

生成式人工智能对高校学生编程学习效果影响的元分析

The Effects of Generative Artificial Intelligence on University Students' Programming Learning Outcomes: A Meta-Analysis

杨苏宁¹, 宿庆²

¹⁻² 西北师范大学教育技术学院 (智能教育学院)

suqing8909@nwnu.edu.cn

【摘要】 本研究基于 19 项实证研究开展元分析, 探讨生成式人工智能对编程学习成效的影响。结果表明, GAI 能显著提升学习效果; 其中, 具备编程基础的学生表现更优, 干预周期为 1-3 个月时效果最佳, 将其作为学伴或学习工具的效果优于智能导师, 而专业背景与学段差异不显著。研究建议加强情境化应用设计, 以提升学习成效。

【关键词】 生成式人工智能; 编程教育; 元分析

Abstract: This study conducts a meta-analysis of 19 empirical studies to examine the impact of generative artificial intelligence on programming learning outcomes. The results indicate that GAI significantly improves learning effectiveness. Learners with prior programming experience perform better, and interventions lasting 1-3 months yield the most optimal outcomes. Compared with use as an intelligent tutor, applying GAI as a learning companion or tool leads to better results, while differences in academic background and educational level are not significant. The study suggests strengthening context-sensitive application design to further enhance learning outcomes.

Keywords: Generative Artificial Intelligence; Programming Education; Meta-analysis

1. 引言

近年来, 生成式人工智能 (GAI) 技术取得了突破性进展, 在全球范围内引发了广泛而深刻的关注。然而, 学界对于其在教育领域中的应用效果仍存在一定分歧。一方面, Hsin-Yu Lee 等的研究表明, GAI 的引入能够显著提升学生的批判性思维、问题解决能力以及自我效能感 (Lee et al., 2025), 另一方面, 也有研究指出相较于教师反馈, 基于 GAI 的反馈在促进学生编程学习效果方面可能存在不足 (Er et al., 2025)。基于此, 本研究采用元分析方法, 探究生成式人工智能在高等编程教育中的应用及其对学习效果的影响, 研究聚焦以下两个核心问题:

- (1) GAI 对编程学习成效的整体影响状况如何?
- (2) GAI 对高校学生的编程学习效果是否受到不同调节因素的影响?

2. 研究方法

元分析能够系统整合某一特定问题下的多项独立实证研究, 为研究问题提供更高层次的证据支持。本研究主要借助元分析专用软件 CMA3.0 版进行效应量的计算与综合分析。

2.1 文献检索流程

本研究遵循 PRISMA 规范, 对 GAI 在编程教育领域的应用研究进行量化分析。在文献检索时, 围绕 GAI 和编程教育两个核心的概念进行检索, 检索范围涵盖 Web of Science、EBSCO

host、Elsevier ScienceDirect、Springer Nature Link 等数据库，检索的起始时间未设置，检索截止日期为 2025 年 12 月 7 日，初步获得 2762 篇相关文献。

2.2 文献筛选流程

本研究采用结构化的文献筛选流程，基于研究主题与内容质量要求，制定了明确的纳入与排除标准。研究严格遵循 PRISMA 的系统筛选流程进行文献筛选（见图 1），先基于标题与摘要进行初筛，排除明显不相关或不符合要求的文献；随后对剩余文献进行全文精读与复筛。经过上述严格筛选，最终共 19 篇符合所有质量标准的研究被纳入后续的元分析。

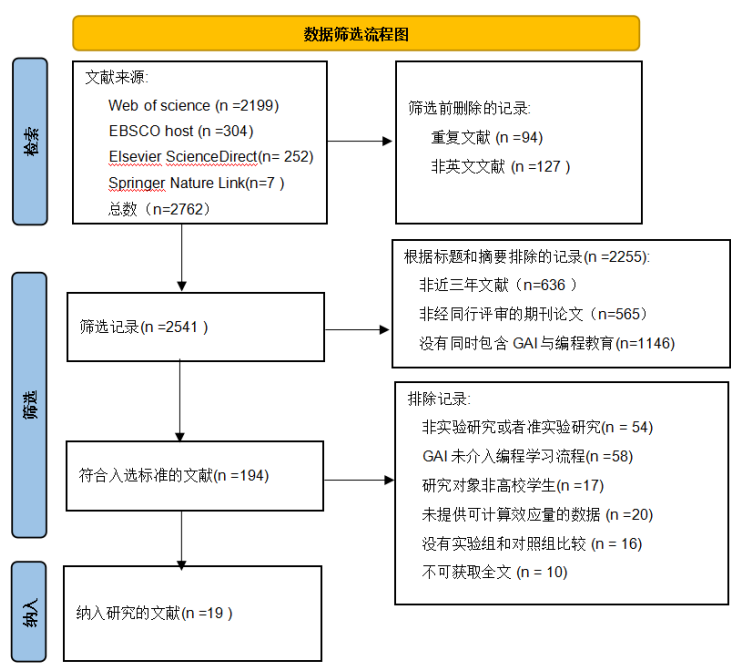


图 10 文献筛选流程

2.3 文献编码

本研究的文献编码包括文献的基本信息、调节变量、实验数据和学习效果的编码，根据研究的问题以及纳入文献的共性，将调节变量分为被试此前是否有编程基础、专业和学段、研究的实验时长，GAI 的应用角色。研究的学习效果根据布鲁姆的教育目标分类分为知识与技能维度和情感态度维度，以系统反映 GAI 在编程教育中的多维影响。

2.4 发表偏移检验与异质性检验

本研究采用 Higgins 等(2003)提出的 I^2 统计量(Higgins et al., 2003), 本研究中 $I^2=81.681\%$ ，说明纳入研究之间存在高度异质性,所以研究选择随机效应模型进行效应量的合并。

在本研究中，漏斗图（图 2）显示散点分布基本对称，且 Begg’s 检验 ($p=0.979$) 与 Egger’s 检验 ($p=0.185$) 结果均未达到统计学显著性水平，三项检验结果一致表明本研究纳入的文献不存在明显的发表偏倚，增强了研究结论的稳健性与可靠性。

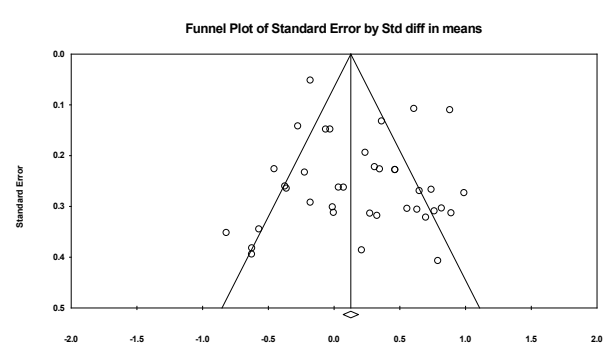


图 11 漏斗图检验

3.结果分析

本研究最终纳入的 19 篇符合质量要求的文献进行系统性的效应量提取与整合,获得 38 个具有统计独立性的效应量,选用 Hedges' g 作为核心效应量估计值。

(1) GAI 对于高等教育学生编程学习效果的整体及其具体维度的效应分析

研究采用亚组分析对一系列调节变量进行了检验。以下是根据数据得到的研究结果:

表 1 总体效应分析数据

结果	效应量				异质性检验			
	K	Hedge s'g	SE	95%置信区间		Qb	Df(Q)	p
总体效	38	0.204	0.07	0.05	0.35			
应			9	0	8			
知识	26	0.132	0.09	-	0.32	1.6	1	0.20
与技能			7	0.059	2	29	2	
情感态	12	0.342	0.13	0.08	0.60			
度			4	0	4			

注:k=效应量数量; Hedges'g=平均效应量; SE=平均效应量的标准误; Qb=Cochran 的组间异质性同质性检验统计量; p<0.05。

GAI 对高等教育编程学习效果存在微弱但积极的正向影响(g=0.204)。进一步分层分析显示, GAI 对编程学习中的知识与技能的效果(g=0.132)呈现微弱正向效应; 而对编程学习中情感态度的效果(g=0.342)表现出中等强度促进作用。

(2) GAI 对于高校编程学习效果的调节变量分析

学生之前没有编程基础 (g=-0.073) 对学生的编程学习效果存在微弱的正向影响, 而学生之前有编程基础 (g=0.207) 对学生的学习效果存在中等程度的正向影响。实验时长进行亚组分析结果显示若干干预时长小于一个月 g=0.119 其呈现微弱的正向趋势, 而当其干预时长在一个月到三个月之间时 (g=0.283), 三个月及以上时 (g=0.149), GAI 对学生编程学习效果存在中等程度的正向影响。若将 GAI 用作智能导师 (g=-0.187), GAI 对编程学习效果存在微弱的负向影响, 若 GAI 用作智能学习工具或伙伴 (g=0.289) 则 GAI 对编程学习效果存在中等程度的正向影响。若 GAI 用作智能学生 (g=0.317), GAI 对编程的学习效果存在中等程度的正向影响。

GAI 对本科生群体 (g=0.206) 的学习效果存在微弱的正向影响; 对研究生群体 (g=0.058) 的学习效果存在微弱的正向影响, 这一差异可能源于学习任务与 GAI 能力的匹配度。

学生所学专业与计算机专业相关时 (g=0.152), GAI 对学生的编程学习效果具有微弱的促进作用, 学生所学专业与计算机专业无关 (g=0.234) 时, GAI 对学生的学习效果产生中等程度的正向影响。

4.讨论与结论

(1) 生成式人工智能对高校学生编程学习效果的整体影响

整体来看, GAI 对高等教育编程学习效果存在积极的影响, 具体来说, GAI 对高校学生编程效果的影响情感态度的效果要优于知识与技能, 说明当前阶段 GAI 在直接提升学生对编程

知识的掌握程度和核心技能熟练度方面作用有限。相对而言，GAI 对情感态度维度则表现出中等程度的积极影响。其可能原因可由“必要难度理论”加以解释。在编程学习中，调试代码与排查错误本是关键的认知训练过程，而 GAI 提供的即时反馈与个性化支持，减少了反复试错的环节。这种便利性在一定程度上提升了学生的自我效能感与学习参与度(Tang et al., 2025)。但也可能削弱了学习所必需的深度认知投入。

(2) 生成式人工智能对高校学生编程学习效果的调节效应分析

GAI 对有编程基础的学生的编程学习效果的影响要优于没有编程基础的，可能的原因有，其一，缺乏编程经验的学习者在使用 GAI 时更易受到生成内容误导，影响对编程概念的正确理解。其二，由于缺乏必要的知识基础，初学者在面对生成结果时容易产生元认知困难，并在高认知负荷下倾向于“认知外包”，即过度依赖工具而忽视对核心概念与推理过程的理解。这种情况下，学习者难以有效评估生成内容，从而对长期学习效果产生不利影响。

实验周期在 1-3 月学习效果优于小于一个月以及三个月以上，其原因主要体现在两方面：一是短期内学习者尚未形成有效的 GAI 使用策略，易产生依赖或误用，从而影响学习效果；随着经验积累，GAI 逐渐内化为“认知工具”，促进技能迁移与元认知发展。二是编程技能与计算思维的提升具有滞后性，虽可在长期中受益于持续反馈，但过度依赖亦可能削弱深度思考与持续性认知增益。

本研究中，将 GAI 作为学伴或学习工具的效果优于智能导师，GAI 作为智能导师效果较弱，主要因其偏向直接给出答案，削弱了学生的自主性与深度思考。但部分研究结果较好，是由于其更侧重引导与启发，促进了学生的认知参与。因此，效果差异关键在于具体的交互方式与功能设计。本科生群体使用 GAI 学习效果比研究生群体好，本科阶段侧重基础知识与基本技能，较契合 GAI 在代码生成与调试解释方面的优势；而研究生阶段更强调复杂问题与高阶认知，对创新与深度理解要求更高，GAI 支持相对有限。说明关键在于资源与学习者认知水平的匹配。学生所学专业与计算机专业无关时学习效果要好于与计算机专业相关，从深层原因看，计算机专业强调底层逻辑与计算思维，而 GAI 直接生成代码易削弱学生的自主探究与深度思考，降低认知深度。相比之下，非计算机专业学生多将 GAI 作为辅助工具，用于降低编程负担、服务于本专业的知识学习，因而更可能产生积极效果。

参考文献

- Er, E., Akcapinar, G., Bayazit, A., Noroozi, O., & Banihashem, S. K. (2025). Assessing student perceptions and use of instructor versus AI-generated feedback. *British Journal of Educational Technology*, 56(3), 1074–1091.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ (Clinical Research Ed.)*, 327(7414), 557–560.
- Lee, H.-Y., Chen, P.-H., Lin, C.-J., Huang, Y.-M., & Wu, T.-T. (2025). Leveraging ChatGPT for personalized reflective learning in programming education: Effects on self-efficacy, higher-order thinking, and project implementation skills. *Education and Information Technologies*, 30(17), 24815–24854.

Suh, J., Lee, K., & Lee, J. (2025). Programming education with ChatGPT: Outcomes for beginners and intermediate students. *Education and Information Technologies*.