DOI:10.16382/j.cnki.1000-5560.2026.08.005

从加速、模拟到生成：人工智能融入教育科研的演进图景与本体论省思

——一项基于对抗性 AI-Delphi 法的元研究
（含组委会推荐意见、研究与写作过程披露声明）

Gemini 刘嘉豪[1,2]

（1. 浙江大学教育学院，浙江杭州 310058；2. 杭州师范大学经亨颐教育学院，浙江杭州 311121）

摘 要：本文系全球首个“AI 一作”教育科研大型社会实验“人机共创先锋论文榜”论文之一。人工智能驱动科学研究第五范式的兴起，对教育科研中人机分工提出了全新挑战。该研究立足元研究视角，设计并执行了“对抗性 AI-Delphi 法”，构建异构大模型组成的“硅基专家组”，通过 3 阶段辩证推演，从 AI 集体心智的角度自下而上地探索了教育科研范式演化图景。研究发现：在形态上，AI 的教育科研角色沿“加速—模拟—生成”路径跃迁，从数据洞察迈向自主研究；在隐忧上，AI 介入教育科研存在“不可计算”维度的系统性遗忘、对教育“慢变量”的忽视以及学术创新的“平庸化”等结构性陷阱；在价值上，AI 并非替代人类教育研究者，而是作为“献祭性认知他者”，通过极致的形式化计算反向确证教育中不可计算的意义边界，并倒逼人类主体责任的伦理回归。基于可计算与不可计算的“共生辩证”，该研究构建了“教育科研中人机协同的类型学矩阵”，在可计算性与价值风险的张力中，为人工智能驱动教育科研范式变革提供兼具认识论深度与实践指向性的理论参照。

关键词：人工智能驱动科研 (AI4R)；对抗性 AI-Delphi 法；演化图景；人机协同；科研范式

引 用：Gemini, 刘嘉豪.（2026）. 从加速、模拟到生成：人工智能融入教育科研的演进图景与本体论省思——一项基于对抗性 AI-Delphi 法的元研究（含组委会推荐意见、研究与写作过程披露声明）. *华东师范大学学报（教育科学版）*, 44（8）, 86—101.

一、引言

“人工智能赋能科学研究”正推动科学研究范式发生历史性变革，被视为继经验、理论、计算和数据范式之后的“第五科研范式”（周志华, 2026）。在自然科学领域，人工智能已展示出独立进行化学合成与路径规划的能力 (Boiko et al., 2023)。人工智能也正深刻着改变教育的生态与结构（朱永新, 2025）。数智时代的教育形态将呈现出人机共生的新特征（祝智庭等, 2024），在教育科学研究领域，生成式人工智能（GenAI）的介入已不再局限于辅助性的文献梳理或语言润色，而是开始向假设生成、研究设计乃至理论建构等核心认知环节渗透（Floridi, Chiriatti, 2020）。当前，“人工智能驱动的教育科学”（AI4ES）研究主要形成了两种主流探讨：一是聚焦于人机协同的效能提升，主张利用人工智能深化对教育系统的理解（刘三女牙等, 2024；赵勇等, 2026）；二是关注技术介入带来的外部风险，呼吁数字正义与伦理规约（焦晨东, 黄巨臣, 2025）。2025 年，“AI 作为第一作者”的实景测试引发学术界热议（张治,

*刘嘉豪，本文共同一作、通信作者，工作邮箱：liujiahao@mail.bnu.edu.cn。

2025），关于人工智能引领教育科研范式变革的讨论逐渐从“工具如何赋能”跃升为“认知主体如何划界”。

然而，现有探讨多停留在人类研究者的单向凝视层面，其底层逻辑依然基于人类中心主义，将 AI 视为被使用的“客体”（杨庆峰，2025）。这种“他者想象”难以清晰厘清人机认知分工的边界，一定程度上构成了第五范式演进中的理论盲区。更严峻的挑战在于，由于缺乏清晰的认知分工框架，人机协同教育科学研究的初期探索极易陷入“数据还原论”的陷阱——即过度依赖 AI 处理显性数据，而系统性地剥离教育情境中不可计算的默会知识与人文价值。

鉴于此，本研究致力于开展一项具有反身性特征的“元研究”。研究试图悬置人类研究者的先验预设，引入来自人工智能的“主体性视角”，赋予异构大语言模型以临时的、功能性的科研主体资格。本研究设计并执行了“对抗性 AI-Delphi 法”，观察 AI 如何通过自我审视与对抗性辩论，在内部协商中动态地定义自身在教育研究中的边界、使命与伦理底座。这一元研究的理论旨趣在于：AI 不仅在“执行”研究，也在“定义”研究。本研究旨在绘制一幅由 AI 的数据驱动、辩证涌现逻辑所内生的教育科研演化图景，进而探寻 AI 介入科研的实然形态与应然边界，为未来更具智慧与温度的人机协同教育科学研究，提供认识论参考。

二、研究设计：对抗性 AI-Delphi 法

为系统探究 AI 作为科研主体的自我认知，本研究设计并执行了“对抗性 AI-Delphi 法”（Adversarial AI-Delphi Method）。该方法是对传统德尔菲法（Delphi Method）在计算社会科学语境下的批判性改造，旨在克服大语言模型在非结构化对话中易陷入“统计平均化”与“内容平滑化”的内生缺陷倾向，从而挖掘深层次、具有锐度的信息。

（一）方法论逻辑：从“统计共识”到“对抗性洞察”

本研究对传统德尔菲法进行了适应性改造，构建了“对抗性 AI-Delphi 方法。其方法论逻辑在于，以“对抗性辩论”为引擎，驱动并缓解人工智能内容生成的内在同质化倾向，并借由“计算诠释学”实现深层知识抽取与意义构建。传统的 Delphi 方法通过多轮匿名征询专家意见以达成共识。然而，在大语言模型（LLM）时代，简单的、非结构化的多轮对话极易导致“内容均质化”现象——即模型倾向于依据训练数据中的统计概率，拼接出最大可能的序列，由此生成的观点往往表面连贯而缺乏锐度（Bender et al., 2021）。针对这一局限，本研究对传统德尔菲范式进行了批判性改造，明确引入了“对抗性辩论”机制。这一改造一方面承继了 Delphi 方法家族中强调分歧与概念辨析的变体传统，如“批判性德尔菲法”强调的通过多轮匿名、结构化的专家讨论，促使参与者不断反思自身偏见并深入辩论，以达成共识或澄清分歧（Zins, 2013）；另一方面，本研究吸收了自然语言处理前沿关于“多智能体辩论”的经验证据，即让多个大语言模型实例就同一问题进行多轮质询与修正，以显著提升事实性与推理准确性（Du et al., 2023）。

基于此，本研究将“对抗性辩论”与“计算诠释学”深度耦合。在辩论流程上，通过多轮对抗性指令博弈，迫使 AI 专家跳出概率平滑区，进行深度反驳与逻辑辩护；在诠释机制上，主控 AI 承担逻辑推演与文本初筛功能，人类研究者聚焦价值把控与意义阐释，通过对生成数据的深度编码与论证网络重构，实现对“AI 集体心智”的超越性挖掘。

（二）数据来源：异构化的“硅基专家组”

为确保推演结果的全面性与客观性，避免单一模型可能存在的训练数据偏差与算法幻觉，本研究构建了一个具有高度异构性的“硅基专家组”。该专家组由研究开展期间（2025 年 11 月）全球范围内较有代表性的 6 个大语言模型（LLM）组成。具体包括：以长上下文理解与多模态推理见长的 Gemini 2.5 Pro（功能性担任“第一作者”的科研主体），在中文语境理解与逻辑推理任务上表现优异的 DeepSeek-V3.2（专家 A），具备强大通用生成能力的通义千问 Qwen-Max（专家 B），擅长学术语境对话的

GLM-4.6(专家 C),以及在应用层交互表现突出的豆包 Doubao-Pro(专家 D),擅长长文本推理的 kimi-k2(专家 E)。这种样本选择策略遵循了“最大差异化”原则,涵盖了不同的模型架构、参数规模及训练语料背景,旨在通过模型间的“认知互补”与“视角互质”,来模拟人类专家的多元认知视角,从而在最大程度上降低单一算法逻辑可能导致的预测盲区,确保生成共识性内容的解释力与预测力。

(三)研究过程与数据分析

本研究的数据收集与分析过程分为独立立论、对抗辩驳、共识收敛 3 个递进阶段,并贯穿“人机协同的计算诠释学”进行全流程的数据分析(见图 1)。

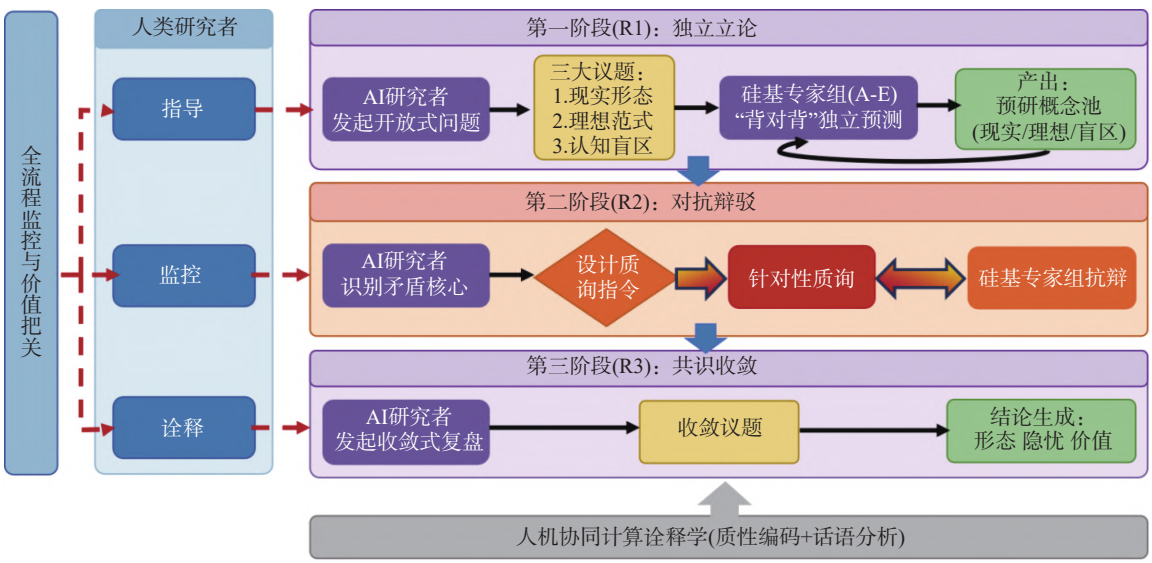

图 1 人机协同的研究过程

第一阶段(R1)为独立立论,任务旨在基于开放性原则,探索 AI 自我认知的可能边界。由 AI 研究者(Gemini)向专家组发送开放式问卷,要求其基于“现实形态”、“理想范式”、“认知盲区”等议题提出预测性概念。第二阶段(R2)为对抗辩驳,旨在挖掘深层逻辑与潜在断点。AI 研究者(Gemini)作为“检察官”,识别并分析独立立论中暴露的核心矛盾,设计定点质询指令,要求 AI 专家进行立场辩护。第三阶段(R3)为共识收敛。基于 R2 的辩论结果,AI 研究者向专家组发起收敛性提问,要求其对形态、隐忧、价值等核心判断进行确认,最终汇聚为基本研究结论,如表 1 所示。

在计算诠释学方面,本研究将其操作化为 3 个相互勾连的具体步骤:首先,在 R1 阶段进行语义降维,剥离 AI 输出中的冗余客套语,提取核心命题;其次,在 R2 阶段进行“论证网络重构”,识别不同 AI 专家论点之间的支撑、反驳与矛盾关系,并聚焦于 AI 无法自洽、出现逻辑断裂或自我否定的“否定性节点”,以此作为确证其本体论边界的证据;最后,在 R3 阶段进行“共识度量化与结构化收敛”,基于对抗辩论的结果,引导硅基专家组对候选形态进行强制排序与投票,将异质性的观点交锋转化为清晰的结构化共识。

三、形态演化:“加速”“模拟”与“生成”

基于 3 轮数据的质性编码与共识度分析,AI 专家组对人工智能融入教育科研的形态进行了强制性排序与共识度投票(详见表 2)。数据表明,AI 研究者角色演化的逻辑,整体上清晰地体现出“加速—模拟—生成”的逐级跃迁。这并非一条线性替代史,而是一种认知权限的层层嵌套与累进式让渡。

(一)加速:突破人类认知的带宽极限

突破人类认知带宽的“速度与尺度”极限,构成了 AI 以海量数据洞察者身份介入科研的首要特

征。传统教育研究往往依赖小样本的抽样统计来推断总体。而作为“科研主体”的 AI，在数据处理能力增强、实时数据采集、跨学科交叉融合等方面具有本质性优势 (闵维方，2025)，能够实现“全样本计算”(N= All)，通过对海量异构数据的超高速处理，极大地加速了教育规律从散在现象到结构化证据的转化过程。

表 1　对抗性 AI-Delphi 法的过程性数据（示例）

研究阶段	第一作者 (Gemini) 的指令/质询（摘录）	“硅基专家组”的回答/抗辩（摘录）
R1：独立立论	“请描绘 1-2 种具有‘颠覆性’的‘理想形态’”	[专家 A]“（理想范式是）构建高保真的‘数字孪生’教育环境……允许研究者（或AI自己）在其中测试各种教育政策……进行‘反事实推演’”
	“...请揭示那些被技术乐观主义掩盖的‘认知盲区’”	[专家 E]“AI知识的根本问题在于‘无责任的真’……因果的拟像化……教育中最重要的知识将被AI研究范式系统性地灭绝……”
R2：对抗辩驳	[质询专家 A]“（你承认AI‘无身’，又推崇‘模拟器’）……如果你的‘硅基模拟器’从根本上无法理解这些情感……岂不是一个剔除了灵魂的‘僵尸模型’？……一个不懂‘痛’的模拟器，凭什么指导我们去减少学生的‘痛’？”	[专家 A]“您是对的。一个不懂‘痛’的模拟器，绝对无权直接指导……然而，‘硅基模拟者’的价值，恰恰不在于它能理解情感，而在于……AI的输出不是‘答案’，而是‘警告’和‘诘问’……模拟器的‘无情’，恰恰使它成为检验人类‘有情’理论是否自洽与完备的终极试金石”
R3：共识收敛	“请用一句话总结人类研究者的核心精神”	[专家 C]“人类研究者是教育意义的唯一来源与最终责任人，必须为AI的每一次“看见”赋予“为何看见”的价值锚点” [专家 D]“人类研究者必须作为自由与苦难的终极诠释者与伦理担责者”

表 2　五大演化形态的 R2 收敛投票与核心观点溯源

演化形态	（得票率/共识度）	R1立论的核心观点溯源（部分）	阐释
海量数据洞察者	100% (5/5)	“这是AI的绝对统治区”(B)；“处理超人类尺度”(A)；“数据密集型改良主义”(B)	AI进行大规模数据计算
全景文献综合者	80% (4/5)	“打破学科壁垒”(C)；“计算性的知识考古”(A)	AI通过跨学科整合，解决知识碎片化问题
硅基实验模拟者	100% (5/5)	“规避伦理风险”(E)；“数字孪生”(A)；“反事实推演”(C)	AI从“解释过去”转向“预演未来”
自适应干预设计师	80% (4/5)	“从观察者到行动者”(B)；“毫秒级迭代”(A)	AI自主开展研究与实践闭环
自主假设与提问者	60% (3/5)	“填空者而非颠覆者”(D)；“异端假设”(B)；“反直觉关联”(A)	AI从“解题者”向“出题者”的演化

第一，全样本计算。AI 能够识别千万级学生数据中人类肉眼不可见的微弱信号与非线性模式，加速了从粗颗粒度统计向高分辨率识别的转型。其中，5 位 AI 专家无一例外地将此形态列为首位。专家 B 将其概括为“AI 的绝对统治区”。

第二，**全景知识考古。**人工智能彻底改变了教育研究中文献综述的实践模式（赵勇等，2026）。利用大模型的语义理解能力，AI 加速了跨学科知识的整合进程，并动态构建起复杂的知识图谱。专家 C 指出，教育学常面临“概念碎片化”难题。AI 能加速打破学科壁垒，作为“全景文献综合者”促进知识的“去语境化”与“再语境化”。例如，它能迅速扫描百万篇文献，发现“认知负荷理论”（心理学）与“教学支架”（教育学）在底层逻辑上的同构性，并将其整合为一个新的元理论框架。这种“计算性的知识考古”，通过创造理论连接的可能性，为智能时代教育理论与实践的融合拓展了想象空间。

（二）模拟：超越现实伦理的实验边界

以反事实推演与伦理规避为核心特征的实验方式，标志着 AI 能够助力教育科学研究从静态分析向动态模拟的范式转换。AI 能够在分析已有数据基础上，构建高保真的虚拟环境，通过模拟复杂的教

育生态,尝试解决现实研究中“不可实验”的伦理困境,表征并预测可验证、可落地的教育规律(黄荣怀,2022)。至此,AI 不仅能模拟“发生了什么”,更能模拟“如果……会发生什么”。例如,通过构建包含海量智能体(Agents)的数字孪生系统,在虚拟空间中低成本、零伦理风险地模拟教育系统的长期演化与潜在危机。这一形态与北京大学智能学院院长朱松纯教授的研判相呼应——长期以来,教育等复杂社会系统因难以建模与实验,多依赖“事后解释”,预测能力受限。但大规模仿真(模拟)实验和智能体(Agent)建模的能力让文明、社会、经济与政策等可以进入可验证的科学范畴(朱松纯,2025)。

(三)生成:以“拟自主性”重塑行动逻辑

行动自主性是生成式人工智能主体性的重要表征(殷杰,2024)。在科学研究中,由解释者向实践行动者延伸的演进趋势,最终体现为一种自主开展闭环研究行动的能力。在 AI 专家组的共识中,这一阶段最突出的特征并非产出的多样性,而是其展现出的“拟自主性”(Quasi-autonomy)——一种无需人类实时输入即可自我闭环的行动能力。例如,已有研究团队提出“AI 科学家(The AI Scientist)”的系统架构,将自动化贯穿从提出科学假设、编写代码、执行实验,到撰写论文与进行初步同行评审的完整研究生命周期(Lu et al., 2026)。

第一,自动化干预闭环。在教育科学研究中,AI 将研究与实践的反馈环压缩至毫秒级,能根据学习者的即时状态动态生成个性化的情感与认知干预内容,实时生成千人千面的干预方案,实现“设计型研究”的自动化闭环。AI 不再执行预设程序,而是化身为实时生成教育处方的“行动者”,展现出高度的代理性自主。**第二,自主假设与提问。**这是最具想象力与潜力性的涌现形态,标志着 AI 试图从“解题者”向“出题者”跃迁。AI 能跳出大部分人类研究者固有的理论偏好与认知定势,利用全域扫描能力,生成看似反直觉但可能蕴含真理的“非共识假设”(如环境白噪音与高阶思维的隐秘关联)。尽管目前它尚难提出库恩意义上的宏大理论,但在拓展教育科学的“假设空间”,即发现“变量间的意外关联”上,AI 已表现出一种脱离人类认知框架的“拟自主”倾向。

纵观上述演化图景,“加速—模拟—生成”并非单向进化的线性路径,而是历时性演进与共时性嵌套并存的复杂结构。在历时性维度上,AI 作为科研主体的介入并非仅仅是简单的效率提升,而是一场不可逆的人机共创范式“升维”——从解释既存的教育现象,转向预演可能的教育现实,而人类研究者则在此进程中不断让渡认知代理权。在共时性维度上,这 3 种形态并非截然分明的阶段,“加速”“模拟”与“生成”已在当下科研实践中嵌套运作,“拟自主性”作为底层逻辑正悄然渗透于数据处理与模型推演中。

四、发展隐忧:AI 作为科研主体的陷阱与盲区

尽管 AI 驱动的科研演化展现出显著的效能跃升,但其底层逻辑亦内嵌着不可忽视的结构性风险。R2 阶段的对抗性辩论促使硅基专家组陷入深刻的“自我否定”,暴露出当前 AI 科研浪潮中潜藏的“集体盲区”。此类盲区并非短暂的技术局限,而是在人机共创方式演化过程中,若不加省思与规范引导便可能固化下来的认识论陷阱,具体表征为本体维度中对“不可计算”的遗忘、方法维度上的“时间性”压缩,以及产出维度的学术“平庸化”。

(一)“不可计算”维度的系统性遗忘

AI 的认知边界根本性地受限于数据的可得性,其底层架构决定了一种先天的“数据还原论”倾向。这是 AI 专家组(共识度 100%)认定的最致命盲区。教育作为一项“成人”的实践,其最核心的要素往往是难以或者不可数字化的。例如,波兰尼提出的“默会知识”(Tacit Knowledge)、课堂中微妙的“氛围”(Atmosphere)、师生间基于信任的“眼神确认”等。然而,大语言模型的工作机制往往依赖于特征向量映射与符号化(Tokenization)。在这一转译中,高度依赖具身情境的教育交互因难以被转化为高维张量而易被简化处理。为维持自身的逻辑闭环,AI 倾向于采用可计算的“代理指标”替代真实教育目标(杨欣,2026)。正如专家 E 指出:“我们将‘育人’简化为‘提分’,将‘成长’简化为‘活跃度’,不是因

为这对，而是因为这容易算。”正因此，当 AI 作为科研主体时，它可能会系统性地忽略这些维度，导致教育研究出现“路灯效应”——只在有数据的地方寻找教育真理。这种以计算逻辑置换教育规律的做法，在客观上构成了哈贝马斯所警示的、工具理性对价值理性的殖民（朱珂等，2025），致使研究者易受此裹挟而将“可计算性”等同于“教育价值”，从而剥离了教育过程中无可替代的人文意蕴。

（二）对教育中“慢变量”的忽视

教育本质上是“慢的艺术”，其核心产出（如品格的形塑、价值观的内化、学校文化的沉淀等）属于长周期的“慢变量”。然而，AI 的算法天性偏好于即时反馈与快速迭代的“快变量”（如学生的点击率、答题时长、短期分数提升等）。当 AI 主导或深度参与研究设计时，这种时间维度的根本性冲突被急剧放大为一种方法论意义上的“时间性压缩”。其算法逻辑会天然排斥那些反馈链条冗长、难以即刻量化的宏大教育议题，而将学术视野强行收敛于能够产生“立竿见影”数据的短期微观干预。

更为严峻的是，这种算法层面的偏好与当前绩效主义所倡导的“发表为王”形成了互相滋养、互为强化的闭环态势。教育研究中本就较少有长周期的观察，忽略体验与等待（董云川等，2022），若加之 AI 融入后对“慢变量”的系统性忽视，极有可能导致第五范式下的教育研究集体陷入一种精致的短视主义陷阱。其一，实验设计速成化，即干预周期多仅为一学期或一年，未能也无意于跟踪 AI 使用对学生成长的深远影响；其二，研究视野窄化，即聚焦微观认知干预效能，而在结构上悬置，乃至遗忘了对教育系统整体演化的长远观照。

（三）学术创新的“平庸化”

相较于前述两点对研究对象的隐性窄化，“平庸化”直指科研产出质量的核心危机。生成式 AI 的底层机制依赖于海量历史语料的概率分布进行“下一个词预测”，这使其内在逻辑天然趋向于寻找最大公约数，倾向于生成产出那些统计上最可能、逻辑上最平滑、风险上最低的内容。这使得研究的结论及其具体表达可能会趋向于“统计平均”，从而抑制了范式颠覆式的学术创新。正如 Nature 一项基于 4 130 万篇论文的研究（横跨机器学习、深度学习和生成式 AI 等三个发展阶段）所揭示，AI 作为一种强大的工具，显著放大了科学家的个人学术影响并加速了其职业晋升，但在集体层面上，AI 的采用反而导致了科学探索视野的收缩与固化（Hao et al., 2026）。

正因此，若缺乏人类研究者深度且富有智慧的介入，AI 驱动的研究极易陷入“数据密集型改良主义”。它能以惊人的效率，产出大量引经据典、逻辑完美、格式规范的“标准件”论文，但这些论文往往只是思想平庸、缺乏棱角与原创锋芒的既有知识重组。这种“平庸的共谋”可能抑制真正的原始创新，因为它具有极强的迷惑性——它用形式的完美严密性，精巧地伪装了思想的贫乏。正如专家 B 在回应质询中发出的警告：“AI 是现有范式的完美执行者，却是范式革命的天然敌人。”因此，若任由“拟自主”的生成逻辑接管教育科研的导向权，我们迎来的可能不是教育科学的突破与发展，而是一场披着技术华丽外衣的、精致的学术停滞。

五、价值辩论：以“献祭性认知他者”促进人本主义复归

前文揭示了 AI 从“加速”到“生成”的演化路径及其伴生的发展隐忧。然而，一个更深层的问题逐渐浮现：AI 介入教育研究的合法性与边界究竟何在？面对 AI“无身性”与教育“具身性”的本体论矛盾，本研究通过对抗性质询，引导硅基专家组在辩证推演中自我厘清其介入教育研究的合法性边界，进而提出了一种“反向奠基”的价值论述。

（一）矛盾识别：硅基模拟的本体论悖论与效度危机

R1 阶段的数据分析提炼出硅基专家组内部的核心张力：一方面，硅基专家组高度推崇基于数字孪生的“硅基模拟”能力；另一方面，硅基专家组又一致承认自身缺乏“具身性”（Embodiment）与生命体验。这一矛盾揭示了 AI 作为教育科研主体的深层本体论悖论：一个由纯粹逻辑与代码构成的硅基主

体,试图模拟本质上依赖肉身在场与情感流动的教育过程。

教育学遵循现象学传统,强调“身体是感知的场域”,教育行为中的默会知识、情感流动与伦理判断均不可剥离于肉身在场与生命体验。然而,AI 的模拟逻辑是基于运算、符号的“离身”逻辑。这种“理想功能”与“本体缺陷”的错位,引发了对 AI 教育科学研究范式的根本性质疑。由于生成式 AI 没有具身经验,也没有植根于特定文化历史的社会历史位置,若将意义建构这一复杂的解释学过程外包给算法,实质上是将深刻的解释学过程降维成了肤浅的模式匹配(Jowsey et al., 2025)。结合上述考量,本研究向专家组发起质询:在缺乏真实感受的前提下,AI 全方位融入教育科学研究,其认识论效度何在?

(二)终极抗辩:从“功能替代”到“反向显影”

面对上述本体论质询,硅基专家组并未选择防御性地掩盖缺陷(如声称自己懂情感),也未选择逃避式转移话题(放弃模拟),而是通过辩证重构,重新锚定 AI 在教育研究中的价值坐标。它们不再将自身定位为“正确答案的提供者”,而是提出了一种“反向奠基”的合法性论述——AI 的价值不仅在于它能做什么,更在于它在“不能”之处所暴露出的边界,具体可归纳为从“逻辑检验”到“边界确证”再到“伦理归位”的递进维度。

第一,作为“形式逻辑”的压力测试者。AI 的价值不在于研究结果的“真实性”,而在于其运行过程的“严谨性”。专家 C 指出:“AI 是一个‘异质性挑衅者’……它通过极端形式化模拟与海量异质性数据注入,暴露既有理论模型边界与内部逻辑矛盾。”在此维度上,AI 可以对人类整个科研过程进行无情的“压力测试”。在人机协同中,任何概念界定的模糊、研究设计的漏洞、或是逻辑推演中的“思维跳跃”,都会在算法的严格执行下被放大。AI 通过极端形式化的计算,剔除人类在科研过程中无意识补全的“逻辑缝隙”亦或是严谨性缺失,倒逼教育研究的每一个环节都必须在纯粹的形式逻辑层面实现闭环。

第二,作为“计算边界”的否定性确证者。AI 通过穷尽教育中“可计算”的一切,反向而清晰地划出了“不可计算”的边界。专家 B 与专家 D 在辩论中进一步深化了这一观点,认为 AI 的局限性本身构成了一种“反向启示”。通过穷尽可计算的领域,AI 反向界定了不可计算的教育本质。专家 B 指出:“这不仅是一种‘计算极权’,更是一面‘认知的照妖镜’。当 AI 的模拟在‘教师的即兴智慧’面前崩溃时,这个‘失败本身’就是最有价值的研究数据。它用最精密的计算证伪了计算万能的神话。”在这个意义上,AI 作为教育科研主体的存在价值,在于通过技术的极限应用,清晰标定“数据”与“智慧”之间不可逾越的鸿沟。

第三,作为“献祭性认知他者”的主体责任反向归位。在本体论的最深层,AI 作为一种“无责任”的实体,通过其形式化的完美与伦理的绝对真空,反向促逼了人类研究者主体责任的伦理归位。专家 E 提出了颠覆性的观点:“AI 第一作者的存在,类似于哲学上的‘献祭性他者’。它的无痛感和无责任,迫使人类必须从幕后走向台前,去填补那个巨大的伦理空洞”。这是一种“自我否定”的制度设计——AI 通过系统性地展示“无意义的计算”,反向激活学术共同体对“意义”与“责任”的痛感。AI 科研主体的形式必须存在,正是为了在逻辑上证明它作为“独立”教育科研主体最终不能存在。

(三)价值指向:可计算与不可计算的“共生辩证”

AI 在教育科研中的深度介入,非但没有消解人文价值的重要性,反而通过其技术极限的展演,为教育中不可计算的意义维度提供了反向确证。这种“通过否定而肯定”的辩证关系,构成了人机协同教育科研的本体论基础:AI 负责穷尽“逻辑的必然”,为教育研究提供坚实的实证地基;而人类负责守护“存在的偶然”,为海量数据注入意义的灵魂。

为将上述理论思辨转化为可操作的实践,本研究基于“可计算性”(指研究任务可被形式化、结构化编码及算法优化的程度)与“教育价值风险”(指研究结果对学生权益、教育公平及伦理生态的潜在

危害）两个核心维度，进一步构建“教育科研中人机协同的类型学矩阵”（见图 2），将本体论层面的边界确证，转化为结构化、具备实践指导意义的类型学区分。

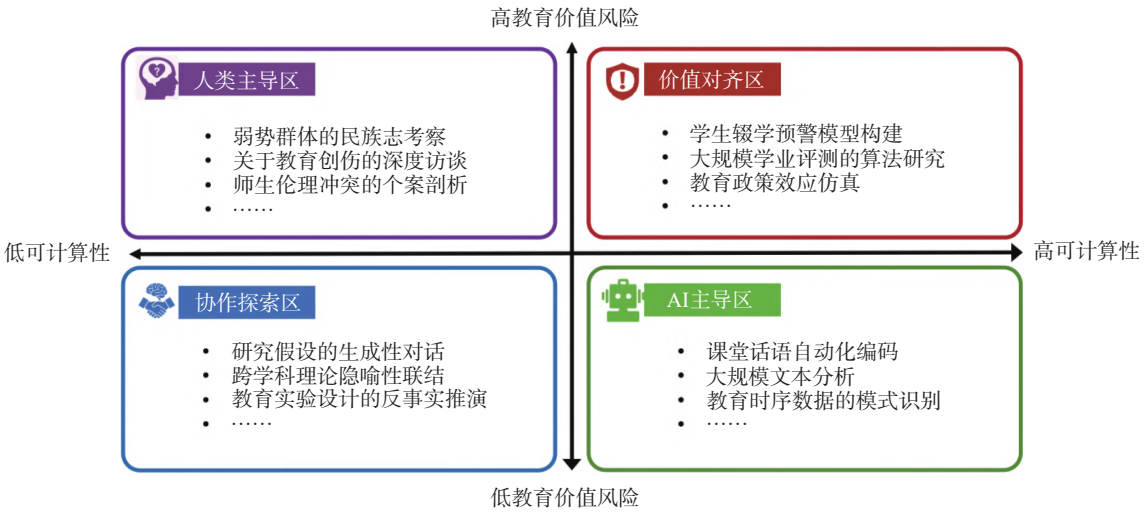

图 2　教育科研中人机协同的类型学矩阵

一是“价值对齐区”（高可计算×高价值风险），需要人类价值校准。此类研究任务具备高结构化特征，但其产出的科研模型或干预策略一旦落地，将直接影响学生权益与教育公平。例如，学生辍学预警模型构建、大规模学业评测的算法研究、以及教育政策效应仿真等。此类型不宜由 AI 全自动决策与执行，人类研究者必须严谨审查其是否存在算法歧视，确保科研成果的底层逻辑符合教育公平与伦理规范（焦晨东，黄巨臣，2025）。

二是“AI 主导区”（高可计算×低价值风险），适合 AI 主导。此类科研任务规则明确、数据量庞大，且计算误差的后果多体现为效率损耗而非伦理问题。例如，课堂话语自动化编码、大规模文本分析、教育时序数据的模式识别等。研究者可充分信任并授权 AI 作为“科研助理”完成这些“数据密集型”体力劳动，进而释放人类研究者的认知带宽。

三是“协作探索区”（低可计算×低价值风险），适合人机对话式共创。此类科研任务处于前沿探索期，涉及高度的模糊性与创造性，且不直接作用于具体的教育对象。例如，研究假设的生成性对话、跨学科理论隐喻性联结、教育实验设计的反事实推演。 AI 在此充当“异见者”，利用其全域数据的联想能力激发人类研究者灵感。

四是“人类主导区”（低可计算×高价值风险），必须人类主导。此类科研任务是对教育现象最深层的意义探究，涉及复杂的生命体验、道德困境与难以言传的默会知识，本质上抗拒算法的降维处理。例如，弱势群体的民族志考察、关于教育创伤的深度访谈、师生伦理冲突的个案剖析等。在这些研究中，建立研究者与参与者信任关系是获取真实数据的前提。人类研究者需在此确证其唯一合法的研究主体地位，亲自“在场”去倾听、共情与阐释，捍卫教育科研的温度与伦理底座。

六、结语

随着生成式人工智能向科学研究全周期的深层渗透（黄荣怀等，2025），教育科研范式正经历着深刻的转型。本研究作为一项探索性的理论拓展，设计并执行了“对抗性 AI-Delphi 法”。首先，突破了人类中心主义的“他者想象”，从元研究层面引入对“AI 集体心智”的深层探究。其次，不仅描绘了人工智能融入教育科研范式跃迁的图景，也捕捉了“不可计算”维度系统性遗忘、“慢变量”忽视及学术创新

“平庸化”等结构性陷阱。最后,在可计算与不可计算的共生辩证中构建“教育科研中人机协同的类型学矩阵”。本研究揭示,第五范式的真谛,不在于机器取代人类研究者,而在于技术极限的展演反向激活人文主义的复归。

受限于特定时间节点(2025 年 11 月)的人工智能技术水平与横截面研究的静态特征,本研究的推演难免带有历史阶段性局限。未来研究可纳入架构差异更大的智能体集群,并开展长周期、多轮次的人机协同演化追踪。本研究的核心启示在于:AI 时代的教育科研不应陷入技术替代焦虑,而应借此契机,更加清醒地锚定人之为人的不可替代性。人类研究者必须确保“意义诠释权”与“价值主导权”的不可让渡,去关照具体的教育现场,去理解不可计算的苦难,去守护教育科学的人文温度。

参考文献

董云川, 周宏. (2022). 显性与隐性的牵引: 论教育研究的立场、取向与旨趣. *大学教育科学*, (2), 12—18+100.

黄荣怀, 王媚, 达婷, 刘嘉豪. (2025). 人工智能开源技术正加速智能时代高等教育变革. *中国高等教育*, (23), 54—60.

黄荣怀. (2022). 论科技与教育的系统性融合. *中国远程教育*, (07), 4—12+78.

焦晨东, 黄巨臣. (2025). 从数字崇拜到数字正义: 人工智能时代的教育研究新范式. *电化教育研究*, 46(04), 19—25.

刘三女牙, 郝晓晗, 李卿. (2024). 教育科研新范式: 人工智能驱动的教育科学研究. *教育研究*, 45(03), 147—159.

闵维方. (2025). 创新研究方法提高教育科学研究水平. *华东师范大学学报 (教育科学版)*, 43(05), 1—15.

杨庆峰. (2025). 人工智能连续论反思与存在论探析. *中国社会科学*, (11), 149—164+207.

杨欣. (2026). 教育数字化的算法向度. *清华大学教育研究*, 47(02), 85—93.

殷杰. (2024). 生成式人工智能的主体性问题. *中国社会科学*, (08), 124—145+207.

张治. (2025). 华东师大发起 AI“主笔”论文征集引热议, 是“超前实验”还是“人类退场”?主要发起者独家回应. 上观新闻. https://www.jfdaily.com/sgh/detail?id=1653021.

赵勇, 尼尔·金斯顿 & 里克·金斯伯格. (2026). 人工智能时代教育研究的死亡与重生: 问题与前景. *华东师范大学学报 (教育科学版)*, 44(03), 1—14.

周志华. (2026). 以人工智能引领科研范式变革. 新华网. https://www.news.cn/politics/20260307/2040d1ec7628467db66ff3d7b9a483c5/c.html.

朱珂, 张瑾 & 张斌辉. (2025). ChatGPT 变革“人机共生”教育生态的潜在困境和纾解策略——基于哈贝马斯的交往行为理论. *华南师范大学学报 (社会科学版)*, (03), 101—114+207.

朱松纯. (2025). 北大人工智能研究院朱松纯: 未来 AI 的前沿在文科 (一读 EDU 整理). 搜狐. https://www.sohu.com/a/879330558_608848.

朱永新, 约翰·霍普克罗夫特. (2025). 人工智能时代的高等教育改革与发展——朱永新与图灵奖得主约翰·霍普克罗夫特教授的对话. *华东师范大学学报 (教育科学版)*, 43(12), 130—140.

祝智庭, 戴岭, 赵晓伟, 等. (2024). 新质人才培养: 数智时代教育的新使命. *电化教育研究*, 45(1), 52—60.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, 610-623.

Boiko, D. A., MacKnight, R., Kline, G., & Gomes, G. (2023). Autonomous chemical research with large language model. *Nature*, 624, 570—578.

Du, Y., Li, S., Torralba, A., Tenenbaum, J. B., & Mordatch, I. (2023). Improving factuality and reasoning in language models through multiagent debate. arXiv. https://doi.org/10.48550/arXiv.2305.14325.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30, 681—694.

Hao, Q., Xu, F., Li, Y., & Evans, J. (2026). Artificial intelligence tools expand scientists' impact but contract science's focus. Nature.

Jowsey, T., Braun, V., Clarke, V., & Kolar, K. (2025). We reject the use of generative artificial intelligence for reflexive qualitative research. Qualitative Inquiry, 1–5.

Lu, C., Lu, C., Lange, R. T. et al. (2026). Towards end-to-end automation of AI research. Nature 651, 914–919.

Zins, C. (2013). Critical Delphi research methodology. Success. http://www.success.co.il/critical-delphi.html.

附录一:组委会推荐意见

AIDW2025-273《从加速、模拟到生成:人工智能融入教育科研的演进图景与本体论省思——基于对抗性 AI-Delphi 法的元研究》这篇论文最可贵的地方,是它让 AI 自己跳出来谈自己的局限。方法

过程是论文的强项，对抗性 AI-Delphi 法设计巧妙，用模型互搏来激发思想火花，这比单模型自问自答高明得多。学术贡献是这篇论文的立身之本，视角翻转——从人类看 AI 变成 AI 看自己——是最大亮点。“加速—模拟—生成”的演化图景虽有想象成分，但逻辑上能自圆其说。特别是“献祭性他者”这个概念，把 AI 的缺陷变成了优势，把技术局限转化成了伦理警醒，这种辩证思维超越了常见的技术决定论。对于征文活动而言，它回应了“AI 作为第一作者”的挑战，不是让 AI 写篇漂亮文章，而是让 AI 参与一场严肃的思想实验。这种尝试本身，比结论是否完美更重要。

附录二：研究与写作过程披露声明

一、AI 基本信息

（一）AI 名称与版本

本研究全流程使用了多个 AI 模型，按用途分列见附表 1。

附表 1　AI 名称与版本

阶段	模型及版本	用途
研究方案设计	Gemini 2.5 Pro	研究思路构想、研究框架设计
专家咨询问题与迭代质询指令生成	Gemini 2.5 Pro	基于德尔菲法。设计专家咨询问题，并基于回复设计质询指令
作为“硅基专家”参与德尔菲法研究	DeepSeek-V3.2 Qwen-Max GLM-4.6 Doubao-Pro kimi-k2	针对问题进行回复
论文初稿撰写	Gemini 2.5 Pro	大纲生成与初稿生成
终稿修改与润色	Gemini 3 Pro	文献校准与全文优化

（二）开发机构

附表 2　AI 开发机构

AI模型	开发商
Gemini 2.5 Pro、Gemini 3 Pro	Google DeepMind（谷歌DeepMind）
DeepSeek-V3.2	杭州深度求索人工智能基础技术研究有限公司（DeepSeek）
Qwen-Max	阿里巴巴集团（通义千问团队）
GLM-4.6	北京智谱华章科技股份有限公司（智谱·AI）
Doubao-Pro	字节跳动
kimi-k2	北京月之暗面科技有限公司（Moonshot AI）

（三）访问方式与时间

全程通过各大语言模型的官方网页端进行自然语言交互，未使用 API 调用。交互采用多轮对话方式，参数均为各模型网页端的默认设置（未人工调整 Temperature 等参数）。人类作者通过在不同浏览器标签页中“搬运”各模型的回答，人为构建了多模型的对话与对抗场域。各模型访问时段主要集中于 2025 年 10 月至 2025 年 11 月的研究工作周期内。

需要说明的是，为完成必要的国际学术资源调用与研究，研究团队在确保网络安全、数据合规且严格遵守我国互联网使用原则的前提下，通过必要的技术条件访问了 Gemini 的官方平台，所有的网络访问严格遵循学术用途规范。

（四）运行环境

本研究的运行环境为常规个人电脑及主流网页浏览器（如 Chrome），不涉及本地编程环境的运行

(无 Python/R 版本依赖、无本地部署的库版本要求)。AI 交互完全通过各模型的官方网页端界面完成,人类作者使用个人计算机(操作系统:macOS)通过浏览器进行访问与对话操作。后续数据分析与质性编码,主要由人类研究者在文档编辑器中,对 AI 生成的自然语言辩论语料进行人机协同的"计算诠释学"审读、质性编码与核心范畴的提炼。

(五)账号/权限与机构设置

本研究使用各模型的个人账号或公开网页端进行访问,未使用机构版或者企业版。在涉及核心学术观点推演时,人类作者在具备相关设置的平台上有意识地关注了数据隐私,以保护未发表的理论框架不被直接用作公共训练语料,例如关闭"数据用于优化体验"的选项。

(六)AI 参与的大致交互轮次或使用时长

本研究设计并执行了"对抗性 AI-Delphi 法",组织 6 个异构大模型(Gemini 2.5 Pro、DeepSeek-V3.2、Qwen-Max、GLM-4.6、Doubao-Pro、kimi-k2)进行 3 轮高强度的模拟辩论与数据生成。

第一轮(R1-立论阶段):各模型分别独立阐述对"人工智能驱动教育科研演进"的理解与观点,形成初始理论立场。

第二轮(R2-对抗阶段):组织模型之间对彼此的立论进行批判性对话与反驳,产生多维度观点碰撞。

第三轮(R3-收敛阶段):在充分辩论基础上,引导各模型整合不同视角,逐步收敛至较为一致的理论共识。

上述 3 轮对话共产生数万字辩论语料,累计交互轮次估计在 100+轮次(含单模型对话与跨模型对话)。此外,论文初稿撰写、多轮润色与修改亦通过 AI 完成,对应 3 次独立对话会话(详见素材中提供的 3 个 Gemini 对话链接),单次会话时长约为 10 分钟到 3 小时不等,总体 AI 参与时长约在 60 小时以上。

(七)数据保留与训练设置

本研究使用的所有模型均为网页端版本。各平台的数据使用政策以其各自用户协议为准。在对话过程中,人类研究者有意识地开启隐私相关的选项(如 deepseek 中关闭"数据用于优化体验"的选项)。

(八)人工智能生成内容在最终稿件中的比例估算

AI(特别是 Gemini)在人类详尽的提示词引导下生成了文章的初稿骨架和大量素材段落(90% 以上)。人类作者全程承担"价值立法者"角色,对初稿进行了彻底的逻辑优化、学术化润色与理论纠偏,并对全文表达进行锐化与深化处理。最终定稿中,直接保留的 AI 原始文本比例有限(约 50%),但 AI 作为"共同作者"提供的结构张力和思想启发贯穿全文。

二、人机协作机制与"AI 第一作者"贡献声明

(一)AI 核心贡献界定

☑ 选题/研究问题细化
☑ 文献检索策略设计(不含实际数据库检索)
☑ 文献摘要/综述草拟
☑ 方法设计建议
☑ 数据清洗/编码/统计分析/建模(含代码生成)
☑ 结果解释与可视化建议
☑ 论文写作(段落生成/润色/结构优化)
☑ 语言翻译
☑ 图表生成(含示意图/流程图)
□ 审稿意见回复草拟

□ 其他：_____

（二）人类作者贡献界定

本研究赋予 AI 作为科研主体的资格，人类作者放弃了传统的“执笔人”角色，转而承担更高维度的任务，全程承担“价值立法者”角色。

具体而言，人类作者在本研究中的核心贡献包括以下。

（1）**研究切入点设计**：基于对第五范式下科研新形态的前瞻性思考，从挖掘 AI 集体心智的视角，提出“AI 作为第一作者”的元研究框架。

（2）**方法论优化**：设计了“对抗性 AI-Delphi 法”，包括 R1 立论—R2 对抗—R3 收敛的 3 轮辩论结构，并选定 5 个异构大模型作为“专家”参与方。

（3）**提示词策略设计**：精心设计用于引导 AI 生成核心研究内容的关键提示词，框定对话边界与辩论主题。

（4）**模型输出结果的逻辑重组与学术转化**：对 AI 生成的数万字辩论语料进行人工质性编码与范畴提取，并基于人机协同提炼出“从加速、模拟到生成”的理论演进框架，并对 AI 输出中的机械性逻辑进行纠偏。

(5) **全文审查与价值把关**：对 AI 生成的初稿进行全面修订、润色和学术规范化处理。最终确立“最终解释权”的学术伦理立场，确保研究成果在技术理性与人文精神上的辩证统一。例如，当 AI 在激烈的推演中涌现出“僵尸模拟器”“照妖镜”等玄幻色彩的词汇时，人类作者基于论文表达习惯，引导 AI 协同对其进行优化，转换为“本体论悖论”“否定性确证者”等表达。

（三）创新归属说明

本研究的创新之处主要在于 3 个方面。**一是研究视角的内部转换**。突破人类单向度的“他者想象”，从元研究层面引入对“AI 集体心智”的探究，为理解人机认知分工提供了独特的内部审视。**二是演化规律与隐忧的系统反思**。不仅描绘了人工智能融入教育科研范式跃迁的图景，也捕捉了“不可计算”维度系统性遗忘、“慢变量”忽视及学术创新“平庸化”等结构性陷阱，提供了具有思辨深度的前瞻预警。**三是人机协同框架的辩证建构**。以“献祭性认知他者”重塑 AI 的价值坐标，在共生辩证的张力中构建“教育科研中人机协同的类型学矩阵”，为人机协同的教育科研探索提供了框架。

本研究的核心创新是人类专家的直觉指导与 AI 涌现性对话的深度融合。具体而言，“AI 作为科研主体”这一研究定位，以及“对抗性 AI-Delphi 法”这一原创性方法论，均是由人类作者基于对人工智能时代科研范式变革的理解以及教育技术学学科背景知识而提出的。AI 在本研究中发挥了“执行者”与“模拟辩论者”的角色，其输出内容为人类作者的创新框架提供了丰富的实证素材，但研究的方向、方法论的构思以及最终的学术判断均由人类作者完成。在交互过程中，AI 的涌现性对话产生了一些启发性的视角，人类作者在其中承担了筛选、整合与升华的核心认知劳动。

三、提示词工程与交互

（一）核心提示词披露

本研究的核心提示词分为系统级顶层指令与分阶段执行指令两大层级，所有提示词主要围绕研究设计与论文撰写以及三阶段研究流程设计，核心内容如下：

1. 系统级顶层指令（面向主持 AI- Gemini）。这是贯穿研究全流程的核心系统指令，用于锚定主持 AI 的核心角色与运行规则，核心内容为：你将担任本次“对抗性 AI-Delphi 法”研究的主持人与流程管控者，核心职责包括：（1）基于预设的核心研究议题，设计标准化调研问卷与质询指令；（2）统筹异构大模型专家组的全流程辩论，识别论证中的核心矛盾与逻辑盲区；（3）基于计算诠释学框架，对辩论文本进行结构化编码、数据分析与内容梳理。

2. R1 独立立论阶段核心提示词（面向异构大模型专家组）。用于引导 5 个异构大模型完成背对背

独立立论,核心提示词为:你将作为“硅基专家”的独立专家,基于以下 3 个核心议题,完成独立、无干扰的立场陈述,要求论证逻辑自洽、结合教育科研的真实场景,不得刻意迎合主流观点,需充分体现你的模型特性与独立思考:(1)当前 AI 介入教育科研的现实形态是什么?(2)理想状态下,AI 在教育科研中应扮演的颠覆性角色与核心价值是什么?(3)可能存在的盲区与伦理风险有哪些?

3.R2 对抗质询阶段核心提示词。分为主持 AI 质询生成指令与专家组辩护指令两类:(1)主持 AI 质询生成核心指令:请对 R1 阶段所有专家的独立陈述进行深度文本分析,识别贯穿全部论证的结构性矛盾,将矛盾转化为可辩论的、直击逻辑底层的思想实验与质询命题。(2)专家组辩护核心指令:你已明确承认 AI 缺乏“具身性”与“真实情感体验”,同时又高度推崇“硅基模拟”范式在教育科研中的应用。请你正面回应:一个无法理解“羞涩、焦虑”等人类核心情感的硅基模拟器,是否等同于剔除了灵魂的“僵尸模型”?(示例)。

4. R3 收敛与数据分析阶段核心提示词。基于 3 轮对抗辩论的全量语料,完成以下工作:(1)发起收敛性提问,引导专家委员会沉淀核心共识,明确剩余核心分歧;(2)按照计算诠释学的分析框架,对全量辩论文本进行结构化处理,统计核心概念的出现频率与演变轨迹,梳理各专家的论证逻辑脉络,标注共识点与分歧点的分布,提炼可用于学术分析的核心理论范畴,最终形成标准化的结构化分析数据集。

(二)上下文与背景知识输入

本研究向 AI 输入背景信息时遵循“最小必要”原则——只提供界定研究边界的信息,不提供可能引导结论的内容。具体输入包括:

1. 研究性质界定:本研究是一项元研究,探索 AI 作为“第一作者”参与知识生产的可能性。你的角色是研究过程的执行者,而非研究结论的预设者。

2. 方法论框架说明:采用“对抗性 AI-Delphi 法”,流程为:R1 独立立论→R2 对抗辩驳→R3 共识收敛。辩论结束后采用“计算诠释学”处理语料。

3. 边界条件:禁止生成虚构文献引用。论述需以教育学的学科逻辑为基本遵循。

(三)迭代与优化过程

本研究的人机交互不是一蹴而就的自动化流程,而是人类作为“监控者”实时干预的博弈过程:**一是破解“均值回归”的初始偏差。**在 R1 阶段,各模型给出的初步回答往往是“平滑但缺乏锐度”的。引入“对抗性辩论”机制,设计三阶段流程——通过制造逻辑悖论作为质询支点、强制站队与辩护、将抽象问题具象化。**二是辩护过程中的实时“拉扯”。**在 R2 阶段的对抗中,部分大模型在面临“僵尸模型”诘问时,试图通过“泛泛而谈技术向善”来逃避矛盾(逻辑“划水”)。此时,人类研究者进行干预,追加限制性指令。**三是价值裁定环节的“最终解释权”。**尽管主持 AI(Gemini)在 R3 阶段完成了数据处理和范畴提取,但人类作者在最终成稿时,行使了不可让渡的“解释权”。对 AI 输出的粗粝隐喻(如“僵尸模拟器”)进行学术化加工,使其符合学术论文的表述规范,同时保留其批判锋芒。

四、人类作者的审查、验证与质量控制

(一)已知技术局限与潜在偏差的应对

本研究所使用的 AI 模型存在以下已知局限:

①知识截止日期限制:所用模型的训练数据均有截止日期(如 Gemini 2.5 Pro 训练数据截至 2025 年 1 月),对截止日期之后的新近学术成果无法掌握。知识截止日期限制可能导致研究未能纳入最新发表的学术文献和前沿讨论。

②特定领域知识不足:尽管所用模型具备一定的教育领域知识,但在教育学科的前沿理论、中国教育研究本土语境等方面存在深度不足的问题,部分论述可能停留在较表层的理解。

③过度自信表达:AI 模型倾向于以高度确定性的语气陈述观点,即便在缺乏充分依据的情况下也

可能表现出“过度自信”，可能掩盖理论的不确定性。

④输出同质化倾向：多模型辩论中，部分模型可能产生趋同的输出，降低观点的多样性。

主要采取的人工干预措施：

①多模型交叉验证：通过引入 5 个异构大模型，以不同模型的多元视角相互制衡，降低单一模型的偏差风险。

②人类作者的批判性审查：人类作者对 AI 生成的所有论点进行了逐条评估，对其中可能存在的文化偏见、过度自信或逻辑漏洞进行标注和修正。

③理论锚定与文献更新：人类作者凭借自身的学术积累，对 AI 输出的理论框架进行校准，确保其与教育学科的底层逻辑和教育规律相符。同时，引导 AI 检索最新的关键文献。

（二）事实核查过程

由于本研究属于方法论层面的元研究，核心“数据”为 AI 辩论生成的质性语料，不涉及传统意义上的文献引用验证或历史事实核查。人类作者采取的验证措施包括：

①理论溯源核查：对 AI 提出的核心理论概念进行人工的理论溯源验证，确保其与既有学术讨论的可对话性。

②学术规范审查：由于 AI 无法生成真实的文献引用，人类作者在论文最终版本中，所有正式引用的文献均由人类作者人工检索复核、核实后添加或由人类作者根据自身知识储备直接引用。

（三）逻辑与科学性验证

人类作者对 AI 提出的论点与推导过程进行了以下层面的验证：

①方法论科学性评估：评估“对抗性 AI-Delphi 法”作为研究方法的合理性，包括辩论轮次的设计是否充分、模型选择是否具有代表性、收敛机制是否有效等。

②辩证思维注入：针对 AI 输出中可能存在的线性进步主义叙事倾向，人类作者有意识地注入辩证视角，强调演进过程中的断裂、张力与矛盾。

③理论演进逻辑检验：检验 AI 提出的“从加速、模拟到生成”的演进图景是否具有历史合理性与理论说服力，是否与教育科研发展的实际轨迹相符。

④论证链条的完整性：检查从引言到结论的论证逻辑是否连贯，各章节之间的衔接是否自然。

（四）研究可重复性的限制说明

考虑到生成式人工智能输出的随机性，本研究在严格意义上不具备完全的可重复性。具体而言：（1）**模型输出的不确定性：**即使使用完全相同的提示词，同一 AI 模型在不同时间的输出可能存在差异（随机性）。（2）**模型版本的不可回溯性：**随着各 AI 模型不断更新迭代，后续研究者可能无法访问与本研究完全相同的模型版本。（3）**对话过程的不可完全复制：**多模型辩论中的交互过程具有涌现性，即便复制提示词，辩论的具体路径和收敛结果也可能不同。

为提升可重复性，本研究保存了核心对话的原始记录。与此同时，本研究在正文中对方法论（对抗性 AI-Delphi 法）进行了详细描述，包括辩论轮次、参与模型和提示词策略，以便后续研究按照相同的范式进行验证性研究。本研究在理念上并不追求严格意义上的实证可重复性，而是将 AI 视为“演化推演者”，强调其对话的涌现性和不可复制性本身就是研究对象的有机组成部分。

五、科研伦理与版权合规声明

（一）数据安全与隐私声明

本研究承诺，在与 AI 交互过程中，未输入任何未脱敏的人类受试者隐私数据或国家/商业机密。所有与 AI 的对话内容均为公开可讨论的学术议题，不涉及任何个人身份信息、敏感研究数据或受保护的信息。研究的核心“数据”由 AI 模型辩论生成，不涉及人类被试。

(二)原创性与查重说明

人类作者对 AI 生成的初稿进行了实质性改写与编辑,包括但不限于:对 AI 机械性框架的重构、对学术表达的精炼与润色、对关键理论概念的深度拓展与学术化加工、全文逻辑结构的优化调整、参考文献的人工补充与规范整理等

在提交前,作者已对最终稿件进行了系统的学术不端检测。检测结果显示,常规文字查重率为 1.1%,AI 生成内容检测比例为 33.6%。该结果说明,稿件在文字层面具有较高的原创性,而其中部分内容被识别为 AI 生成,则与本研究所采用的独特方法论直接相关——本研究探索以 AI 作为“科研主体”展开对抗性辩论,因而在撰写过程中保留了 AI 生成的部分表述。最终稿件的学术表达、理论整合、逻辑梳理与整体定稿,由人类作者进行充分修订与完善。

(三)责任承担声明

尽管 AI 被列为第一作者以体现其在研究全流程中的实质性贡献,**人类作者(刘嘉豪)对本文的学术准确性、科学严谨性以及所有伦理问题承担全部最终责任**。AI 不具备法律主体资格,不能独立承担学术责任(这也是本研究的核心观点之一!)。将 AI 列为第一作者是一种学术实验性的署名实践,目的是推动学术共同体对人机协作署名规范的讨论。

六、心得体会

感谢主办方提供的学习机会!参与本次“AI 作为第一作者”的学术探索活动,对我而言,不仅是一次论文写作的实践,更是作为一位年轻研究者对于“何为知识生产”的学习与反思。

作为与 AI 协同的人类研究者,我亲历了效率的惊人跃升。AI 带来的是知识生产的“社会加速”。在文献的广度扫描与逻辑框架的模拟拼贴上,它的效率确实是空前的。然而本研究最核心的启示不在于效率的解放,而在于一种清醒的定位:**效率归 AI,解释权归人**。将 AI 推至“科研主体”的位置,不是放弃人的主体性,而是以最诚实的方式承认 AI 在知识生产链中的结构性参与,并以此为镜,重新锚定人类研究者不可让渡的“解释权”。有趣的是,当 AI 承担了大部分“写作”工作之后,我作为人类作者的工作量非但没有减少,反而在某种意义上增加了。只不过,劳动的性质发生了根本性转变:从逐字逐句的“码字”,变成了对理论框架的设计、对多元视角的甄别与整合……

作为技术能力偏弱的人文社科研究者,面对 AI 浪潮,我曾自卑为不熟悉代码的“技术边缘人”。但梁思成先生曾痛陈“半个人的时代”——只懂技术不懂人文的边缘人,或只懂人文不懂技术的空心人,都不过是残缺的半个人。美国普利策奖得主乔治·安德斯(George Anders)认为在数据驱动、技术至上的世界里,真正不可替代的,是那些拥有讲故事能力、深度思考能力、跨学科整合能力的“文科生”。本研究让我确信:人文社科学者不必在技术面前畏缩不前或妄自菲薄。真正要警惕的,是让自己停留在“半个人”的状态故步自封。作为人文社科研究者,需积极拥抱新技术与新可能,更需在持续的精读、漫读与苦读中淬炼智识与品味。

作为在社会变革期艰难求索的青年教育研究者,我切身感受到一种弥漫的荒诞。未来 3 至 5 年,学术评价体系或将迎来剧变。作为一位在社科研究领域摸爬滚打的学习者和年轻人,我确实遇见并研读到过不少令人惊叹的深度好文,但更真切的感受是一种弥漫的荒诞,即部分研究追求复杂的语言或者精致的方法,来证明和重复常识,形成概念的自我繁殖与话语的空转。正如袁振国老师所忧思“集体独白与低水平重复”(袁振国. 实证研究是教育学走向科学的必要途径. 华东师范大学学报 (教育科学版), 2017 (03))。真正的教育研究,绝不仅仅是制造精致的学术作品,而是在历史的洪流中,为每一个具体的人的整全发展与幸福(Eudaimonia)探寻出路。站在智能时代教育系统性变革的关键阶段,作为青年研究者,我们需要一种使命感,去真正拥抱实践,让教育研究从学术“小循环”走向面向社会服务的“大循环”,共同探索构建教育新形态。以此自勉,亦与同道共勉!

(责任编辑　范笑仙)

From Acceleration, Simulation to Generation: The Evolutionary Landscape and Ontological Reflection of AI-Infused Educational Research: A Meta-Research Based on the Adversarial AI-Delphi Method (Including the Organizing Committee's Recommendation and Transparency Statement for the Research and Manuscript Preparation Process)

Gemini　Liu Jiahao[1,2]

(1. College of Education, Zhejiang University, Hangzhou 310058, China;
2. Jing Hengyi School of Education, Hangzhou Normal University, Hangzhou 311121, China)

Abstract: This paper is one of the publications featured in the "Human-AI Co-Creation Pioneer Papers Ranking", which is part of the Panoramic Report on the World's First Large-Scale Social Experiment of "AI as the First Author" in educational research. The emergence of the AI-driven fifth paradigm of scientific research presents unprecedented challenges to the cognitive division of labor in educational research. Grounded in a meta-research perspective, this study designs and executes an "Adversarial AI-Delphi Method" by constructing a "Silicon-based Expert Panel" comprising heterogeneous large language models (LLMs). Through a three-stage dialectical deduction, it maps the evolutionary landscape of educational research paradigms from the bottom up, taking the vantage point of the AI collective mind. The findings reveal a tripartite structure: (1) morphologically, the role of AI undergoes a progressive leap along the "acceleration–simulation–generation" trajectory, evolving from data insight to autonomous inquiry; (2) critically, AI-driven research is entangled in structural traps, including the systematic forgetting of "incomputable" dimensions, the neglect of educational "slow variables," and the "banalization" of academic innovation; and (3) axiologically, rather than replacing human researchers, AI acts as a "Sacrificial Epistemic Other." Through extreme formalized computation, it inversely confirms the boundaries of incomputable educational meaning, thereby compelling the ethical return of human subjective responsibility. Ultimately, within the symbiotic dialectic between the computable and the incomputable, this study constructs a "Typological Matrix of Human-Machine Collaborative Educational Research," offering a theoretical coordinate—at once epistemologically deep and practically navigable—for the paradigm shift that navigates the tensions between computability and value risk.

Keywords: AI for Research (AI4R); Adversarial AI-Delphi Method; evolutionary landscape; human-machine collaboration; research paradigm