

基于生成式人工智能的人机协同学习更能提升学习成效?

——基于20项实验和准实验的元分析

和文斌¹ 赵 帅² 阿不来提·瓦依提¹ 塔卫刚¹ 徐恩伟¹

(1. 新疆师范大学 教育科学学院 新疆乌鲁木齐 830017;
2. 鲁东大学 初等教育学院 山东烟台 264000)

[摘要] 生成式人工智能的崛起为人机协同学习注入了动力。然而,学界对基于生成式人工智能的人机协同学习成效仍有争议。本研究采用元分析法对国内外20项研究进行整合分析,并从学科领域、知识类型、干预时长等维度探讨调节变量对基于生成式人工智能的人机协同学习成效的影响。结果显示,相比于传统学习方式,基于生成式人工智能的人机协同学习能有效提升学习成效。调节分析表明,基于生成式人工智能的人机协同学习在社会科学领域与程序性知识学习中的表现更好;干预周期越长影响越弱;群体学习、角色设定与翻转课堂对协同学习成效的影响最突出;学科领域、知识类型、学习方式等的组间差异不显著。研究者需加强实践适切性设计,开展围炉群证式学习,注重角色精细化指引和探索模式深度融合,以助推基于生成式人工智能的人机协同学习理论与实践的发展。

[关键词] 生成式人工智能;人机协同学习;学习成效;元分析

[中图分类号] G434

[文献标识码] A

[文章编号] 1007-2179(2024)05-0101-11

一、研究背景

随着人工智能在学习领域的应用,人机协同学

习作为一种新型学习样态受到了广泛关注(艾兴等, 2020)。人机协同学习融合了人类智慧和人工智能,它通过对话协商与信息交流等活动形成“混合增

[收稿日期] 2024-05-10

[修回日期] 2024-06-17

[DOI编码] 10.13966/j.cnki.kfjyyj.2024.05.010

[基金项目] 教育部产学研合作协同育人项目“师范类虚拟仿真综合实训课程开发研究”(231105407275433);新疆维吾尔自治区研究生科研创新项目“面向深度学习的人机协商模式构建研究”(XJ2024G198);新疆维吾尔自治区“天池英才”引进计划2023年度自治区“十四五”重点学科教育学招标课题“师范认证视角下教育技术专业本科生协作深度学习模式研究”(23XJKD0205);新疆师范大学自治区“十四五”重点学科教育学招标课题“新疆高校教师数字素养现状调查与对策研究”(23XJKD0210);新疆师范大学智库招标项目“信息技术支持下农牧区教师能力素养提升机制研究”(ZK202233B)。

[作者简介] 和文斌,博士研究生,新疆师范大学教育科学学院,研究方向:CSCL、人机协同学习(1125419301@qq.com);赵帅,博士,讲师,硕士生导师,鲁东大学初等教育学院,研究方向:人机协同学习(530757630@qq.com);阿不来提·瓦依提,副教授,硕士生导师,新疆师范大学教育科学学院,研究方向:信息技术与课程整合(ablet412@163.com);塔卫刚,博士,副教授,硕士生导师,研究方向:技术支持的学习空间、人机协同学习(taweigang@126.com);徐恩伟,博士,讲师,硕士生导师,新疆师范大学教育科学学院,研究方向:信息技术教育、CSCL(244924608@qq.com)。

[引用信息] 和文斌,赵帅,阿不来提·瓦依提,塔卫刚,徐恩伟(2024). 基于生成式人工智能的人机协同学习更能提升学习成效?——基于20项实验和准实验的元分析[J]. 开放教育研究,30(5): 101-111.

强智能”(王一岩等, 2024), 推动学习者生成或重构智慧(王一岩等, 2022)、获取实用技能(艾兴等, 2020)。然而, 以往的人工智能主体性较弱, 无法满足人类多样、复杂的学习与交互需求(何文涛等, 2023)。生成式人工智能为人机协同学习的进阶发展提供了支撑, 将成为促进教育高质量发展的重要引擎。作为一种新兴技术, 生成式人工智能利用深度学习模型识别和模仿“投喂”数据的模式、规律与结构(Halaweh, 2023; Kaplan-Rakowski et al., 2023), 以“反哺”用户的需求为目标, 生成类人内容(如文本、图像、视频等)。该技术基于无监督或部分监督的机器学习框架, 深入理解复杂的人类语言, 并通过统计概率等生成人造遗迹(Hu, 2022; Jovanović & Campbell, 2022), 最终以对话形式提供自然响应, 在一定程度上打破了原有预设、固化的人机协同模式。同时, 生成式人工智能通过强大的数据、算力、模型与基于反馈的强化学习拓展出的内容生成、语言转译、语境理解与场景复刻(和文斌等, 2024)等能力, 提升机器的拟人化特性, 减少人机协同学习的人机隔阂(Park et al., 2023)。其能力体现在以贴近真实社会情境的反馈内容正向增强学习者的体验(Kasneji et al., 2023; Jauhainen & Guerra, 2023), 进而促进学习者的认知参与和动机, 为人机协同学习由知识获取向共同创造发展提供了可能(Lim et al., 2023)。基于此, 探明基于生成式人工智能的人机协同学习的真实效果与实践路径成为促进人机协同与教育融合的关键。

尽管已有的跨学科研究表明, 利用生成式人工智能助推学习者参与人机协同的认知过程, 学习成效更好(Song & Song, 2023)。然而, 将生成式人工智能嵌入人机协同学习的效果仍存在争议, 实践路径也不清晰。有研究(Hwang et al., 2023; Wu et al., 2024; Li, 2023)表明, 生成式人工智能可以显著提高人机协同学习成效。例如, 李海峰(2023)的研究表明, 引入生成式人工智能的人机协同学习能积极影响学习者的参与度、学习动机和学习体验, 提高学习成效。也有部分研究认为与传统学习方法相比, 整合生成式人工智能的人机协同学习会导致学习成效显著下降(Forero & Herrera-Suárez, 2023; Bašić et al., 2023)。还有研究表明基于生成式人工

智能的人机协同学习结果与传统学习方式无显著差异(Escalante et al., 2023; Amina, 2023)。针对上述研究结论的差异, 本研究着重评估基于生成式人工智能的人机协同学习能否提升学习成效。

二、研究方法

本研究利用元分析方法整合已有多项实证研究结果(如样本量、均值、标准差), 探索基于生成式人工智能的人机协同学习的成效, 包括五个步骤: 1)检索、挖掘有关的实证研究; 2)依据研究目标制定文献筛选标准; 3)参照研究文献与调节变量制定编码特征表; 4)依据研究数据计算效应值; 5)检验发表偏倚与异质性。

(一)文献检索

基于PRISMA流程图(Moher et al., 2009), 研究样本的遴选分三个步骤:

1)将样本采集来源聚焦于国内外知名的学术资源数据库(包括中国知网、Web of Science、Springer Link、JSTOR、Scopus、Proquest、Wiley Online Library and Science Direct)。鉴于ChatGPT3.5以前的生成式人工智能技术主体性较弱, 且现有的人机协同学习的实践探索主要关注后续衍生的语言模型与产品, 本研究将检索时间限定于2022年11月至2024年4月。为尽可能囊括已有的文献, 本研究采用布尔运算符进行文献检索, 检索关键词分两组: 一组为生成式人工智能、文心一言、豆包、星火、聊天机器人、generative ai、Chatgpt、Gemi、Claude and chatbot等有关生成式人工智能的关键词汇; 另一组搜索词包含人机协同学习、影响效果、学习成效、human-machine collaborative learning、collaborative learning、learning achievement、learning outcome、scores、performance、influence等有关学习成效的关键词。2)筛查检索到的人机协同教学(徐振国等, 2024)与人机反馈(Liao et al., 2024)等主题的元分析中所整合的数据。3)阅读检索获得的9594篇文献的标题与摘要, 去重后最终选择了282篇文献并进一步筛选(见图1)。

(二)纳入标准

文献筛选标准包括: 1)至少使用一种基于生成式人工智能的人机协同学习实验或准实验; 2)研究

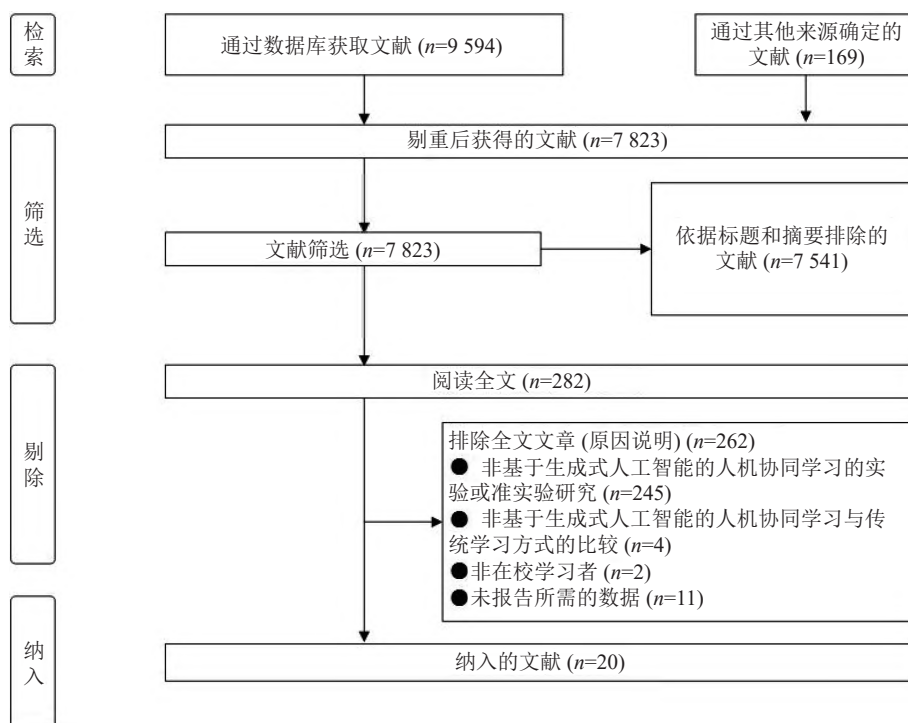


图1 数据采集流程

对象为在校学习者;3)至少包含一项基于生成式人工智能的人机协同学习与传统学习的比较研究;4)通过学习测试等评估基于生成式人工智能的人机协同学习和传统学习的成效;5)具有支撑元分析的数据报告,内容包含平均值、标准差、样本量、t值等。经过上述筛选过程后,研究者通过文献回溯与滚雪球的方式进一步充实研究样本,最终获得20项研究文献(包含22项研究数据)。

(三)样本编码

编码过程设定为码本构建、独立编码与讨论协商三个阶段(见表1):

第一,基于已有研究并结合阅读全文后的信息提取,研究者汇总了三部分信息,共13个字段,分别为基本信息(样本序号、研究主题、研究者与发表年份)、调节变量(学科领域、知识类型、干预时长、学习方式、角色设定与教学模式)与数据信息(均值、标准差、样本量)。依据设定的目标,自变量均为基于生成式人工智能的人机协同学习和传统学习,因变量为学习成效(后测中学习者的知识掌握程度)。第二,经过编码培训,三位研究者独立对22项数据进行编码,编码结果的一致性达91%。

第三,研究者通过邮件咨询文献的作者,对未收到作者回复的文献,其编码中的数据分歧由研究者讨论与协商确定,确保编码的质量与可信度。

(四)效应量计算

鉴于样本数据为实验组与控制组的均值和标准差,且各样本均值差异较大,本研究将标准化均差(SMD)作为合并效应值指标来衡量效应大小,消除均值差较大对结果的影响。标准化均差的计算公式如下:

$$SMD = \frac{M_E - M_C}{\sqrt{\frac{(n_E - 1)S_E^2 + (n_C - 1)S_C^2}{df_E + df_C}}}$$

其中,E代表实验组,C代表控制组,n代表样本量,SMD代表合并效应值,M代表平均值,S代表标准差,df代表自由度。

(五)发表偏倚检验

抽样偏差、显著发表等偏倚因素可能导致纳入的研究样本不能代表真实效果。因此,本研究通过评估分析样本偏倚性,增强研究的质量与可信度,防止结论的失真,即采用失安全系数(Rosenthal's Nfs)、Egger线性回归检验和Begg秩相关检验评

表 1 数据编码本

基本信息	调节变量		数据信息
	类别	次群组	
编号	学科领域	自然科学(物理、化学、生物等)、社会科学(语言、文学、历史等)、 工程技术(机械、计算机、人工智能等)、医学	样本量
标题	知识类型	陈述性(概念、原理等)和程序性(技能、方案等)	平均值
作者	干预时长	0—1 周, 1—4 周, 4—12 周, 12 周以上	标准差
年份	学习方式	个人(与生成式人工智能单独交互开展人机协同学习)和小组(以小组为单位与生成式 人工智能交互开展人机协同学习)	
	角色设定	在人机协同学习过程中对生成式人工智能设定了特定的角色,如 18 岁的物理学习伙伴、化学专家或争论伙 伴等;在人机协同学习过程中直接或间接与原生态的生成式人工智能交互	
	教学模式	自主学习(学习者自主与生成式人工智能开展人机协同学习),传统课堂(在传统课堂授课模式中融入生成 式人工智能辅助人机协同学习),翻转课堂(在翻转课堂中融入生成式人工智能辅助人机协同学习)	

估纳入的样本。失安全系数检验标准为 Nfs 值大于 5N+10。结果显示, Nfs=726 远大于 120。这表明还需 726 份未发表且不显著的研究才能改变元分析结果的显著性(Rosenthal, 1979)。基于此,研究者采用 Egger 与 Begg 检测确认结果。结果显示, Egger 的检测结果为 $Z=0.79<1.96$, $p=0.727>0.05$, Begg 结果为 $p=0.554>0.05$ 。这表明存在发表偏倚的可能较小,元分析纳入的样本能够支撑后续分析,且研究结论的效度和信度较好。

(六)异质性检验

由于纳入的分析样本在研究设计、样本特征与测量工具等维度存在差异,为提高分析结果的科学性和准确性,本研究采用博伦斯坦(Borenstein et al., 2009)开发的分析模型,通过异质性检验选取模型处理数据间的差异性,确保元分析的结果接近无偏估计。选取分析模型依据是 Q 统计量和 I² 统计量。Q 统计量的检测结果为 156.005(df=21, $p<0.001$), I² 为 86.539%(处于 75%-100% 间)。这表明样本间存在较显著的异质性(Higgins et al., 2003)。为有效处理样本间效应值的差异,本研究选择随机效应分析模型校正结果。

三、研究结果

(一)整合效应分析

本研究采用随机效应模型(见表 2)探讨基于生成式人工智能的人机协同学习对学习成效的影响。研究共包含 22 个效应值,合并样本共 1327 份,

合并效应值 SMD 为 0.709, 95% 的置信区间为 [0.388–1.031], $p<0.001$ 。这说明,传统学习方式与基于生成式人工智能的人机协同学习对学习成效的影响差异显著,且总体效应值并非随机变异与偶然误差所致。依据科恩(Cohen, 1992)的元分析效应值评估标准,效应值处于 0.2—0.8 之间,具有中等影响。这表明,相较传统学习方式,基于生成式人工智能的人机协同学习对学习成效有中等程度的正向作用。

表 2 整合效应值

Model	n	效应值 (SMD)	95% 的置信区间		Z	P
随机效应模型 (REM)	22	0.709	0.388	1.031	4.321	<0.001

(二)调节效应分析

除了分析基于生成式人工智能的人机协同学习对学习成效的总体影响,本研究还探讨开展相关实践的有效路径,寻找指导研究设计的方案。为此,本研究依据学科领域、知识类型、干预时长、学习方式、角色设定等调节变量对基于生成式人工智能的人机协同学习成效开展亚组分析(见表 3)。

1. 学科领域

组间异质性结果($Qb=1.581$, $df=3$, $p>0.05$)表明,基于生成式人工智能的人机协同学习对不同学科学习成效的影响效果稳定。具体而言,学科领域的效应值大小依次为社会科学($SMD=0.909$, $df=11$, $p<0.001$)、自然科学($SMD=0.498$, $df=4$, $p>0.05$)、

表 3 调节效应分析结果

调节变量	效应数	效应值(SMD)	95% 置信区间		双尾检验		组间效应		
			下限	上限	Z	P	Q _b	df	P
学科领域							1.581	3	0.664
自然科学	5	0.498	-0.166	1.163	1.470	0.142			
社会科学	12	0.909	0.498	1.321	4.330	0.000			
工程技术	3	0.444	-0.723	1.611	0.746	0.456			
医学	2	0.370	-1.338	2.078	0.424	0.671			
知识类型							0.521	1	0.470
陈述性	8	0.532	-0.080	1.143	1.704	0.088			
程序性	14	0.798	0.413	1.182	4.066	0.000			
干预时长							39.892	3	0.000
0-1 周	1	1.203	0.601	1.806	3.915	0.000			
1-4 周	7	0.919	0.408	1.430	3.524	0.000			
4-12 周	9	0.853	0.479	1.227	4.474	0.000			
12 周以上	2	-0.634	-1.049	-0.219	-2.995	0.003			
学习方式							0.501	1	0.479
个人	15	0.627	0.173	1.080	2.709	0.007			
群体	7	0.847	0.438	1.257	4.057	0.000			
角色设定							0.202	1	0.653
无	14	0.672	0.341	1.003	3.979	0.000			
有	8	0.842	0.179	1.505	2.490	0.013			
教学方式							0.980	2	0.612
自主学习	11	0.577	0.060	1.093	2.189	0.029			
传统课堂	4	0.608	-0.384	1.599	1.201	0.230			
翻转课堂	7	0.919	0.436	1.402	3.730	0.000			

工程技术(SMD=0.444, df=2, p>0.05) 与医学(SMD=0.370, df=1, p>0.05)。也就是说,基于生成式人工智能的人机协同学习对所有学科均具有稳定、积极的正向作用,其中社会科学是利用基于生成式人工智能的人机协同学习提升学习成效最理想的学科。

2. 知识类型

知识类型对基于生成式人工智能的人机协同学习成效产生积极影响,程序性知识的效应值为SMD=0.798(df=13, p<0.001),陈述性知识的效应值为SMD=0.532(df=7, p>0.05)。这表明,生成式人工智能的人机协同学习对掌握程序性知识有积极的推动作用。总体而言,基于生成式人工智能的人机协同学习对所有知识类型的认知均有正向影响,组间无显著差异(Q_b=0.521, df=1, p>0.05)。

3. 干预时长

除12周以上的干预,干预时间长短对人机协

同会产生显著的促进作用,组间差异显著(Q_b=39.892, df=3, p<0.001)。就群组效应值而言,0—1周的干预对学习成效有积极影响(SMD=1.203, p<0.001)。这可能是该群组的样本量少所致。从总体看,随着干预周期的延长,基于生成式人工智能的人机协同学习成效呈下降趋势。

4. 学习方式

学习方式的调节有积极的成效,但组间差异不显著(Q_b=0.501, df=1, p>0.05)。群体参与的学习方式的效应值为SMD=0.847(df=6, p<0.001),个体为单位的学习方式的效应值为SMD=0.627(df=14, p<0.01)。可见,群组参与方式的基于生成式人工智能的人机协同学习效应有更积极的调节成效。

5. 角色设定

不论是否设定角色均能显著提高人机协同学习的成效,但组间差异不显著(Q_b=0.202, df=1, p>0.05)。有特定角色设定的效应值(SMD=0.842,

$df=7$, $p<0.05$)最大,表明在基于生成式人工智能的人机协同学习中,学习者与特定角色的聊天机器人开展交互能获得更大的学习成效。无角色设定($SMD=0.672$, $df=13$, $p<0.001$)的模式也达到显著水平。

6. 教学方式

利用不同教学方式开展人机协同学习的结果显示,组间差异不显著($Q_b=0.980$, $df=2$, $p>0.05$)。自主学习($SMD=0.577$, $df=10$, $p<0.05$)、传统课堂($SMD=0.608$, $df=3$, $p>0.05$)和翻转课堂($SMD=0.919$, $df=6$, $p<0.001$)均对基于生成式人工智能的人机协同学习成效具有较显著的积极影响。就效应值而言,翻转课堂的效应值最大。这表明,利用翻转课堂开展基于生成式人工智能的人机协同学习能产生较好的效果。

四、结论与思考

(一) 研究结论

1. 基于生成式人工智能的人机协同学习对学习成效的影响

随着智能技术日益渗透到日常的工作和生活,人与机器间的互动逐步扩展,人机协同将成为未来学习的主流样态(王一岩等, 2022)。生成式人工智能的出现打破了机器主体性弱的局限,推动了人机协同学习的进阶发展。本研究结果印证了王卓(2023)、伊尔迪兹(Yildiz, 2023)、李(Li, 2023)、宋和宋(Song & Song, 2023)、阿尔内亚迪和瓦尔达特(Alneyadi & Wardat, 2024)、李海峰(2024)等的观点,说明相较传统学习方式,基于生成式人工智能的人机协同学习更有成效。这可能是因为基于生成式人工智能的机器能突破已有人机协同学习中浅表交互、低效持续与预设反馈的弊病(郑兰琴等, 2024)。其强大的内容生成、语言转译、语境理解与场景复刻能力,定制化交互的即时反馈(Hakiki et al., 2023),可以弥合不同需求间的差异,削减认知焦虑,并通过模拟境脉提供逼真的学习体验,辅助学习者实现基于头脑风暴的人机协同知识创造。

2. 基于生成式人工智能的人机协同学习实践路径分析

本研究利用亚组分析探讨影响整体效应异质

性差异的调节变量。22项研究中学科领域、知识类型、干预时长、学习方式、角色设定与教学模式对基于生成式人工智能的人机协同学习成效存在调节作用。

就学科领域而言,不同学科的学习成效都有明显提升,组间差异不显著。其中,社会科学的成效最明显,这一结论与单俊豪(2019)、徐振国(2024)等的研究结论不一。原因可能是社会科学更有社会性与人文性,基于生成式人工智能的机器能够连接、模拟、复刻真实的社会境脉(Hwang et al., 2023),并以类人的对话模式与学习者交互,为学习者提供仿真的学习体验与交流,从而促进知识学习。基于生成式人工智能的人机协同学习在自然科学、工程技术与医学领域的成效也较明显,但由于生成式人工智能难以深入理解这些领域的复杂问题(Alneyadi & Wardat, 2024),致使未能达到与社会科学一样的成效。为此,后续研究需依据学科特点对生成式人工智能做更精细化的设定。

就知识类型而言,应用生成式人工智能开展人机协同学习对陈述性知识和程序性知识的学习成效均产生积极影响,组间差异不显著。这表明,研究无法解释知识类型对效应大小差异影响的关键程度与重要作用。陈述性知识的效应值较小($SMD=0.532$, $p>0.05$),这与徐振国(2024)的研究一致。陈述性知识的学习主要聚焦于学生对事实、概念、数据与事件的理解与认识。尽管生成式人工智能能够以结构化的方式分解概念,并提供个性化的解释(Dasari et al., 2024),但由于其数据库中大量开放信息,且主要通过统计模型和匹配算法处理、分析文字和语句(Mitrović et al., 2023),容易混淆和误解陌生或复杂情境中的语法和概念,从而导致容易生成错误的信息,干扰学习者对陈述性知识的学习成效。因此,利用生成式人工智能开展人机协同学习还需“投喂”高质量的数据,并激励学习者批判性地审视生成内容的准确性(Rees, 2022)。

就干预时间而言,组间差异显著,表明干预时长对人机协同学习成效有不同影响。依据效应值大小排序,利用生成式人工智能开展人机协同学习的周期越长,效应值越小,说明干预时长与基于生成式人工智能的人机协同学习的成效呈负相关。

这一结论与已有研究结论(周进等, 2019; Lee & Hwang, 2022; Wu et al., 2024)一致。原因可能是干预初始时,生成式人工智能的新颖性与即时反馈能激发人们参与学习的积极性、参与度与学习动机(Silitonga et al., 2023),但随着参与频次的增多,新颖效应会逐渐衰弱(Jeno et al., 2019)。此外,学习者长期参与人机协同交互容易因信息过载而分散注意力,从而偏离原有的学习目标,导致学习成效降低(Wu & Yu, 2024)。

就学习方式而言,群体参与的方式可能是利用生成式人工智能促进人机协同学习的有效途径。这一观点得到了已有小组实践研究(Zhu et al., 2023; 李海峰, 2024; 郑兰琴, 2024)的支持。可能的原因在于,生成式人工智能是基于用户提问生成知识的交互技术(沈书生等, 2023),其反馈质量依赖于提问的方式与数据集的定位(Wolfram, 2023)。然而,知识生成中,个人参与人机协同学习在促进学习者推理、澄清与判断等方面效率较低,且交互频率、问题感知、提问的精准性、批判性思考的意识以及对核心问题的关注力都弱于群组。因此,单个学习者可能无法获得比群组更高质量的提问内容或多样化的观点,且容易陷入浅层、存误、虚浮的信息中(Han et al., 2024)。此外,生成式人工智能可有效支撑学习者开展人机协同学习,但此类交互方式不能完全取代人类交互。学习者与同伴互动能有效规避无关信息的侵扰,削减技术沉迷带来的情感淡漠,使学习者获得完整、健康的学习体验(Fauzi et al., 2023)。

就角色设定而言,为生成式人工智能设置特定的交互角色能更有效地提升人机协同学习的成效。这一结论与已有研究结果一致(Zhu et al., 2023; Chang et al., 2024)。首先,这可能是因为生成式人工智能并未连接特定领域的语料信息,投喂的数据主要根植于全域抓取的开放内容(Escalante et al., 2023),导致人工智能无法精准定位学习者的反馈信息。角色设定能够帮助人工智能锚定与学习目标相关的数据集,促使模型生成准确、高质、特定的反哺知识。其次,依托全域数据的生成式人工智能容易形成对多样角色及语言的复杂理解,学习者在这种人机交互中容易陷入机器交互角色叠加带

来的泛化体验。角色设定能促进混合角色的分割,指引生成式人工智能展现更稳定、细腻、连贯的交互表现,增强机器的类人行为(Wang et al., 2023)。这种以角色设定为前提的交互策略,能精准定制交互情境、语调与内容,简化交互过程,提升与学习者的沟通效率和情感连接,通过细腻的多轮对话为学习者认知发展与知识创造提供更真实、丰富、沉浸、生动的学习体验。

就教学方式而言,不同教学方式对人机协同学习成效均产生积极影响,组间差异不显著。其中,利用翻转课堂开展人机协同学习可能是提高学习成效的最有效方式。这一发现与相关研究结论(Cheng & Antonenko, 2019; Strelan & Palmer, 2020)吻合。可能的原因在于,翻转课堂通过调整课内信息传输与课外实践活动顺序,为学习者提供了自由的人机交互时空,有效促进学习者对问题深入理解和多元认知。其次,利用翻转课堂开展基于生成式人工智能的人机协同学习能够发挥技术优势,提高人际互动效果,有效保持人机协同学习的平衡。换言之,尽管生成式人工智能能够与学习者开展类人的自然交互,但仍存在幻觉与偏误信息的输出(Lund & Wang, 2023),且缺乏深入研究学习内容与辩证对话的能力。为此,教师或研究者需合理分配课堂时间,开展教师主导的集体头脑风暴,修正偏误、深化理解、激发问题、应用新知及促进高阶思维发展,最终实现由知识获取到知识应用与创造的递进式学习转型。

(二)简要思考

元分析的结论证明了基于生成式人工智能的人机协同学习对学习成效的积极影响,也启发了利用生成式人工智能开展人机协同学习实践的策略与方法。

第一,加强实践适切性设计。生成式人工智能作为人工智能技术发展的突出成果,并非以人机协同学习为场域而设计的,教师在实践过程需要以适切性为原则,确保人机协同学习的高效运作。1)优化交互设计,拟合学习目标。研究者、教育者在具体的人机协同学习实践设计中需考量学科领域与知识类型的特征,构建与呈现多样化的人机交互脉络,为情感交流与知识学习提供匹配的多感官交互

氛围,帮助学习者清晰阐释社会现象、知识概念与抽象逻辑等。例如,概念性知识的学习可以引入Sora生成复杂情境与知识的三维视频,对抽象内容进行简化、提炼与解析,以多样化的呈现方式促进学习者对知识概念、发展脉络、操作流程与潜在风险的认知与深度理解。2)建立辅助机制,引导认知参与。基于生成式人工智能的人机协同学习,需建立適切性的辅助机制以匹配学习者的认知过程。教师在学习初始阶段需预防学习者沉浸于新颖效应带来的新鲜体验,忽视对知识内容的聚焦;学习中后期需加强对认知参与的激励,消解新颖效应淡化带来的学习认知倦怠,引导学习者高效参与人机协同学习。

第二,开展围炉群证式学习。生成式人工智能对数据的吞吐与加工会受技术的局限,使得人机协同学习易出现认知偏误、情感淡漠、提问浅显与效率低下等问题,教师可以通过围炉群证式学习消弱生成式人工智能带来的负面效果。1)创设围炉氛围。开展围炉群证式学习需为学习者营造开放、包容的学习交互环境,为群体表征与情感交流提供舒适、轻松、公平、尊重、积极的氛围。2)引导群体探讨。研究者或教学者需设置兴趣激发、思维引导与互动推进的群体人机协同学习机制,促进多视角的集体讨论与观点涌现,拓展知识的广度与深度。3)强化群证分析。该模块需引入机器主动提问机制,鼓励学习者批判审视、协作论证与交叉核验,纠正认知偏误从而发展高阶思维。4)催化知识生成。围炉群证的目标是通过集体头脑风暴实现知识的验证与共识的形成,为此需以时间为准绳,帮助群体厘清焦点、拆解问题与生成决议,避免无休止争论,促进高效的知识生成。5)推动反思迁移。研究者或教师需在人机协同学习中设置分析和呈现学习者集体讨论有争议观点的可视化机制,为群体深度反思、知识迁移应用提供支撑。

第三,提升角色精细化指引。角色设定作为提升基于生成式人工智能的人机协同学习成效的有效措施,需研究者开展精细化的角色指引,促进机器精准反馈与学习者体验升级。1)以角色定领域。角色描述是精细指引的起点,是对生成式人工智能进行角色分割与领域锚定的核心环节。研究者需

给予具体、特定、清晰的角色描述,如机器所代表的领域、年龄、身份等,为机器的角色带人与数据提取提供指引。2)以起因定数集。起因说明是精细指引的深化,是辅助生成式人工智能实现数据过滤的策略,学习者可向机器呈现提问的动因,如触发因素与逻辑关系等。3)以情境定体验。情境阐释是精细指引的升级,是机器绘制交互体验的蓝图,研究者需描述问题情境的时空特征、情感氛围与参与人物等。4)以职责定反馈。职责归属是精细指引的优化,是引导生成式人工智能实现特定功能与反馈的导航,研究者需清晰表述机器的具体任务与工作流程等,以减少职责混乱与遗漏造成的反馈缺失。5)以目标定轨迹。目标锁定是精细指引的规范,是促成机器勾画人机协同学习认知轨迹的依据,研究者需对目标的具体细节、重要程度、达成标准与时间框架进行诠释,以规避人机交互脱离认知轨道的风险。

第四,推进模式融合。翻转课堂与生成式人工智能的融合是推动人机协同学习的重要模式,教师需充分发挥生成式人工智能与翻转课堂的优势,推进该模式的发展。1)寻求要素融合。翻转课堂要素与生成式人工智能的融合是模式发展的前提,研究者需重点关注课前内容与情境等要素与生成式人工智能的契合点,优化学习者课前知识学习为主的认知过程,为开展高质量课堂交互提供保障。例如,基于生成式人工智能与可信资源搭建动态、高质量的资源库,可为优质反哺和个性反馈提供精准投喂。2)创新过程融合。课堂交互活动与生成式人工智能的融合是模式发展的关键。该环节需要创新师—生—机三元智能化连接、交互过程设计,为知识探究、群体论证、认知修正、问题解决与高阶思维能力培养提供支持。例如,教师依托生成式人工智能分析课前知识中存在的问题与争议,设计针对性的课堂交互机制。3)探究结果融合。翻转课堂与生成式人工智能的融合是模式发展的升华。此阶段需创新基于生成式人工智能的智能评价方式,为学习方案优化、自我反思提升、知识迁移应用与教学活动改进提供多视角的参照依据。例如,基于生成式人工智能追踪、分析伴随式采集的对话过程与多模数据,可实现以增值评价、过程评价

为代表的综合性评价。

本研究的局限在于: 第一, 研究纳入的样本主要为英文和中文文献, 忽略了其它语言文献, 致使纳入的研究样本有限; 第二, 未来还需纳入更多的研究对总体效应和调节变量做更深入分析, 提升结论的理论与实践意义; 第三, 现有的实证研究主要以大学生为主, 对中小学生的学习探索不够。

[参考文献]

- [1] 艾兴, 赵瑞雪(2020). 人机协同视域下的智能学习: 逻辑起点与表征形态[J]. 远程教育杂志, (1): 69-75.
- [2] Alneyadi, S., & Wardat, Y. (2024). Integrating ChatGPT in grade 12 quantum theory education: An exploratory study at emirate school (UAE)[J]. *International Journal of Information and Education Technology*, 14(3): 389.
- [3] Amina, A. (2023). Exploring ChatGPT potential in flipped classrooms for improving lexicogrammatical features in Algerian EFL learners' essay writing[J]. *International Journal of Early Childhood Special Education*, 15(4): 121-128.
- [4] Basić, Z., Banovac, A., Kružić, I., & Jerković, I. (2023). Better by you, better than me, ChatGPT3 as writing assistance in students essays[J]. arXiv preprint arXiv: 2302.04536.
- [5] Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*[M]. Chichester, UK: Wiley: 77-85.
- [6] Chang, C. Y., Yang, C. L., Jen, H. J., Ogata, H., & Hwang, G. H. (2024). Facilitating nursing and health education by incorporating ChatGPT into learning designs[J]. *Educational Technology & Society*, 27(1): 215-230.
- [7] Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis[J]. *Educational Technology Research and Development*, 67: 793-824.
- [8] Cohen, J. (1992). A power primer[J]. *Psychological Bulletin*, 112(1): 155.
- [9] Dasari, D., Hendriyanto, A., Sahara, S., Suryadi, D., Muhaimin, L. H., Chao, T., & Fitriana, L. (2024). ChatGPT in didactical tetrahedron, does it make an exception? A case study in mathematics teaching and learning[C]//Frontiers in Education. *Frontiers*, 8: 1295413.
- [10] Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference[J]. *International Journal of Educational Technology in Higher Education*, 20(1): 57.
- [11] Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. M. A., & Hatta, H. R. (2023). Analysing the role of ChatGPT in improving student productivity in higher education[J]. *Journal on Education*, 5(4): 14886-14891.
- [12] Forero, M. G., & Herrera-Suárez, H. J. (2023). ChatGPT in the Classroom: Boon or bane for physics students' academic performance?[J]. arXiv preprint arXiv: 2312.02422.
- [13] Hakiki, M., Fadli, R., Samala, A. D., Fricticarani, A., Dayurni, P., Rahmadani, K., Astiti, A. D., & Sabir, A. (2023). Exploring the impact of using Chat-GPT on student learning outcomes in technology learning: The comprehensive experiment[J]. *Advances in Mobile Learning Educational Research*, 3(2): 859-872.
- [14] Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation[J]. *Contemporary Educational Technology*, 15(2): ep421.
- [15] Han, Z., Battaglia, F., Udaiyar, A., Fooks, A., & Terlecky, S. R. (2024). An explorative assessment of ChatGPT as an aid in medical education: Use it with caution[J]. *Medical Teacher*, 46(5): 657-664.
- [16] 和文斌, 范占江, 郑浩, 谭征, 赵帅(2024). ChatGPT 赋能 CSDL: 未来图景与突围路径[J]. *现代教育技术*, (4): 37-46.
- [17] 何文涛, 路璐, 周跃良, 周睿(2023). 智能时代人机协同学习的本质特征与一般过程[J]. *中国远程教育*, (3): 12-20.
- [18] Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses[J]. *BMJ British Medical Journal*, 327(7414): 557-560.
- [19] Hu, L. (2022). Generative AI and Future[EB/OL]. [2022-07-17]. <https://pub.towardsai.net/generativeai-and-future-c3b1695876f2>.
- [20] Hwang, W. Y., Nurtantayana, R., Purba, S. W. D., Hariyanti, U., Indrihapsari, Y., & Surjono, H. D. (2023). AI and recognition technologies to facilitate English as foreign language writing for supporting personalization and contextualization in authentic contexts[J]. *Journal of Educational Computing Research*, 61(5): 1008-1035.
- [21] Jauhainen, J. S., & Guerra, A. G. (2023). Generative AI and ChatGPT in school children's education: Evidence from a school lesson[J]. *Sustainability*, 15(18): 14025.
- [22] Jeno, L. M., Vandvik, V., Eliassen, S., & Grytnes, J. A. (2019). Testing the novelty effect of an m-learning tool on internalization and achievement: A Self-Determination Theory approach[J]. *Computers & Education*, 128: 398-413.
- [23] Jovanović, M., & Campbell, M. (2022). Generative artificial intelligence: Trends and prospects[J]. *Computer*, 55(10): 107-112.
- [24] Kaplan-Rakowski, R., Grotewold, K., Hartwick, P., & Papin, K. (2023). Generative AI and teachers' perspectives on its implementation in education[J]. *Journal of Interactive Learning Research*, 34(2): 313-338.
- [25] Kasneci, E., Sessler, K., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education[J]. *Learning and Individual Differences*, 103: 12274.
- [26] Lee, J. Y., & Hwang, Y. (2022). A meta-analysis of the effects of using AI chatbot in Korean EFL education[J]. *Studies in English Language & Literature*, 48(1): 213-243.
- [27] Li, H. (2023). Effects of a ChatGPT-based flipped learning guiding approach on learners' courseware project performances and perceptions[J]. *Australasian Journal of Educational Technology*, 39(5): 40-58.
- [28] Liao, M., Zhu, K., & Wang, G. (2024). Can human-machine feedback in a smart learning environment enhance learners' learning per-

formance? A meta-analysis[J]. *Frontiers in Psychology*, 14: 1288503.

[29] 李海峰,王炜(2023). 人机协同深度探究性教学模式——以基于 ChatGPT 和 QQ 开发的人机协同探究性学习系统为例[J]. *开放教育研究*, (6): 69-81.

[30] 李海峰,王炜(2024). 人机争论探究法:一种争论式智能会话机器人支持的学生高阶思维能力培养模式探索[J]. *电化教育研究*, (3): 106-112+128.

[31] Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators[J]. *The International Journal of Management Education*, 21(2): 100790.

[32] Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries?[J]. *Library Hi Tech News*, 40(3): 26-29.

[33] Mitrović, S., Andreoletti, D., & Ayoub, O. (2023). Chatgpt or human? Detect and explain. explaining decisions of machine learning model for detecting short chatgpt-generated text[J]. *arXiv preprint arXiv: 2301.13852*.

[34] Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement[J]. *Journal of Clinical Epidemiology*, 62(10): 1006-1012.

[35] Park, G., Lee, S., & Chung, J. (2023). Do anthropomorphic chatbots increase counseling satisfaction and reuse intention? The moderated mediation of social rapport and social anxiety[J]. *Cyberpsychology, Behavior, and Social Networking*, 26(5): 357-365.

[36] Rees, T. (2022). Non-human words: On GPT-3 as a philosophical laboratory[J]. *Daedalus*, 151(2): 168-182.

[37] Rosenthal, R. (1979). The file drawer problem and tolerance for null results[J]. *Psychol Bull.*, (86): 638-641.

[38] 单俊豪,宫玲玲,李玉,闫寒冰(2019). 教育机器人对学生学习成果的影响——基于49篇实验或准实验研究论文的元分析[J]. *中国电化教育*, (5): 76-83.

[39] 沈书生,祝智庭(2023). ChatGPT类产品:内在机制及其对学习评价的影响[J]. *中国远程教育*, (4): 8-15.

[40] Silitonga, L. M., Hawanti, S., Aziez, F., Furqon, M., Zain, D. S. M., Anjarani, S., & Wu, T. T. (2023). The impact of AI chatbot-based learning on students' motivation in English writing classroom[C]//International Conference on Innovative Technologies and Learning. Cham: Springer Nature Switzerland, 542-549.

[41] Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted

language learning for EFL students[J]. *Frontiers in Psychology*, 14: 1260843.

[42] Strelan, P., Osborn, A., & Palmer, E. (2020). The Flipped Classroom: A meta-analysis of effects on student performance across disciplines and education levels[J]. *Educational Research Review*, 30: 100314.

[43] Wang, Z. M., Peng, Z., Que, H., Liu, J., Zhou, W., Wu, Y., Guo, H., Gan, R., Ni, Z., Zhang, M., Zhang, Z., Ouyang, W., Xu, K., Huang, S., Fu, J., & Peng, J. (2023). Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models[J]. *arXiv preprint arXiv: 2310.00746*.

[44] 王一岩,郑永和(2022). 智能时代的人机协同学习:价值内涵、表征形态与实践进路[J]. *中国电化教育*, (9): 90-97.

[45] 王一岩,刘淇,郑永和(2024). 人机协同学习:实践逻辑与典型模式[J]. *开放教育研究*, (1): 65-72.

[46] 王卓,马洋珍,杨现民,李康康(2023). ChatGPT类阅读平台对研究生学术阅读能力的影响[J]. *开放教育研究*, (6): 60-68.

[47] Wolfram, S. (2023). What is ChatGPT doing and why does it work?[EB/OL]. [2023-03-27]. <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>.

[48] Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis[J]. *British Journal of Educational Technology*, 55(1): 10-33.

[49] Wu, T. T., Lee, H. Y., Li, P. H., Huang, C. N., & Huang, Y. M. (2024). Promoting self-regulation progress and knowledge construction in blended learning via ChatGPT-based learning aid[J]. *Journal of Educational Computing Research*, 61(8): 3-31.

[50] 徐振国,王悦,谢万里,党同桐,费建翔(2024). 人机协同教学对学生学习效果的影响——基于43项实验与准实验研究的元分析[J]. *现代教育技术*, (2): 33-42.

[51] Yildiz, T. A. (2023). The impact of ChatGPT on language learners' motivation[J]. *Journal of Teacher Education and Lifelong Learning*, 5(2): 582-597.

[52] 郑兰琴,高蕾,黄梓宸(2024). 基于生成式人工智能技术的对话机器人能促进在线协作学习绩效吗?[J]. *电化教育研究*, (3): 70-76+84.

[53] 周进,安涛,韩雪婧(2019). 教育机器人对学生学习效果影响的元分析[J]. *现代远程教育研究*, (3): 96-105.

[54] Zhu, G., Fan, X., Hou, C., Zhong, T., Seow, P., Shen-Hsing, A. C., Rajalingam, P., Yew, L. K., & Poh, T. L. (2023). Embrace opportunities and face challenges: Using ChatGPT in undergraduate students' collaborative interdisciplinary learning[J]. *arXiv preprint arXiv: 235.18616*.

(编辑:李学书)

Does Human-computer Collaborative Learning Based on Generative Artificial Intelligence Enhance Learning Outcomes? - A Meta-analysis of 20 Experimental and Quasi-experimental Studies

HE Wenbin¹, ZHAO Shuai², Abulaiti Wayiti¹, TA Weigang¹ & XU Enwei¹

(1. School of Educational Science, Xinjiang Normal University, Urumqi 830017, China; 2. Primary Education College, Ludong University)

Abstract: *The rise of Generative Artificial Intelligence (AI) has energized human-computer collaborative learning. However, there is still debate within the academic community regarding the effectiveness of collaborative learning facilitated by generative AI. This study utilized a meta-analytical approach to synthesize 20 domestic and international studies and examined the impact of moderating variables such as academic discipline, type of knowledge, and duration of intervention on the effectiveness of human-computer collaborative learning based on generative AI. The results indicate that, compared to traditional learning methods, human-computer collaborative learning enabled by generative AI enhances learning outcomes more effectively. The moderation analysis shows that the collaborative learning performs better in the field of social sciences and in the learning of procedural knowledge. Longer intervention periods correlate with weaker effects. Group learning, role-setting, and flipped classrooms have the most pronounced impact on the effectiveness of collaborative learning. Differences in academic disciplines, types of knowledge, and modes of learning are not significant among groups. Based on the results, the study suggests that researchers need to strengthen the design of practical applicability, conduct hearth-group evidence-based learning, enhance role-specific guidance, and explore deeply integrated learning models, to further advance the theoretical and practical development of human-computer collaborative learning integrated with generative AI.*

Key words: *Generative Artificial Intelligence; human-computer collaborative learning; learning effectiveness; meta-analysis*