高等教育环境下生成式人工智能何以影响学习结果？—基于27项实验与准实验的元分析

How Does Generative Artificial Intelligence Affect Learning Outcomes in Higher Education?

A Meta-Analysis of 27 Experimental and Quasi-Experimental Studies

王佳伟* 董伟
天津大学教育学院
hbujiaweiwang@163.com

【摘要】针对学习者使用生成式人工智能（GenAI）可能引发的“认知卸载”风险，研究旨在厘清GenAI对高等教育学习结果的影响及边界条件。本研究采用元分析方法，纳入27项实验与准实验研究，检验保持、迁移与延迟测试结果及调节效应。GenAI总体显著提升学习结果，但效果高度依赖条件；其显著促进保持与迁移的即时表现，而延迟效应不稳定。长期干预（>4周）、脚手架引导与支架型交互能有效促进能力跨情境迁移与长效维持。未来应用需从“工具接入”转向“机制设计”，以规避“认知债务”风险。

【关键词】生成式人工智能；认知卸载；学习结果；元分析；教学支架

Abstract: *To address the risk of “cognitive offloading” that learners may face when using generative artificial intelligence (GenAI), this study aims to clarify the impact of GenAI on learning outcomes in higher education and its boundary conditions. A meta-analysis was conducted, incorporating 27 experimental and quasi-experimental studies to examine retention, transfer, and delayed test results, as well as moderating effects. The findings indicate that GenAI significantly improves learning outcomes overall, yet its effects are highly context-dependent. While it significantly enhances immediate performance in retention and transfer, its delayed effects are less stable. Long-term interventions (>4 weeks), scaffolding guidance, and scaffold-based interaction effectively promote cross-contextual transfer and long-term maintenance of competencies. Future applications should shift from mere “tool access” to systematic “mechanism design” in order to mitigate the risk of “cognitive debt”.*

Keywords: Generative Artificial Intelligence (GenAI); Cognitive Offloading; Learning Outcomes; Meta-analysis; Instructional Scaffolding

1.引言

随着ChatGPT为代表的生成式人工智能（Generative Artificial Intelligence，GenAI）深度嵌入高等教育，学习者的加工认知链条正在被深刻重塑（Zhang,2025）。相较于传统数字工具，GenAI不仅提供信息检索，更通过生成解释、重组知识等方式直接介入推理与协作学习过程（Villar-Onrubia et al.,2025）。在此背景下，认知卸载（Cognitive offloading）风险日益凸显：虽然GenAI能释放认知资源并快速提升短期任务绩效，但也极易导致学习者将自我解释与监控等关键环节让渡给系统，进而形成“认知债务”（Kosmyna et al.,2025; Risko et al.,2016）。若教育价值评估仍主要依赖即时测验或课堂产出，短期表现与真实习得之间的界限将被严重弱化(Soderstrom et al.,2015)。

针对GenAI的学习效应，既有实证研究在结论上呈现显著分歧（Bastani et al.,2025; Kestin et al.,2025），这提示其效应具有高度的条件依赖性。尽管近期已有部分元分析探讨了GenAI对学生学习表现的总体影响（Wang et al.,2025），但这些研究多聚焦于即时或短期的产出指标，尚未在“保持—迁移—延迟”等更具深度的认知结果层级上进行细分检验。此外，现

有综合研究未能充分澄清工具介入的边界条件。例如，当交互导向“替代式”完成时，图式建构与策略内化随之受损；而当其作为强调反馈约束的“支架”时，则更可能促发生成性加工并转化为长期收益(Soderstrom et al.,2015;Deng et al.,2025)。

综上，亟需通过系统的量化综合估计，澄清 GenAI 学习效应的边界条件与风险区间，为“认知债务”的出现情境提供更具解释力的证据。本研究旨在回答以下核心问题：(1)GenAI 介入对高等教育学习结果的总体效应如何？其在保持、迁移与延迟测试三个结果层级上的效应大小与稳定性是否存在差异？（2）学习情境、材料特征与交互特征是否对效应异质性具有系统解释力？哪些条件会触发效应的增强、衰减或反转？

2.文献综述

随着 GenAI 深度嵌入高等教育，其功能已直接介入学习者的认知加工过程。当前学界围绕 GenAI 学习效应的机制争议，主要聚焦于其究竟是促进了“生成性加工”还是诱发了“认知卸载”（Bastani et al. 2025;Risko et al.,2016）。一方面，若交互能触发解释与反思，GenAI 便构成认知支架；另一方面，若关键推理长期由系统承担，则可能导致图式建构受损，使学习收益仅停留在即时表现，难以沉淀为跨情境的迁移能力或在延迟测验中维持（Kestin et al., 2025）。因此，教育价值的评估应超越单一的即时成绩，向保持、迁移与延迟测试等更深层的结果层级延伸。

尽管近期实证研究已初步揭示 GenAI 与学习表现的正向关联，但相关结论存在显著异质性（Wang et al., 2025）。更为关键的是，现有元分析在解释效应异质性时，多停留在宏观的学科或学习模式层面，未能深入探究交互限制性、干预时长及教学护栏等对“认知债务”风险的具体调节机制。

在机制层面的理论分歧之外，实证研究对 GenAI 学习效应的报告亦呈现出显著的异质性，提示其效应极可能受制于特定的边界条件。尽管多数研究观察到 GenAI 能正向提升即时学习表现或情感动机，但这些收益并未在所有情境中稳定出现（Wu et al., 2025）。现有证据表明，在缺乏教学护栏的通用交互中，学习者容易出现深层认知投入不足，进而削弱高阶思维及长时距评估中的表现（Bastani et al.,2025）；相反，遵循“必要难度”原则的支架式 AI 辅导则更有利于促进知识的深度内化（Nelson et al,2023）。此外，干预时长、任务复杂性及学习模式等方法学层面的差异，也导致了研究结论的进一步分化。这表明，要全面评估 GenAI 的教育价值，需要系统识别并检验影响其效应转化的调节因素。

因此，既有研究中的关键边界条件可系统归纳为三大调节维度。其一，学习环境特征。在真实教学中，教师的脚手架引导能促使学生将 AI 输出作为解释支架而非答案替代，从而彻底改变知识保持与迁移的轨迹。其二，材料特征。相较于易诱发直接采用路径的简单、良构任务，复杂、劣构任务更迫使学习者进行表征重构，进而触发跨情境的能力迁移。其三，交互特征。具备学习护栏的支架型交互能够维持必要的认知挑战，其对延迟保持稳定性的提升，显著优于无限制的替代型交互。综上所述，本研究将以保持成绩、迁移成绩与延迟测试成绩作为核心评估指标，并基于上述三大维度构建调节变量编码体系，旨在通过量化综合，深度澄清除即时表现外，GenAI 学习效应发生异质性的内在机制。

3.研究设计

3.1 数据来源

基于上述问题，本研究严格遵循 PRISMA 准则开展系统检索，时间范围限定为 2022 年 11 月至 2025 年 12 月。为兼顾学科覆盖与文献完整性，检索数据库包括 Web of Science、PubMed、ERIC 等引文与学科索引数据库，以及 ElsevierScienceDirect、Taylor & Francis、Sprin

ger Nature Link 等出版商数据库。检索利用关键词 Generative AI/GenAI/Large Language Model/LLM/ChatGPT/GPT-/Claude/Gemini/Chatbot/Conversational Agent/Copilot 和 Human-AI interaction/Human-AI collaboration/Practice mode/Pedagogical mode/Scaffold/Tutor/Learning companion/Teaching assistant 和 Learning outcome/Higher-order thinking/HOTS/Critical thinking/Creative thinking/Problem solving/Self-regulat/Skill acquisition 以及 Experiment/Quasi-experiment/Randomized controlled trial/RCT/Control group/Pre-test/Post-test/Comparison group 进行检索。此外，采用滚雪球法对纳入研究的参考文献进行回溯，以补充潜在遗漏的关键实证研究。

文献筛选流程见图 1。初检共获得 893 条记录。经题名、摘要与关键词初筛，剔除重复记录、非期刊学术论文及主题不相关研究 532 篇，纳入 232 篇进入下一轮。进一步按研究类型与设计标准筛除综述、非实证与非教育场景研究，以及不满足随机对照或准实验设计者 56 篇，余 188 篇。全文审查阶段依据元分析统计要求排除效应量不可提取、关键调节变量缺失或样本量差距过大者 132 篇，保留 56 篇。随后进行内容一致性核验，剔除不构成 GenAI 干预或与研究问题不匹配者，最终纳入 27 篇期刊论文。

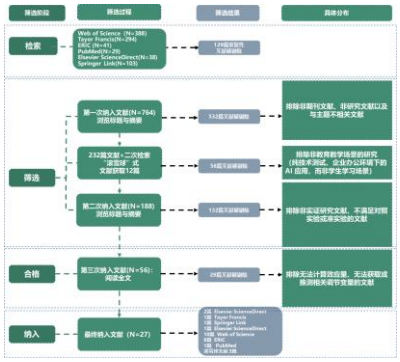

图 1 文献检索与纳入流程

3.2. 文献编码

元分析的文献编码维度包括环境特征、材料特征、交互特征及统计特征等四个核心维度。环境特征涵盖学习环境(实验室环境、在线环境、真实课堂)、干预时长（<1 周、1~4 周、>4 周）与引导方式（自主探索、教师引导/脚手架支持）；材料特征涵盖学科领域(自然学科和人文社科)、任务复杂性（简单/良构任务和复杂/劣构任务）和生成内容类型（文本生成、代码生成、多模态生成）；交互特征涵盖交互角色(助教、学伴)、交互限制性(替代型、支架型)和反馈精细度(结果反馈、揭示反馈)；统计特征涵盖文献作者及出版年份、样本量与实验对照组数据。为确保元分析编码的质量与信度，本研究采用双人独立编码。两名编码者先对随机抽取的 5 篇文献进行试编码，并据此协商完善编码框架，随后对其余文献开展正式编码。编码一致性以 Kappa 系数评估，K=0.88，表明一致性良好；分歧结果经复核与协商达成一致。

3.3. 研究方法

本研究采用元分析方法，使用 Comprehensive Meta Analysis 3.7 软件展开。首先，从纳入元分析的 27 篇文献中提取相关统计量。然后，计算合并效应值，根据合并效应值得出 GenAI 对高等教育学生学习成果的整体效应值，同时对纳入的文献进行发表偏倚检验。最后，进行调节效应检验，确定影响保持成绩、迁移成绩和延迟测试成绩的调节变量。本研究 27 篇样本文献中提取 105 项效应值。此外，为排除极端数值的影响，研究使用元分析中排除某一项研究的方法进行影响分析。经过检验，当前元分析中不存在异常值。

4.研究结果

4.1. 发表偏倚检验与异质性检验

发表偏倚是指统计结果显著的研究较不显著的研究更容易被发表的现象，这可能导致元分析夸大真实的效应估计。为了确保研究结论的稳健性，本研究采用漏斗图、Egger 线性回归检验以及失安全系数三种方法，对纳入效应量进行综合评估(Borenstein,2021)。通过多重证据交叉检验发现，元分析结果具有较高稳健性，不存在严重发表偏倚。尽管直观检视漏斗图在中小样本区域（漏斗中下部）存在微弱非对称性，但更为灵敏的 Egger 线性回归结果显示截距 t 值=0.520，p1=0.302,p2=0.604,t 值<1.96 且 p2>0.05，说明存在发表偏倚的可能性较小。

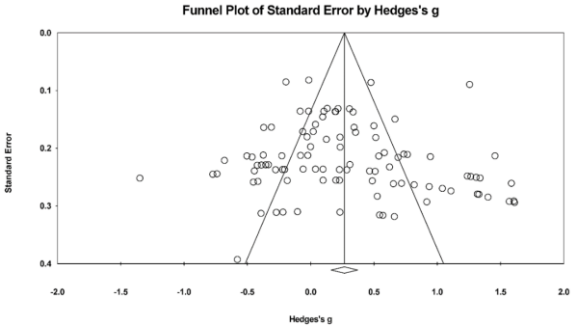

图 2 漏斗图

4.2. 总体效应异质性检验与分析

本研究整体效应检验结果表 1 所示所示。异质性检验结果显示，纳入的 105 个独立效应量之间存在高度显著的异质性（Q=795.735，p<0.001,I^2=86.930%）。这表明高达 86.93%的效应量变异源于研究间样本特征与干预方式的真实差异，而非单纯的抽样误差。鉴于此，本研究采用随机效应模型（Random-effects model）对合并效应量进行稳健估计。

由此可见，GenAI 对学习结果的整体促进作用虽具备统计学意义上的稳定性，但其实际提升幅度相对温和（g=0.267,p<0.001）。同时，极高的 I^2 值揭示了该平均效应背后存在显著的情境波动，表明 GenAI 的实际教学收益高度依赖于特定的边界条件。这一特征为后续开展调节效应分析、进一步识别诱发效应增强或衰减的核心教学要素提供了坚实的统计学基础。

表 1 整体效应检验结果

模型	k	g	95%CI		渐进		异质性检验			
			下限	上限	Z 值	P 值	Q	df	p	I^2
固定效应	105	0.245	0.209	0.282	12.074	0.000	795.735	104	0.000	86.930
随机效应	105	0.267	0.162	0.371	4.992	0.000				

为进一步检验 GenAI 介入效应在不同层级因变量上的差异，本研究分别对学习结果与过程性变量进行了子群分析（如表 2 所示）。总体而言，GenAI 对保持成绩、迁移成绩及部分过程性指标（如认知负荷）具有显著影响，但在延迟测验成绩、学习满意度与学习参与度上未观察到稳定效应。此外，多数结果变量的异质性检验显著且 I^2 值较高，表明效应量在不同研究间存在较大离散度，单一的合并效应不足以完全刻画跨研究的真实效应分布。

学习结果指标呈即时与长效的显著分化。在认知结果层面，GenAI 对保持成绩（g=0.338）与迁移成绩（g=0.381）均具有显著的正向促进作用，且子群内部异质性显著（保持成绩 Q(18)=172.994；迁移成绩 Q(37)=268.981）。相较之下，延迟测试成绩的合并效应未达显著水平（g=0.053），其置信区间跨越零点且异质性显著（Q(16)=65.270）。这充分说明，GenAI 介入带来的即时或跨情境学习增益，尚未在长时距的延迟测量中转化为稳定的总体优势。

过程性指标改善显著但高度依赖实施情境。在学习过程层面，GenAI 对认知负荷呈现显著的降低效应（g=-0.411），但组内异质性显著（Q(4)=16.511）。这种高异质性往往源于任务复杂度与支架设计的差异。当 GenAI 能够有效分解复杂概念时，可显著降低学习者的内在负荷；反之，若缺乏必要的教学护栏，学习者在甄别海量生成结果时极易增加无关的外部认知

负荷（Bastani et al.,2025）。同时，GenAI 显著提升了学习动机（g=0.347），但该指标的异质性同样处于高位（Q(20)=162.909）。这表明，基于技术新奇感与任务便利性带来的初期动机增益，在高强度评估或长期干预下能否转化为稳定的内在认知需求，深度受制于特定的研究情境。

至于学习参与度与满意度，鉴于独立研究数量过少且置信区间过宽，当前证据尚无法支持得出稳健的总体结论。综合上述结果可见，GenAI 对高等教育学习效果的影响呈现出显著的维度敏感性与时间尺度差异。即时性的保持与迁移指标呈显著正效应，而延迟测验未见稳定增益；过程性体验虽在负荷与动机上有所改善，但缺乏跨研究的普遍稳定性。此外，各子群极高的 I^2 值再次印证，这些效应分化绝非偶然抽样误差，而是由更深层的系统性边界条件所驱动。这为本文后续开展调节效应检验指明了方向，即需要进一步剖析何种学习情境、材料属性与交互特征能够真正促使 GenAI 的短期表现红利转化为长期的稳定习得。

表 2 因变量检验结果

结果变量	K	g	95%CI		渐进		异质性检验			
			下限	上限	Z 值	P 值	Q	df	P	I2
认知负荷	5	-0.411	-0.786	-0.035	-2.144	0.032	16.511	4	0.002	75.774
学习参与度	2	0.652	-0.682	1.986	0.958	0.652	17.070	1	0.000	94.142
学习动机	20	0.347	0.069	0.626	2.446	0.014	162.909	19	0.000	88.337
学习满意度	4	0.058	-0.218	0.333	0.409	0.682	8.148	3	0.043	63.180
保持成绩	19	0.338	0.098	0.579	2.763	0.006	172.994	18	0.000	89.595
迁移成绩	38	0.381	0.212	0.550	4.420	0.000	268.981	37	0.000	86.244
延迟测试成绩	17	0.053	-0.177	0.284	0.452	0.652	65.270	16	0.000	75.487

4.3. 不同调节变量对学生学习结果的影响

4.3.1. 能力迁移维度的调节效应分析：干预时长与支架引导的关键作用

能力迁移维度受干预时长与引导方式的显著调节。子群调节分析结果显示，GenAI 对迁移能力的促进作用深度受制于特定教学条件的系统性调节（如表 3 所示）。具体而言，干预时长表现出显著的组间差异（Q_B=13.742,p=0.001）。短期介入（少于 1 周）未能产生稳定的迁移增益，中期介入（1 至 4 周）效应虽转为正向但置信区间仍跨越零点（g=0.272,p=0.090），唯有长期介入（大于 4 周）才获得了显著且幅度可观的迁移增益（g=0.503,p<0.001）。这表明迁移能力的形成高度依赖持续性的练习反馈循环与跨任务应用机会，短期的工具接入不足以转化为稳定的跨情境应用能力。此外，引导方式对迁移效应亦具有高度显著的调节作用（Q_B=14.605,p<0.001）。在教师引导或脚手架支持条件下，迁移成绩呈现显著正效应（g=0.449），而在缺乏护栏的自主探索条件下，效应则转为负向且不显著（g=-0.171）。这一对比充分证实，迁移收益应在结构化的教学设计中才能被稳定激发。只有当学习者获得明确的任务要求、过程性提示与监控支持时，GenAI 的输出才更可能被转化为推理与整合的素材，进而服务于跨情境应用。值得注意的是，学习环境、学科领域及任务复杂度等其余变量未表现出显著的组间差异。鉴于部分子组纳入的研究数量偏少且统计功效受限，这种不显著不应直接等同于无调节作用，后续研究仍需通过扩大样本量或采用元回归方法对这些潜在边界进行精细检验。

表 3 GenAI 对迁移成绩的调节效应分析

一级维度	二级维度	调节变量分类	k	g	95%CI		渐进		组间效应值
					下限	上限	Z 值	P 值	Q
环境特征	学习环境	实验室环境	7	0.277	-0.118	0.673	1.375	0.169	$Q_{B=2.099}$ $P_{=0.350}$
		在线环境	3	0.135	-0.232	0.503	0.721	0.471	
		真实课堂	28	0.437	0.227	0.646	4.088	0.000	

表 3 GenAI 对迁移成绩的调节效应分析(续)

	干预时长	<1 周	5	-1.101	-0.327	0.125	-0.878	0.380	$Q_{B=13.742}$ $P=_{0.001}$
		1~4 周	6	0.272	-0.042	0.585	1.696	0.090	
		>4 周	27	0.503	0.274	0.731	4.312	0.000	
	引导方式	自主探索	4	-0.171	-0.432	0.091	-1.278	0.201	$Q_{B=14.605}$ $P=_{0.000}$
		教师引导/脚手架支持	34	0.449	0.269	0.629	4.879	0.000	
材料特征	学科领域	自然学科	12	0.288	0.022	0.001	1.965	0.049	$Q_{B=0.568}$ $P=_{0.451}$
		人文社科	26	0.425	0.011	0.217	4.015	0.000	
	任务复杂度	简单/良构任务	1	0.311	-0.137	0.759	1.360	0.174	$Q_{B=0.088}$ $P=_{0.767}$
		复杂/劣构任务	37	0.384	0.211	0.556	4.346	0.000	
	生成内容类型	文本生成	30	0.359	0.182	0.537	3.969	0.000	$Q_{B=0.225}$ $P=_{0.893}$
		代码生成	7	0.494	-0.126	1.114	1.561	0.118	
		多模态生成	1	0.311	-0.137	0.759	1.360	0.174	
交互特征	交互角色	助教	7	0.156	-0.278	0.590	0.705	0.481	$Q_{B=1.318}$ $P=_{0.251}$
		学伴	31	0.433	0.246	0.619	4.550	0.000	
	交互限制性	替代型	11	0.226	-0.049	0.501	1.610	0.107	$Q_{B=1.548}$ $P=_{0.213}$
		支架型	27	0.445	0.236	0.655	4.169	0.000	
	反馈精细度	结果反馈	2	0.014	-0.474	0.501	0.055	0.956	$Q_{B=2.163}$ $P=_{0.141}$
		揭示反馈	36	0.403	0.226	0.580	4.450	0.000	

4.3.2. 延迟测试维度的调节效应分析：引导方式与交互限制性作为关键条件

针对延迟测试维度（如表 4 所示），虽然整体合并效应未达统计学显著，但调节效应分析揭示了引导方式与交互限制性存在显著的组间差异（Q_B=9.871,p=0.002）。具体而言，在缺乏结构化支持或呈现完全替代型交互的条件下，延迟测试表现出显著的负向效应（g=-0.680）。而在具备支架属性或护栏引导的条件下，效应点估计虽转为正向但未达显著（g=0.102）。这表明 GenAI 介入带来的即时增益在长时距测量中存在极大的衰减风险，长效收益的沉淀与维持深度依赖于系统交互能否强制保留并激活学习者关键的深层加工与自我监控环节。

表 4 GenAI 对迁移成绩的调节效应分析

一级维度	二级维度	调节变量分类	k	g	95%CI		渐进		组间效应值
					下限	上限	Z 值	P 值	Q
环境特征	学习环境	实验室环境	2	0.161	-0.587	0.909	0.423	0.672	QB=5.048 P=0.080
		在线环境	1	0.654	0.141	1.167	2.500	0.012	
		真实课堂	14	-0.000	-0.252	0.252	-0.003	0.998	
	干预时长	<1 周	4	0.222	-0.265	0.708	0.893	0.372	QB=0.754 P=0.686
		1~4 周	4	0.064	-0.614	0.742	0.184	0.854	
		>4 周	9	-0.027	-0.312	0.258	-0.186	0.852	
	引导方式	自主探索	1	-0.680	-1.114	-0.247	-3.076	0.002	QB=9.871 P=0.002
		教师引导/脚手架支持	16	0.102	-0.122	0.325	0.890	0.373	
材料特征	学科领域	自然学科	—						QB=0.000 P=1.000
		人文社科	17	0.053	-0.177	0.284	0.452	0.652	
	任务复杂度	简单/良构任务	2	0.086	-1.008	1.180	0.154	0.878	QB=0.004 P=0.951
		复杂/劣构任务	15	0.051	-0.189	0.291	0.417	0.677	

表 4　GenAI 对迁移成绩的调节效应分析（续）

	生成内容类型	文本生成	16	0.017	-0.217	0.251	0.142	0.887	QB=4.911 P=0.027
		代码生成	—						
		多模态生成	1	0.654	0.141	1.167	2.500	0.012	
交互特征	交互角色	助教	6	0.140	-0.322	0.601	0.594	0.552	QB=0.261 P=0.609
		学伴	11	0.002	-0.254	0.258	0.017	0.986	
	交互限制性	替代型	1	-0.680	-1.114	-0.247	-3.076	0.002	QB=9.871 P=0.002
		支架型	16	0.102	-0.122	0.325	0.890	0.373	
	反馈精细度	结果反馈	—						QB=0.000 P=1.000
		揭示反馈	17	0.053	-0.177	0.284	0.452	0.652	

5.讨论与启示

5.1. 研究讨论

本研究的元分析结果表明，GenAI 对高等教育学习结果总体呈现显著但幅度温和的正向促进作用。然而，高度的研究间异质性揭示了该效应具有极强的条件依赖性。

在学习结果的维度分化上，即时增益尚未稳定转化为长效习得。认知结果方面，GenAI 在干预阶段能有效提升保持与迁移表现，但延迟测试的总体合并效应不显著且置信区间跨越零点。这一结果印证了学习科学的经典论断，即技术赋能下的高表现并不等同于真实的认知内化(Soderstrom et al.,2015;Deng et al.,2025)。当推理、整合与监控等关键加工过程被工具部分替代时，撤除支持后的延迟测量往往会暴露能力沉淀的不足。过程指标方面，GenAI 虽能显著降低认知负荷并提升学习动机，但学习参与度与满意度并未获得稳定支持，表明过程性体验的改善仍受制于任务压力与交互质量等复合因素。

在效应的边界条件上，长效收益深度依赖系统化的教学设计。调节分析证实，迁移成绩存在显著的“时长效应”趋势，短期介入难以形成稳定增益，而超过 4 周的长期干预则能通过持续的练习与反馈循环激发更为显著的跨情境迁移。同时，引导方式起到决定性调节作用。教师引导与脚手架支持条件下的迁移与延迟表现显著优于自主探索，且无限制的替代型使用在延迟测验中释放出明确的“认知卸载”风险信号。需要指出的是，部分变量未显现统计学差异主要受限于当前子群样本量，不应被过度推论为 GenAI 具备跨情境的绝对普适性。综上所述，GenAI 在高等教育中带来即时收益的前提，是需要将其置于充足的干预时长、结构化的支架引导以及严格的交互限制等特定教学设计架构之内。

5.2. 研究启示

为规避认知卸载风险并提升知识内化概率，GenAI 赋能教学应从单纯的“工具接入”转向系统化的“机制设计”，首要在于重构生成性交互机制与动态干预时序。依据“必要难度”理论(Nelson,2023; Bjork et al.,2020)，教学实践应明确限制 GenAI“成品化输出”的使用边界，避免直接供给答案诱发表现与学习的混淆及能力幻觉。GenAI 学习系统应转为以苏格拉底式追问、反例提示等形式提供认知辅助，强制保留学习者的核验与修正步骤，将外部智能可靠地转化为内部心理表征。同时，为防范长期无差别依赖诱发的元认知惰性(Fan et al.,2025)，干预设计需遵循脚手架原则实施动态褪除。从新手期的细粒度支持，逐步向发展期的诊断反思过渡，并在巩固复盘环节最终撤除外部代理，通过认知责任的动态交接，推动学习者完成从外部调节向自我调节的实质性跨越(ATKINSON R K,2003;Ryan et al.,2000)。

在评价范式与主体建构层面，教学体系需协同发力以规避即时表现的遮蔽效应与技术依赖引发的动机置换。鉴于单一的产出导向极易高估真实的学习效果，有必要构建“协同与独立”并重的双轨评价链条。在评估人机协作任务分解与推理透明度的同时，可以通过变式任务强

化脱机状态下近/远迁移能力的独立评估，促使学习者深度重组生成信息，消解协同绩效对独立习得能力的测量遮蔽(Bastani et al.,2025)。此外，过度自动化极易导致学习者在关键环节处于“被替代”地位，引发主体性弱化与心理所有权衰退(Pierce et al.,2001)。因此，整合应用应确立“人为主导、机为辅助”的规范边界，依托理由陈述与选择解释等约束机制(Chatzichristofis,2026; Iranmanesh,2026)，确保学习者在证据判断与最终决策中保有一定控制权，借此维持深层加工的内在动机，使技术真正成为促进独立思维发展的长效支撑。

参考文献

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. Journal of Educational Psychology, 95(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences, 122(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. Journal of Applied Research in Memory and Cognition, 9(4), 475. https://doi.org/10.1016/j.jarmac.2020.09.003

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). Introduction to meta-analysis. John Wiley & Sons.

Chatzichristofis, S. A. (2026). Reframing AI in education: A pedagogical opportunity, not a technocraticthreat. Journal of Research in Science Teaching, 63(1), 97–103. https://doi.org/10.1002/tea.70024

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., … & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology, 56(2), 489–530. https://doi.org/10.1111/bjet.13544

Iranmanesh, A. (2026). Reflection in co-action: AI joins the crit. International Journal of Technology and Design Education, 1–14. https://doi.org/10.1007/s10798-026-10079-6

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. Scientific Reports, 15(1), 17458. https://doi.org/10.1038/s41598-025-97652-6

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V.,& Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. arXiv preprint arXiv:2506.08872. https://doi.org/10.48550/arXiv.2506.08872

Nelson, A., & Eliasz, K. L. (2023). Desirable difficulty: Theory and application of intentionally challenging learning. Medical Education, 57(2), 123–130. https://doi.org/10.1111/medu.14916

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230–253. https://doi.org/10.1518/001872097778543886

Pierce, J. L., Kostova, T., & Dirks, K. T. (2001). Toward a theory of psychological ownership in organizations. Academy of Management Review, 26(2), 298–310.https://doi.org/10.5465/amr.2001.4378028

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68-78. https://doi.org/10.1037/0003-066X.55.1.68

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. Perspectives on Psychological Science, 10(2), 176–199. https://doi.org/10.1177/1745691615569000

Villar-Onrubia, D., Cachia, R., Rietz, C., Feltrero, R., Niemi, H., Hallissy, M., & Reuter, R. (2025). Generative artificial intelligence in secondary education: Uses and perceptions from the perspective of early adopters across five EU member states. https://doi.org/10.2760/8636621

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. Humanities and Social Sciences Communications, 12(1), 1–21. https://doi.org/10.1057/s41599-025-04787-y

Wu, S., Liu, Y., Ruan, M., Chen, S., & Xie, X. Y. (2025). Human-generative AI collaboration enhances task performance but undermines human's intrinsic motivation. Scientific Reports, 15(1), 15105. https://doi.org/10.1038/s41598-025-98385-2

Zhang, A. (2025). Information retrieval in the age of generative AI: A mismatch that matters. Legal Reference Services Quarterly, 44(3), 297–306. https://doi.org/10.1080/0270319X.2025.2536920

教育智能体从模型驱动走向意图驱动的编排范式

From Model-Driven to Intention-Driven: An Orchestration Paradigm for Educational Agents

杨海亮[1]，黄慧鹏[1]，吴永和[1]*
[1]华东师范大学教育信息技术学系
*yhwu@deit.ecnu.edu.cn

【摘要】随着生成式人工智能由内容生成走向任务执行，教育智能体开始进入教学设计、课堂支持与学习评价等更复杂的教育场景。但现有应用仍多停留于答疑、推荐、评分等局部环节，缺乏面向课堂全过程的协同机制，难以支撑教学目标、学习活动与评价证据的一体化联动。基于此，本文提出从模型驱动到意图驱动的教育智能体编排范式，构建目标，规划，行动，反馈的编排主链，并从目标层、机制层与治理层阐释其教育意涵。研究认为，教育智能体的核心不在于增强模型输出能力，而在于将教学意图转化为可解释、可执行、可审计的课堂过程。本文可为教育智能体走向教学编排提供理论参考。
【关键词】 教育智能体；意图驱动；模型驱动；编排范式；智能课堂

Abstract: *As generative artificial intelligence evolves from content generation to task execution, educational agents are increasingly entering more complex educational scenarios such as instructional design, classroom support, and learning assessment. However, most existing applications still remain limited to localized functions such as question answering, recommendation, and grading, lacking a collaborative mechanism that supports the entire classroom process and thus failing to enable integrated alignment among instructional goals, learning activities, and assessment evidence. In response, this paper proposes an orchestration paradigm for educational agents that shifts from model-driven to intent-driven logic. It further constructs an orchestration chain of goal–planning–action–feedback and explains its educational implications from the perspectives of the goal layer, mechanism layer, and governance layer. The study argues that the core value of educational agents lies not in enhancing model output capabilities, but in transforming pedagogical intent into classroom processes that are interpretable, executable, and auditable. This paper provides a theoretical reference for moving educational agents from tool-based application toward instructional orchestration.*
Keywords: Educational Agents; Intent-driven; model-driven; Orchestration paradigm; Intelligent classroom

1.引言

生成式人工智能的发展推动智能系统由生成文本走向完成任务，智能体开始具备更强的任务理解、流程分解与工具调用能力。在教育领域，教育智能体也由早期的答疑助手、教学代理和数字化助教，逐步扩展为生成式人工智能支持下的多样化教育支持形态（Bozkurt, 2023）。然而，从当前实践来看，绝大多数教育智能体仍然停留在局部功能增强阶段，其核心仍是围绕模型能力提供服务，即通过更强的语言生成、内容推荐与交互反馈来提升单点任务的完成效率。但教学活动并不是若干孤立任务的简单叠加，而是一个围绕课程目标展开的连续组织过程。它既涉及目标设定、任务实施、课堂互动，也涉及证据收集、形成性评价与结果反馈。若教育智能体仅仅作为一个更强的对话工具存在，那么它固然能够提升信息供给效率，却难以真正进入课堂流程，也难以回应教学中对过程可见、结果可证、责任可溯的现