认知活动，则必须保障学生的真实心智投入[6]。

其三是 AI 使用方式边界。大致有如下三种类型：①生成式使用，让 AI 直接产出答案最易造成认知外包；②检索式使用，将 AI 作为信息检索工具风险较低；③对话式使用，与 AI 辩论讨论等则可能反向激发思维。

其四是透明度与主导权边界。AI 以黑箱形式输出答案，学生完全照搬，属于过度外包；若 AI 披露推理过程、学生保留最终修改主导权，则可以称之为良性人机协同。由于 AI 的内容生成过程缺乏透明性和可解释性，犹如一个黑箱，故而这些错误信息难以被及时识别与精准纠正[8]。

由此可见，边界条件的核心价值，在于将讨论“该不该用 AI”的二分法，扩展为“什么条件下使用会损害、什

么条件下使用会提升”的条件分析，旨在为合理使用 AI 提供清晰的条件边界与可操作的判定标准。

（二）过度认知外包的风险表现

当认知外包突破合理边界、从辅助性工具使用滑向替代性思维让渡时，便会在认知、行为与发展等维度对学习者产生系统性的负面影响，其本质是突破合理边界后对学生认知体系与学习主体性的侵蚀。具体表现在下述方面：

首先是导致学生学习思维能力的退化。当学生在多次任务中体验到 AI 秒级生成带来的效率红利后，会逐步形成对 AI 的认知依赖。从认知卸载理论而言，AI 提供的平滑交互契合了人类信息加工的认知经济性原则，倾向于以最小的认知负荷完成任务的省力路径。这种路径一旦成为习惯，学生便会在面对学习任务时优先寻求 AI 的直接输出，而非启动自主思考[6]。并且，长期将复杂认知任务转交给外部智能系统处理，内部认知网络因缺乏必要的训练与激活而逐渐萎缩，从而导致思维懒惰与幼稚化，学习者丧失独立完成高阶认知活动的能力与意愿，思维能力停留在浅层的信息接收层面，难以形成深度的理解与批判性的判断。

其次是使学生丧失学习主体性。过度依赖 AI 会使学生从认知活动的主体建构者沦为 AI 输出的被动接收者，学习决策完全受制于 AI 的生成的结果，而非基于自身内在思考的自我判断，最终导致学习异化，丧失自主学习与自我发展的核心能力。正如研究者所指出，当外部智能系统成为认知的主导力量时，学习者在认知活动中从执行者蜕变为旁观者，其认知地位被边缘化，作为认知主体的自我并未获得相应的成长[5]。

最后是滋生伦理与认知偏见。学生倘若缺乏对 AI 输出结果的甄别与质疑意识，直接采纳 AI 生成的内容，不仅可能因 AI 的信息失真、逻辑漏洞引发学术诚信问题，还会被动接受 AI 隐含的算法偏见，固化自身认知局限，形成认知极端化等问题。

综合而论，上述风险表征共同指向的核心问题在于，AI 在教育中的介入一旦越过辅助边界、进入替代与主导性区域，不仅无益于教育目标的实现，而且会从根本上侵蚀学习者的认知能力、自主意识和成长潜力，效果适得其反，走向辅助学习的对立面。

三、生成式 AI 教育评价的应用及其路向缺陷

伴随人工智能的高速发展和普遍推广，生成式 AI 在教育领域的运用日渐普及和深入，并以之作为教育评价的重要方式。如果说认知外包风险仅限于学生层面的话，那么生成式 AI 教育评价则直接关系教育效果的评判，后者也是客观审视人工智能应用于教育领域不容忽视的因素。

（一）生成式 AI 在教学评价中的应用

人工智能正在深度介入教学评价的各个环节。从人工智能赋能教育评价的角度来看，AI 技术的深入应用正在推动教育评价由经验性评价向数字化评价转变、由单一性评价向综合性评价转变、由结果性评价向过程性评价转变、由诊断性评价向反馈性评价转变的四重转型[3]，其核心逻辑是用智能技术替代传统评价中依赖教师个人经验与有限精力的环节，实现对多源教育数据的全方位分析。

在上述趋势下，AIGC 检测工具的引入，成为 AI 融入教学评价标准最直观且具有争议性的表现。随着生成式 AI 在学生写作、作业及论文中的大规模应用，高校学术诚信面临前所未有的挑战，知网 AIGC 检测系统等工具被迅速引入教学评价体系，通过分析文本语言特征、句式结构与生成概率分布，输出量化的检测率指标，辅助判断文本中 AI 参与生成的比例。

实际上，AIGC 检测率最初被定位为教师判断学生学术诚信的参考信息，教师结合自身专业判断，将检测结果作为辅助依据使用。然而，在实际应用中 AIGC 检测指标的定位正在快速提升，权重也在不断提高：部分高校已将 AIGC 检测率设为论文答辩的前置门槛，检测率超过一定标准则判定为 AI 代写。简而言之，AIGC 检测率已经从参考信息升级为判定标准，从辅助教师判断走向替代教师判断，其实质是教育评价侧出现了技术僭越问题，AIGC 检测的参考辅助功能被核心评价功能所取代，教育评价的技术性角色被无限放大，本质则被淡化和遮蔽。这种偏离教育本质的评价结果，势必对实际的教学产生误导，挫伤教师和学生的主观能动性和创造性，限制教育效果的提升空间。

（二）AI 作为教学评价方式的路向性缺陷

客观而言，AI 融入教学评价在减轻繁琐任务、提升评价效率、推动评价转型方面，的确较之于传统方式有其积极的一面，但是其结果却是对教育本质的异化，根本缺陷即在于技术至上论，即将技术指标上升到首位，对教育自身的难以计算和衡量的核心要素置于次要地位。正如研究者所指出，AI 技术的引入使得教育评价的设计逻辑从以教育规律为参照，转向以数据可得性和算法效率为驱动，那些难以量化但在教育上至关重要的素养维度，比如学习动机的深层变化、思维品质的内在成长，则因不可计算而在评价中被系统性忽视[3]。这种评价路向性缺陷的根源在于技术至上僭越教育主体底线，当技术从评价的辅助工具升级为评价的主导逻辑时，教育评价便从服务于人的成长，异化成服务技术的效率[7]。

首先，将 AI 作为教学评价方式，也必然产生量化指标与教育质量的结构性错位问题，AIGC 检测是这一结构性错位最直观的体现。AIGC 检测率衡量的实际是文本与已知 AI 生成模式的相似度，并非论文本身的学术质量，但在实践中 AIGC 检测率却在目前被当作论文通过答辩的前置标准，而仅仅依靠高度简约的算法模型显然不足以还原教育活动的全部内涵，甚至会裂解教育的复杂性[8]。

其次，AI 赋能教学评价将导致算法偏见与评价公平的问题。AI 评价系统的训练数据与算法设计，都嵌入了特定的价值取向与文化假设。受一定条件约束形成的算法偏见，往往超越人本伦理防线的控制，而一旦带有偏见的算法被用于教育评价，其后果便不仅是技术层面的误差，而是对人本价值的系统性侵害[7]。由于 AI 的内容生成与判断过程缺乏透明性和可解释性，故在 AI 的教学评价结果中，学生面对的是一串不可追问、不可争辩的算法输出，难以理解评价结果的生成逻辑，缺乏判断的依据和过程，进而影响评价的公信力[8]。

最后，技术僭越风险会弱化教师的评价主体地位。教师如果从从评价的主导者沦为技术的执行者，便会丧失对评价过程的把控与对学生的个性化反馈能力，违背人机协同的核心逻辑。技术依赖冲破教育责任防火墙，或可被视为智能教育评价的终极伦理困境：当评价决策由算法而非教师做出时，为不公正评价后果担责的主体则由人沦为技术。这种责任的模糊化更意味着教育评价中最重要的伦理维度被技术中介所稀释[7]。

四、生成式 AI 教育应用的理性边界建构

生成式 AI 教育应用的理性边界，是区分良性人机协同与恶性技术僭越的判定标准，具体包括学生认知侧、教学评价侧双重维度的限定：学生认知侧基于人机分工规律、贴合学生认知发展规律的规范性框架，明确人机协同学习中 AI 辅助与 AI 替代的本质界限，从理论层面确立学生使用生成式 AI 的核心准则，系统建构学生侧合理认知外包的完整理性边界框架；教育评价侧在充分肯定 AI 在教学评价中的赋能价值基础上，强调全流程权责逻辑，重构人机评价分工范式，以改善智能评价技术僭越、标准异化的困境。

（一）生成式 AI 融入教育的理性边界的双重维度限定

1.学生认知侧：认知任务分工的人机边界

学生认知侧维度理性边界的确立重点，在于明确学生过度认知的外包标准，划分外包任务的认知范畴，形成可判定、可参照的认知分工准则，这是界定学生 AI 使用理性边界的规范依据。在学生日常学习的全过程链条中，辅助性、机械性认知任务可适度向生成式 AI 外包，主要包含信息检索、基础数据整理、文本语言润色、格式规范处理等重复性、事务性工作，此类任务无需深度思维参与，借助 AI 完成可有效降低学生认知负荷，实现人机协同增效，让学生将精力聚焦于高阶学习任务；而核心思维性、建构性认知任务需由学生自主完成，涵盖文本深度理解、问题拆解、逻辑推演、观点论证、价值判断等高阶认知环节，此类任务是学生知识内化、思维塑造、能力进阶的核心载体，是核心学习过程，也是坚持学生认知主体性的关键。

2.教学评价侧：评价权责的人机边界

明晰评价权责边界，落实全流环节的权责划分，构建人机分工的判定标准，是评价侧理性边界的核心。生成式 AI 可深度参与评价流程前端的机械性工作，但不可以作为最终的评价标准。在完整的教学评价链条中，AI 具备明确的赋能应用空间，可承担多源数据采集、学习行为模式识别、文本 AI 痕迹筛查、基础数据统计分析、即时性反馈建议生成等重复性、工具性、事务性的前端评价环节，依托算法优势弥补人工评价效率低、数据覆盖不全、时序追踪不足的短板，优化评价实施流程、降低教师评价负荷。但涉及教育本质与育人价值的最终学业判定、学生素养价值判断、学习成长意义解读、评价结果定级裁决等核心环节，需要由教师自主完成。从流程结构

上厘清“机器辅助评价”与“机器主导教育”的本质差异，为智能教育评价提供可参照的边界。

（二）生成式 AI 良性融入教育的理性边界建构

基于前文对学生侧和评价侧双重理性维度限定的分析，生成式 AI 教育应用的理性边界的设立，应重点关注认知建构、指标规范、算法设计、人机互信、伦理规约五个方面的内容，才能为 AI 融入教育场景提供规范合理的落地支撑，构建适配教育场景的实践路径，优化人机协同范式，实现生成式 AI 良性融入教育的效果。

1.强化学生认知

清晰的认知边界，取决于学生自身的认知素养与思维品质，重点在于学生的元认知监控与批判性判断能力，这是设立认知理性边界的主观核心条件。学生需构建完备的内部认知网络，形成结构化、体系化的知识框架，夯实思维基础，能够自主对接、甄别、整合 AI 输出的外部信息，避免沦为智能工具的附庸。

与之相应，实际教学中教师应引导学生建立层次化、结构化、网络化的知识体系，加强知识点之间的语义关联与逻辑联结，形成可迁移、可活用的知识组块；鼓励学生保持思维的批判性、深刻性与系统性，在人机认知交互中主动质疑、校验、修正 AI 生成内容，将外部智能信息有机融入自身认知体系，实现可控、有效的认知外包，而非被动接收现成答案；帮助学生依托元认知判断力自主权衡，清晰区分“借助 AI 辅助思考”与“依赖 AI 替代思考”的认知状态，主动把控认知外包尺度，在合理利用 AI 减负增效的同时，始终坚持自主高阶思维，跨越认知外包带来的教育陷阱，实现人机协同的良性学习形态。

2.完善评价指标体系

评价指标体系是 AI 赋能教学评价的基础支撑，针对现有 AI 评价体系重结果、轻过程，重量化、轻素养的问题，教师需立足教育场景的特殊性，构建层次清晰、贴合教学实际、适配学生发展的精细化评价指标体系。一方面，依托一线教育经验拆解复杂评价内容，将学生思维能力、学习态度、探究素养等隐性育人指标，转化为机器可感知、可建模、可分析的细粒度量化指标，建立教育评价内容与智能数据的精准表现关系；另一方面，兼顾过程性评价与结果性评价，覆盖学生学习全流程，突破传统 AI 评价仅筛查成果、忽视成长过程的局限，精准捕捉学生认知发展、思维成长的动态特征，为智能评价的标准化、规范化实施提供明确依据。

3.优化算法设计逻辑

算法偏见、数据偏差与黑箱问题，是导致 AI 评价效度不足、学生认知过度外包、人机互信缺失的技术诱因。为摆脱技术层面的困境，教师需从数据源头与算法架构两个层面双向发力：在数据层面，拓宽多源异构数据采集渠道，兼顾不同学段、不同基础、不同特质学生的学习数据，丰富数据维度、规避单一数据来源导致的评价片面性；在算法层面，破除固有算法偏见，公开算法运行逻辑，解决 AI 评价黑箱问题。同时引入师生、家长、管理者等多元主体参与算法调试，调整算法参数与决策逻辑，让算法适配教育育人需求，从技术角度捍卫教育的核心育人理念。

4.促进人机双向互信

人机互信危机是制约 AI 智能评价落地见效的瓶颈。鉴于机器评价的可解释性不足导致师生认可度偏低，学生智能素养不足引发过度认知外包两方面的问题，为此需要双向发力构建良性人机信任体系：技术端持续完善智能评价实践逻辑、算法偏见治理、评价逻辑公开等方式；主体端全面培育学生智能评价素养，引导其正确认知 AI 价值与局限，既善于借助智能工具减负增效，又具备独立研判、理性鉴别机器评价结果的能力。从持术端与主体端双向着手，是实现人类主导加机器赋能的人机协同评价新格局的基本途径。

5.完善伦理规约

AI 赋能教育兼具技术优势与伦理风险，数据隐私泄露、人机权责模糊、技术滥用僭越等问题，亟须完善的伦理制度予以约束，为此亟需构建全方位伦理治理机制：其一是严守数据隐私边界，健全数据安全防控体系，杜绝学生泄露隐私的现象；其二是明晰人机权责边界，巩固教师、学生的主体地位，明确 AI 的辅助工具定位，杜绝技术僭越、机器裁决的评价乱象；其三是健全多元监管机制，整合学校、教育部门、技术企业、师生主体的监控力量，对算法质量、数据安全、评价实施、结果应用实施全流程监管，明确伦理问责机制，让生成式 AI 始终在伦理框架与理性边界内向善发展、服务于育人目标。

五、结语

生成式 AI 与教育的深度融合是教育数字化转型的大势所趋，但其核心风险在于技术应用对教育理性边界的突破，具体表现为学生侧的过度认知外包和评价侧的技术僭越两大根本问题。为使生成式 AI 良性融入教育，避免出现过度认知外包的风险和教育评价的异化，必须明确学生认知侧认知任务分工的人机边界和教学评价侧的评价权责的人机边界，并从强化学生认知、完善评价指标体系、优化算法设计逻辑、促进人机双向互信、完善伦理规约方面着手，一定程度上当有助于提升生成式 AI 赋能教育的效果，更好地服务于育人目标的实现。

【参考文献】

[1]姜华,王春秀,杨暑东.生成式 AI 在教育领域的应用潜能、风险挑战及应对策略[J].现代教育管理,2023,(7):66-74.

[2]余胜泉,汪凡淙.人工智能教育应用的认知外包陷阱及其跨越[J].电化教育研究,2023,(12):5-13.

[3]郑永和,王一岩,杨淑豪.人工智能赋能教育评价:价值、挑战与路径[J].开放教育研究,2024,(4):4-10.

[4]刘邦奇,黄倩倩,张金霞.GenAI 教育应用伦理风险多主体协同治理框架及实践策略[J].现代远程教育研究,2025,(5):48-58.

[5]汪凡淙,汤筱玓,余胜泉.基于生成式人工智能的认知外包:交互行为模式与认知结构特征分析[J].心理学报,2025,(6):967-986.

[6]杨晓哲,舒薇澄.人工智能介入课堂教学的边界审视[J].北京教育学院学报,2026,(2):1-6.

[7]吴龙凯,程浩,张珊,等.智能技术赋能教育评价的时代内涵、伦理困境及对策研究[J].电化教育研究,2023,(9):19-25.

[8]张羽,杨子豪,覃菲.人工智能助力教育变革的风险研判、归因分析与生态治理[J].电化教育研究,2025,(7):19-25.

[9]刘泽民,陈向东.人工智能科学(AI4S)引发的范式变革——大语言模型视角下教育研究的问题、方法与过程[J].远程教育杂志,2024,(5):23-34.

[10]黄廷祝.人工智能时代教学形态的主动变革[J].中国大学教学,2025,(Z1):85-91+107.

[11]胡钦太,张彦,伍文燕.人工智能与教育深度融合:逻辑、困境与路径[J].中国电化教育,2026,(6):1-6+23.

[12]周驰亮.数智技术赋能职业教育教学评价的价值表征、伦理挑战与应对策略[J].中国职业技术教育,2024,(35):77-84.

[13]施大宁.学为中心,智能向善——AI 时代背景下教学模式变革的思考[J].中国大学教学,2025,(5):4-8.

[14]魏艳涛,徐琦,王鑫茹,等.一体化推进大中小学人工智能教育的价值意蕴与实施路径[J].中国电化教育,2026,(1):13-20+45.

[15]沈苑,房斯萌,柳晨晨,等.生成式人工智能教育应用治理:案例与反思[J].开放教育研究,2024,(6):39-47.

[16]王一岩,朱陶,杨淑豪,等.人机协同教学:动因、本质与挑战[J].电化教育研究,2024,(8):51-57.

[18]余胜泉.跨越人工智能教育应用的认知外包陷阱[J].中国教育学刊,2025,(4):1.

[19]罗蓉,张安富,王剑.AI 时代高校教育教学数智化变革的理性思考[J].高教发展与评估,2026,(2):1-9+181.

[20]刘洁.AI 重新定义教育:理念重塑、逻辑重建与场景重构——第五届《教育研究》论坛综述[J].教育研究,2025,(8):156-159.