

# 人工智能助力教育变革的风险研判、归因分析与生态治理

张羽, 杨子豪, 覃菲

(清华大学教育研究院, 北京 100084)

**[摘要]** 人工智能助力教育变革已成为全球教育发展的重要议题,其技术潜力不断释放,应用场景日益丰富。然而,这一进程也伴随着复杂的挑战和风险,亟须系统性生态治理。研究剖析了 AI 助力教育变革的三类风险(主体性、安全性、发展性风险),并聚焦四种生态主体(政府、AI 技术研发者、学校和企业)开展归因分析,揭示了三类风险的多层次原因链条,涉及四重逻辑:制度逻辑,法律与专业的约束缺失;技术逻辑,算法与系统的内在局限;教育逻辑,守正与守旧的实践挑战;商业逻辑,效率与利益的价值扭曲。为应对上述问题,研究提出 AI 助力教育变革的生态治理框架,旨在为 AI 与教育的良性共生与可持续发展提供政策指引,主张以政府为主导,完善顶层架构;增加企业参与,规范行业管理;发挥教育引领作用,提升主体自觉。

**[关键词]** 人工智能;教育治理;人工智能教育;深度融合;教育变革

**[中图分类号]** G434 **[文献标志码]** A

**[作者简介]** 张羽(1983—),女,山东曹县人。教授,博士,主要从事教育经济学、心智、脑与教育以及人工智能与教育研究。E-mail:zhangyu2011@tsinghua.edu.cn。

## 一、引言

人工智能(Artificial Intelligence, AI)与教育的深度融合已成为全球教育发展的重要趋势。以 AI 为代表的新一轮科技与教育的关系,已迈入系统性融合的新阶段<sup>[1]</sup>。人工智能助力教育变革不应是“AI 赋能教育”的单向赋能思路,也不应是“AI+教育”的以 AI 技术为主导的融合思路,而应充分发挥教育主体的能动作用,以教育实践的真实需求为主导,引领 AI 技术的发展与应用,实现教育和 AI 的双螺旋式共进。这种深度融合要求从治理层面对 AI 教育进行全面审视。充分考虑教育场景的特殊性,以满足 AI 与教育的深度融合需求。为了达成此目标,应深入剖析 AI 助力教育变革受到的新约束、产生的新风险,挖掘风险生成的深层原因,构建指向 AI 助力教育变革的生态治理体系。

## 二、AI 助力教育变革的风险研判

在 AI 助力教育变革背景下, AI 正在潜移默化地改变传统的师生关系和学习模式。基于技术的权力结构有可能强化 AI 在教育中的“权威性”,进而放大 AI “黑箱”式算法潜藏的技术缺陷,并暴露人类的道德缺陷,最终为教育生态带来多元挑战(如图 1 所示)。



图 1 AI 助力教育变革风险框架

基金项目:2023 年北京市教育科学“十四五”规划优先关注课题“运用现代信息技术深化教育教学质量评价的研究”(课题编号: BGEA23019)

### (一)主体性风险

第一, AI作为一种强大而新颖的“智慧外脑”,为学习者提供了便捷的知识获取渠道,但也带来了路径依赖风险。AI依托于大数据和高算力支撑,其核心在于通过计算路径和统计概率完成相关学习信息的生成与推送。这一算法驱动的嵌入模式在表面上提升了学习的智能性和精准性,但实际上在潜移默化中重构了学习者的认知路径。过度依赖 AI 系统,会使学生的学习决策逐渐受制于 AI 的计算性选择,而非基于内在思考的自我判断。

第二, AI 通过构建“中心化”的共性特征、“引导式”的智能推荐来规范学习者的认知路径,过度使用则会造成认知束缚风险。AI 的“生成”基于群体共性特征的设定,其核心在于通过算法构建出一个统一的“中心”来代表复杂人类群体的共性。学生受制于算法规则框架下的信息循环中,接收与自身偏好一致的信息,容易形成一种“既定认知”的路径,引发认知场域中的“过滤气泡”和“信息茧房”的效应,限制了个体的认知自由与创造力<sup>[2]</sup>。

第三, AI 的算法模型推演主导了诸多学习情境的表征,其不当使用可能带来学习异化风险。正如奥尼尔所言,“没有一种模型能囊括现实世界的所有复杂因素或者人类交流上的所有细微差别”<sup>[3]</sup>,仅依靠高度简约的算法模型难以还原全部的教育内涵,甚至会裂解教育的复杂性<sup>[4]</sup>,这主要体现在学习情境和教学互动两方面。在学习情境方面,某些针对应试教育的 AI 技术可能被滥用于强化标准化测试和机械化训练,忽略了教育中的人文属性<sup>[5]</sup>,导致教育活动的简单化和模式化。在教学互动方面, AI 的情感支持依赖于预设的算法逻辑,难以真正替代人类教师在复杂情感互动中的细腻关怀和价值引导。

### (二)安全性风险

第一,数据泄露风险。AI 技术的研发和应用需要海量数据的支持,带来了潜在的数据泄露风险。特别是教育场景下,师生与 AI 的互动涉及大量个人信息和学习数据,若缺乏系统严密的数据加密、访问控制和防护的机制,容易造成敏感数据被不当访问或泄露,给用户带来严峻的安全隐患<sup>[6]</sup>。

第二,数据偏颇风险。AI 依赖大量的数据集和开发者的设计理念,其中若含有民族、地域等偏见,就可能生成偏离我国学生认知与价值观的教育内容<sup>[7]</sup>。鉴于数据库信息可能存在事实性错误、概念混淆等问题, AI 系统也常会生成一些错误或荒谬的信息,“AI 幻觉”可能导致对学生的潜在误导。并且其内容生成

过程缺乏透明性和可解释性,犹如一个“黑箱”,因此,这些错误信息难以被及时识别与精准纠正。

第三,技术滥用风险。学生可以轻易地利用 AI 进行学术作弊,例如,获得作业答案、生成论文内容甚至编写代码,从而绕过对知识的深入理解和自主学习过程<sup>[8]</sup>。这种滥用行为挑战了基本的学术诚信原则。除此之外,学生利用 AI 从事犯罪活动的问题也应受到关注,如利用 AI 诈骗同学、散布虚假信息等。

### (三)发展性风险

第一,技术悬浮风险。AI 技术逐渐为越来越多的教育情境赋能,但其适用性却面临“悬浮化”困境<sup>[9]</sup>,形成了“试图接近但又无法融入”的若即若离的、悬浮化的状态<sup>[10]</sup>。具体表现为,能够满足信息整合、搜索和简单推荐的功能,能够初步服务于浅层个性化学习但缺乏对教育场景复杂性的深刻理解,无法精准识别和响应教师在教学设计、课堂管理等方面的动态需求,同时,对于学生的学习情绪、元认知能力及长期学业发展等深层次需求,现有技术亦显得力不从心。

第二,教育公平风险。AI 技术的广泛应用需要完善的数字基础设施与资源支持。这些资源在不同地区的可及性差异明显。在农村地区和经济欠发达地区,数字化建设能力薄弱,学生无法获得与城市发达地区同等的 AI 教育支持,从而导致教育资源分配的不平衡。技术的复杂性和高门槛也进一步拉大了不同群体间的差距。在这一过程中,“数智鸿沟”逐渐被隐藏甚至合理化,那些被边缘化的群体难以意识到自身所处的劣势,形成隐形的“数智难民”群体<sup>[11]</sup>。

第三,过度监控风险。AI 对教育场景中行为和表现的实时监控,使得教师和学生在学习活动中被置于高度透明化的“数据监控”环境中,其教学过程也被极度细致地量化评估<sup>[12]</sup>。对于教师而言,如果其每个行为都与评价挂钩,技术支持可能将异化为技术压力,进而弱化了教师的创造性和灵活性。对于学生而言, AI 的持续追踪,可能导致他们在“被监督”的压力下学习,将学习视作一种被审视和量化的任务,而非内在驱动的探索过程,进而减弱其学习意愿和创造力。

## 三、AI 助力教育变革风险的归因分析

AI 助力教育变革风险的产生并非由单一因素驱动,而是深刻根植于多元主体间责任与互动的缺失。在 AI 助力教育变革生态中,政府、AI 技术研发者、学校与企业构成了四个核心主体,而其背后分别体现的制度逻辑、技术逻辑、教育逻辑和商业逻辑,各自蕴含潜在风险。四重逻辑相互交织,共同构成了 AI 助力教育变

革风险的多层次原因链条(如图2所示)。

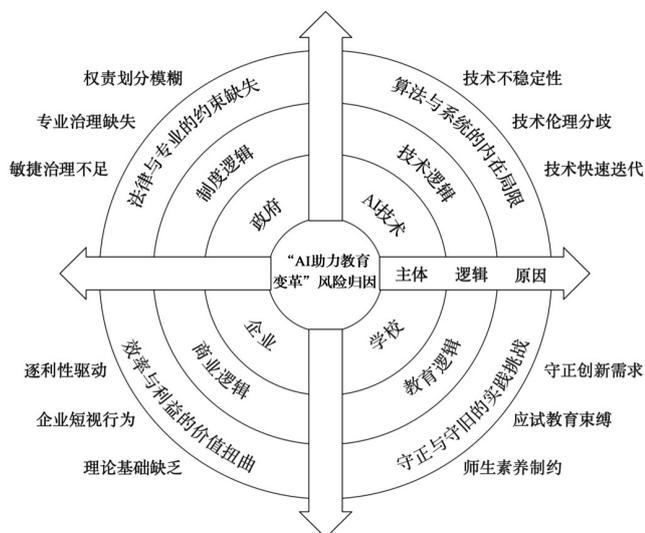


图2 AI助力教育变革风险的归因分析

### (一)制度逻辑:法律与专业的约束缺失

我国的数据治理体系主要建立在《中华人民共和国数据安全法》《中华人民共和国网络安全法》和《中华人民共和国个人信息保护法》的基础之上,但针对AI在教育场景中的具体应用,仍然存在一定空白,这是造成AI助力教育变革风险的根本原因。

在法律层面,目前治理体系和政策中权责划分的模糊性成为引发风险的重要源头。现有政策如《生成式人工智能服务管理暂行办法》主要关注技术企业的合规运营,但未能针对教育场景中学校、技术与政府等主体的权责进行有效分配,导致监管与执行责任不明。此外,现有政策体系尚未建立起分层分级、多主体合作的协同治理体系。以教育数据监管为例,虽然部分区域尝试引入企业与学校共同参与监管,但由于缺乏明确的分级分层监管职责划分,这种合作形式通常流于形式,难以形成有效的风险治理闭环<sup>[13]</sup>。

在专业层面,针对教育场景的专业治理机构和制度设计存在缺位的问题。现有法律框架主要以通用型数据和技术治理为导向,难以完全满足教育领域的学生发展性、多样性、公平性等属性要求,泛化的治理框架也难以有效覆盖AI教育应用中的具体细节。学生主体的特殊性也要求AI在面向学生群体应用时应有特殊的规范。例如,学生学习的个性化并不能沿用消费个性化推荐的底层逻辑,需要基于教育的专业知识和大量科学的实证研究,在制度和标准上提出明确标准和规范。此外,面对当前管理容易出现“一管就死,一放就乱”的问题,政府需要快速组织专业人员对新产品、新应用的科学研究,并制定多样化的标准。

### (二)技术逻辑:算法与系统的内在局限

现阶段AI技术不成熟引发了技术内生的不稳定性和复杂性问题,贯穿于AI系统的研发、设计、制造、应用等多个阶段,是导致AI助力教育变革风险的内生根源。

算法黑箱、数据偏见等不稳定因素对AI应用的安全性及可控性提出挑战。不同于传统基于规则执行的算法,现代机器学习算法具备自我学习和自主决策的特性,能够从庞大的数据集中获取模式,无须人类直接干预<sup>[14]</sup>。然而,这种数据驱动的决策过程由于其“黑箱”特性,使得外界难以理解系统从输入到输出的逻辑,使安全性验证更加困难<sup>[15]</sup>。数据不仅是AI优化的重要基础,更是其实现公平合理决策的关键<sup>[16]</sup>。如果数据质量不佳,AI系统的性能和可靠性也会随之下降。

AI技术的伦理界定问题呈现出深层的理论分歧,增加了治理中权责关系确认难度。传统观点坚持主客二元,将AI视为实现教育目的的辅助物,强调人类主导地位的不可动摇性<sup>[17]</sup>。另一种观点则主张主体间性关系,认为AI在智能化和自动化环境下具有潜在主体性,可能将逐步削弱人类作为唯一教育主体的地位<sup>[18]</sup>。两方争论尚未统一,AI的伦理地位也尚未明确。可能导致在法律和伦理框架下难以准确界定AI责任归属,进而导致治理中的责任转移、伦理真空等深层次问题。

AI技术的快速发展和在教育领域的不断拓展进一步加剧了治理上的难度和风险。当前,AI在教育中的应用从个性化推荐系统到智能教师助手,逐渐渗透到课程设计、教学反馈、学情分析等多重场景,其功能与影响力不断扩展。然而,技术的快速迭代往往伴随着监管机制的滞后,导致AI系统在教育中的应用缺乏统一标准,容易出现不受控的风险<sup>[19]</sup>。

### (三)教育逻辑:守正与守旧的实践挑战

目前,教育改革已经迈入深水区,其追求守正创新的价值目标与长期存在的顽瘴痼疾之间的复杂博弈,共同构成了制约AI助力教育变革在治理上的结构性根源。

学校教育强调守正创新,即在坚守教育的价值规范和社会责任的基础上,以立德树人和培养创新人才为目标,实现五育融合的全方位创新<sup>[20]</sup>。这要求其对新技术的应用保持审慎态度<sup>[21]</sup>,从学生发展视角考量技术应用的正当性和合理性,避免技术对教育进行单向度的改变。然而,在AI融入教育的过程中,尽管已有诸多创新应用和模式,但这些技术在教学安全性和

效果上的实证研究仍较为有限,尚不足以提供全面的证据来支撑其长期应用的合理性与安全性,也难以确保这些应用是否符合教育的核心价值和社会期望,使得学校教育在 AI 应用中存在较大的试错和调整成本。

当前教育体系中仍然存在应试教育倾向的顽瘴痼疾,对 AI 助力教育变革的应用与治理形成了系统性制约。一些 AI 教育产品,如智能作业平台和在线测评系统,主要以提升考试成绩为核心目标,通过精准定位学生薄弱环节并反复强化训练,在短期内优化学业表现。这种“刷题模式”将学习过程简化为单一的应试训练,忽视了对学生批判性思维、探究能力和跨学科能力的培养。同时,AI 系统的成效评价过度依赖量化指标,如正确率与考试分数,这与教育实践中对学业成绩的过度追求高度契合,进一步助推了教师和学校对分数的聚焦,而非对学生全面发展的关注。这种模式使“最先进的技术”服务于“最落后的教育”,不仅与五育融合的教育目标背道而驰,也制约了 AI 技术在教育领域的长远发展。

教师与学生的数字素养发展滞后是当前教育发展中另一个亟须解决的问题。作为教育的核心主体,教师和学生的数字素养和技术适应能力在快速发展的技术环境下相对滞后,不仅拖慢了技术与教育深度融合的步伐,也暴露了教育体系在这方面培养的不足。

#### (四)商业逻辑:效率与利益的价值扭曲

企业作为市场主体的商业性与教育的公益性之间存在天然张力<sup>[22]</sup>。AI 教育企业往往以商业利益为导向,忽视对教育规律的尊重与长期育人目标的实现,这是 AI 助力教育变革风险的外源性原因。

在市场环境下,教育企业在产品开发和提供服务提供中通常以商业利益最大化为核心目标。这种逐利性动机使得企业在产品设计中更倾向于迎合市场需求,追求用户数量和收入的快速增长,而忽视教育的规律和目的。AI 教育企业普遍缺乏扎实的教育理论基础,这进一步加剧了治理风险。许多企业的研发团队主要由计算机科学、数据科学等领域的人员组成,对教育学、心理学等学科的理解相对薄弱。这种技术主导型开发模式容易导致产品设计中教育本质的忽视,使得 AI 教育产品难以从根本上满足教育场景的需求。

行业内激烈的市场竞争往往会加剧企业的短视行为,成为滋养风险的温床。部分企业追求“政策性创新”,只做表面文章而忽视了深层的创新设计<sup>[23]</sup>。为迅速占领市场,许多企业采取“先上线、后优化”的策略,优先追求产品上线速度,而非产品质量和用户体验的

持续优化,导致产品在实际使用中暴露出大量问题,如算法不稳定、数据隐私保护不足等<sup>[24]</sup>。

## 四、AI 助力教育变革的生态治理体系

基于风险研判和归因分析,本研究提出 AI 助力教育变革生态治理体系,以迎接深度融合的机遇与挑战。AI 助力教育变革的生态治理应以坚守教育价值为核心理念,坚持以政府为治理核心,企业和教育用户多方协同参与的治理进路。三方协同联动,构建信息循环与反馈机制,确保治理体系能够根据动态需求持续优化,实现生态系统的自治演进与动态平衡。该治理体系强调三方面的建设:完善顶层架构、规范行业管理和提升主体适应(如图 3 所示)。

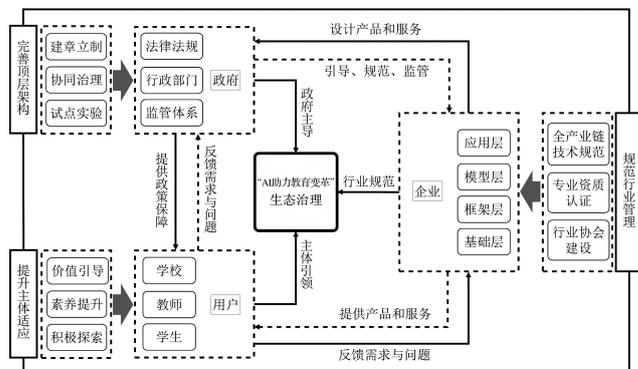


图 3 指向 AI 与教育深度融合的行业治理体系

#### (一)完善以政府为主导的科学管理与协同治理

教育作为一种公共品,AI 助力教育变革生态发展不能完全依赖市场经济规律,需要政府通过宏观调控和政策制定进行全面引导。具体而言,政府应从建章立制、协同治理与试点实验三个层面入手,为 AI 助力教育变革风险治理提供系统性解决方案。

第一,科学建构规避 AI 助力教育变革风险的法治基础。一方面,应加快完善与教育数据相关的隐私保护和风险防控法规,明确规定教育数据的生命周期管理策略,规范数据的知识产权归属<sup>[25]</sup>,确立定期审查制度,确保数据的合法使用与风险可控<sup>[26]</sup>。另一方面,应制定涵盖行业标准与产品规范的市场准入审查制度,对 AI 教育产品的教育性、算法透明性、数据安全性等方面提出明确要求,确保技术供应商在产品开发与提供服务过程中有章可循。可考虑设置 AI 产品在教育领域的准入审查机构,参考新药上市的实验和审查流程,确保对教育领域 AI 产品的有效风险防范。

第二,多方联动的协同治理是政府实现顶层引导的关键通路。政府应发挥统筹协调作用,推动教育领域多方主体在治理中的深度参与,形成动态、联动的

风险防控机制。为此,应设立专业的人工智能应用于教育的技术监管部门,具有专业化的监管理论、规则、方法、技术和流程,依法监管,辐射引领国际,为“AI助力教育变革”提供公共服务。定位上,该部门既是保障教育安全、公平的监管者,同时肩负着国际合作与国家治理的重要职能。该部门应牵头构建智能化教育数据监管体系,整合政府、学校和相关企业的力量,推动教育数据风险的协同监管。

第三,试点实验是政府顶层引导下推动AI助力教育变革的重要实践策略。政府应采取循证的“试点—推广”探索模式,在对小范围试点项目进行充分验证与评估后,再逐步向更大范围推广。这种模式通过实证研究与数据支持,为教育改革提供科学依据,避免盲目推广可能带来的教育隐患,确保新技术应用的可行性、重要性与有效性<sup>[27]</sup>。

## (二)规范行业的技术、资质管理与行业自治

正如“科林格里奇困境”所指出的,当一项技术的社会后果在其早期阶段尚未充分显现时,若未采取有效预防措施,一旦不良后果成为既定事实,技术已深嵌于经济与社会结构中,其治理将面临巨大挑战<sup>[28]</sup>。因此,在AI技术尚未给教育行业造成损害之前,对其潜在风险施行严格的防范极为重要。为规范行业管理,可以从全产业链的技术规范、专业资质考试和行业协会三个方面推进系统化建设。

第一,全产业链的技术规范建设是实现AI助力教育变革企业治理的关键举措。《国家人工智能产业综合标准化体系建设指南(2024版)》中将人工智能产业链划分为四个部分,包括基础层、框架层、模型层、应用层,分别涵盖AI技术从底层架构到场景实践的各个阶段<sup>[29]</sup>。从教育产品开发的全流程来看,行业管理规范可以在AI产业链的四个层次中分别进行构建。在基础层,应提升算法透明性与可解释性,并实施严格的数据治理框架,确保用户隐私与数据安全;在框架层,需加强开源框架的安全检测,制定兼容性标准,构建安全高效的技术开发环境;在模型层,应注重消除算法偏见与伦理风险,推动本土化大模型的研发与应用,以实现公平性与适配性的双重保障;在应用层,需建立动态风险评估体系,完善资源配置机制,确保AI技术在教育场景中的稳定性与包容性,助力教育公平与智能化发展。

第二,建立AI助力教育变革专业资质认证和资格考试体系。作为人工智能与教育深度融合的交叉领域,AI教育对从业者提出了更高的专业化要求,他们不仅要具备AI技术开发的核心能力,还应系统掌

握教育学、心理学等领域的基本理论。通过资质认证与资格考试,可以明确从业人员的资质要求,确保其在技术开发与教育应用中具备理论支持与实践能力。此外,要求教育科技企业设立教研部门并将专业资质作为教学研究团队成员的准入条件,也是提升行业治理体系科学性的重要方法,对AI技术在教育场景中的适用性进行科学论证,避免技术滥用和偏差带来的教育风险。

第三,建设AI助力教育变革行业协会是推动行业规范化与AI助力教育变革高质量发展的关键路径。作为连接企业、教育机构和监管部门的重要平台,行业协会在标准化、监督和资源整合方面具有不可替代的作用。首先,协会可以主导制定覆盖算法透明性、数据伦理规范和教育场景适配性的行业标准,为企业研发和技术应用提供明确指引,确保AI技术与教育需求深度融合。其次,协会应负责推进从业人员资质认证和培训体系建设,提升从业者的专业素养和行业整体水平。此外,行业协会可以定期组织研讨AI教育领域的核心议题,形成行业共识并为政策制定提供科学支持。同时,通过强化行业自治,协会能够推动企业自律机制的建立,构建动态监测与反馈体系,及时应对技术应用中的风险与挑战。

## (三)教育主体在价值、素养与创新的引领作用

AI助力教育变革的人机协同学习不仅依赖于系统本身的优化,更需要教育主体的能动性发挥。为了发挥教育的引领作用,教育主体应从价值、素养和方法创新等方面为“AI助力教育变革”提供强有力的主体层面支持。

第一,坚持以教育价值引领技术创新。“以人为本,智能向善”已成为全球技术治理的共同原则,这一原则不仅是对技术伦理的基本要求,更是保障AI技术服务于教育本质的价值基石。正如奥尼尔所言,算法不应被赋予万能属性,任何将其视为解决教育复杂性与多样性问题的终极手段的科技乌托邦幻想,都是对教育本质的偏离<sup>[3]</sup>。优秀的教育学者和一线工作人员应坚守教育价值认识,将AI技术应用于促进人的全面发展,而非陷入技术至上的狂欢。学校等教育机构需通过制度化手段加强价值引导,例如,开展专题培训、组织工作坊或进行教学案例分析,以此提升师生群体对技术伦理价值的认知,从而促进技术与教育形成良性共生关系。

第二,提升AI素养以确保驾驭技术。教育用户(教师、学生、家长)的AI素养不仅是技术风险治理的重要屏障,更是保障技术在教育中有效应用的核心力

量。AI 素养关乎教育用户在技术生态中的适应力、判断力和行动力,其内涵超越了对技术基本功能的掌握,涵盖了对 AI 技术使用的批判性态度、伦理意识和实践能力<sup>[30]</sup>。具体而言,教师应深刻理解技术在教育中的功能与局限,在教学实践中有效调适技术应用,主动识别并反馈技术风险。学生应具备对 AI 技术的基础认知与应用能力,确保在使用中保持批判性思维与伦理意识。家长应具备选择与监督 AI 教育产品的能力,确保未成年人在技术使用过程中得到适当的保护和引导。

第三,为教师主动创新提供必需的支持。在 AI 技术深度融入教育的过程中,积极探索新方法和新路径是化解技术风险、促进教育创新的重要手段。教师应在教学实践中大胆尝试,将 AI 技术融入课堂教学、作业设计和学情分析等多元场景,探索 AI 在教育应用中的问题与局限,通过反馈机制向学校管理层和技术供应商提供改进建议。同时,教师之间可以通过建立专业学习社群,共享探索成果和实践经验,形成促进

技术创新的群体智慧。此外,学校和教育机构应鼓励学生在使用 AI 工具时保持批判性思维,积极参与技术评估与反馈过程。通过定期组织学生用户的意见征集、体验反馈或创新竞赛等活动,可以为技术改进提供多维度的用户视角。

## 五、结束语

人工智能的迅猛发展正以前所未有的速度深刻嵌入教育体系,其为教育改革注入新动能的同时,也挑战着传统教育的价值根基与治理模式。从“工具理性”的技术嵌入,到“生态治理”的系统重构,AI 助力教育变革的未来图景不再是简单的技术部署问题,而是多方主体共同努力,秉持教育理念、主动而为、形成新的协同机制和治理体系。未来的教育不应是被算法定义的教育,而应由教育定义技术发展的教育。推动 AI 助力教育变革的系统生态治理,既是对智能时代教育本质的深度回应,更是迈向公平、高质量教育未来的关键一步。

### [参考文献]

- [1] 黄荣怀. 论科技与教育的系统性融合[J]. 中国远程教育, 2022(7): 4-12, 78.
- [2] 巫娜. “生成式人工智能+教育”的伪主体间性及其风险[J]. 现代大学教育, 2024, 40(4): 27-37.
- [3] 凯西·奥尼尔. 算法霸权: 数学杀伤性武器的威胁[M]. 马青玲, 译. 北京: 中信出版社, 2018: 164, 243.
- [4] 李荣华, 田友谊. 人工智能教育应用的算法隐忧及应对进路[J]. 当代教育科学, 2024(7): 11-18.
- [5] 张羽, 刘惠琴, 石中英. 教育投入产出的人文属性[J]. 教育研究, 2022, 43(8): 121-140.
- [6] 张惠彬, 许蕾. 生成式人工智能在教育领域的伦理风险与治理路径——基于罗素大学集团的实践考察[J]. 现代教育技术, 2024, 34(6): 25-34.
- [7] 贺樑, 应振宇, 王英英, 等. 教育中的 ChatGPT: 教学能力诊断研究[J]. 华东师范大学学报(教育科学版), 2023, 41(7): 162-176.
- [8] 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.
- [9] 于水, 区小兰, 罗珞峻. 基层治理中的数字悬浮: 内涵表征、形成机理及消解策略[J]. 河海大学学报(哲学社会科学版), 2024, 26(4): 79-88.
- [10] 何东平. 基层数字治理的悬浮化及其克服之策[J]. 领导科学, 2023(1): 45-49.
- [11] 张黎, 鲍文雨, 赵磊磊. “智能鸿沟”的教育镜像: 教育数字化转型的底层视角[J]. 现代教育技术, 2024, 34(7): 51-60.
- [12] 王佑镁, 王欣颖, 柳晨晨. 教育领域生成式人工智能应用的伦理风险管理框架研究[J]. 电化教育研究, 2024, 45(10): 28-34, 42.
- [13] 赵磊磊, 张黎, 王靖. 智能时代教育数据伦理风险: 典型表征与治理路径[J]. 中国远程教育, 2022(3): 17-25, 77.
- [14] 郭胜男, 钱雨, 吴慧娜, 等. 面向未成年人的 AI 安全风险: 风险澄思、根源透析与治理进路[J]. 中国远程教育, 2023(7): 39-46.
- [15] KOOPMAN P, WAGNER M. Challenges in autonomous vehicle testing and validation[J]. SAE international journal of transportation safety, 2016, 4(1): 15-24.
- [16] 刘三女牙, 刘盛英杰, 孙建文, 等. 智能教育发展中的若干关键问题[J]. 中国远程教育, 2021(4): 1-7, 76.
- [17] 余乃忠. 积极的“异化”: 人工智能时代的“人的本质力量”[J]. 南京社会科学, 2018(5): 53-57.
- [18] 张欣, 陈新忠. 人工智能时代教育的转向、价值样态及难点[J]. 电化教育研究, 2021, 42(5): 20-25, 69.
- [19] 刘旭东. 人工智能教育立法: 基于教育法典的体系化编纂路径[J]. 现代远程教育研究, 2024, 36(5): 61-71, 82.
- [20] 石中英, 董玉雪, 仇梦真. 从“五育并举”到“五育融合”: 内涵、合理性与实现路径[J]. 中国教育学刊, 2024(2): 65-69.
- [21] 朱爱芝, 熊才平, 贾靖林, 等. 从教育的保守性看信息技术在教育应用中的滞后性[J]. 中国电化教育, 2009(12): 10-13.

- [22] 蔺海洋,付婷. 我国校外教育治理的风险及其规避[J]. 教育学报,2021,17(4):92-103.
- [23] 黎文靖,郑曼妮. 实质性创新还是策略性创新?——宏观产业政策对微观企业创新的影响[J]. 经济研究,2016,51(4):60-73.
- [24] 吴河江,吴砥. 教育领域通用大模型应用伦理风险的表征、成因与治理[J]. 清华大学教育研究,2024,45(2):33-41.
- [25] 柯清超,米桥伟,鲍婷婷. 生成式人工智能在基础教育领域的应用:机遇、风险与对策[J]. 现代教育技术,2024,34(9):5-13.
- [26] 赵磊磊,张黎,代蕊华. 智能时代教育数据风险治理:实然困境与实践路径[J]. 湖南师范大学教育科学学报,2021,20(6):94-102.
- [27] 孙立会,刘思远,李芒. 面向2035的中国教育信息化发展图景——基于《中国教育现代化2035》的描绘[J]. 中国电化教育,2019(8):1-8,43.
- [28] MAYNE A J. The social control of technology[J]. Journal of the operational research society,1982,33(8):771-772.
- [29] 工业和信息化部,中央网络安全和信息化委员会办公室,国家发展和改革委员会,等. 四部门关于印发国家人工智能产业综合标准化体系建设指南(2024版)的通知[EB/OL]. (2024-06-05) [2024-11-08]. [https://www.gov.cn/zhengce/zhengceku/202407/content\\_6960720.htm](https://www.gov.cn/zhengce/zhengceku/202407/content_6960720.htm).
- [30] 张银荣,杨刚,徐佳艳,等. 人工智能素养模型构建及其实施路径[J]. 现代教育技术,2022,32(3):42-50.

## Risk Research, Attribution Analysis and Ecological Governance of Artificial Intelligence Enabled Educational Transformation

ZHANG Yu, YANG Zihao, QIN Fei

(Institute of Education, Tsinghua University, Beijing 100084)

**[Abstract]** Artificial intelligence(AI)-empowered educational transformation has become an important issue in the development of global education, with its technological potential being continuously released and its application scenarios increasingly enriched. However, this process is also accompanied by complex challenges and risks that require systematic ecological governance. The study analyzes three types of risks (subjectivity, safety and developmental risks) of "AI-empowered educational transformation", and focuses on four types of ecological subjects (government, AI technology developers, schools and enterprises) to carry out attribution analysis, revealing a multi-level chain of causes for the three types of risks, involving a quadruple logic: institutional logic, the lack of legal and professional constraints; technological logic, the inherent limitations of algorithms and systems; educational logic, the practical challenges of abiding by the right and the old; and commercial logic, the value distortion of efficiency and profit. In order to cope with the above problems, the study proposes an ecological governance framework of "AI-empowered educational transformation", which aims to provide policy guidelines for the benign symbiosis and sustainable development of AI and education, advocating that the government should take the lead to improve the top-level structure, increase the participation of enterprises to standardize the management of the industry, and give play to the leading role of education to enhance the subject's self-awareness.

**[Keywords]** Artificial Intelligence; Educational Governance; AI Education; Deep Integration; Educational Transformation