

自适应学习系统在K-12教育中的伦理风险互动机制研究

杜丙新

安阳师范学院文化传播学院, 河南 安阳

收稿日期: 2026年7月27日; 录用日期: 2026年9月9日; 发布日期: 2026年9月17日

摘要

自适应学习系统在K-12教育中的广泛应用, 使技术伦理与教育伦理形成了复杂的互动关系。本文以自适应学习系统为研究对象, 首先从数据伦理与算法伦理两个维度阐释其技术伦理的基本框架, 继而从主体自主、师生关系、教育公平与全面发展四个维度揭示K-12教育伦理的核心关切。在此基础上, 本文提出技术伦理与教育伦理之间存在价值侵入与伦理稀释、风险转移与责任模糊、代理关系与角色重构、尺度扩展与效应放大四种互动机制, 并进一步分析从技术伦理问题到教育伦理风险的嵌入、转化、放大三阶段生成路径。研究表明, 自适应学习系统的伦理风险本质上源于离身的技术逻辑与具身的教育实践之间的结构性张力。纾解风险的路径在于推动二者从冲突走向动态平衡, 在坚守K-12教育育人本质的前提下实现技术的向善应用。

关键词

技术伦理, 教育伦理, 互动机制, 风险生成, 教育人工智能, K-12教育

A Study on the Ethical Risk Interaction Mechanism of Adaptive Learning Systems in K-12 Education

Bingxin Du

School of Culture & Communication, Anyang Normal University, Anyang Henan

Received: July 27, 2026; accepted: September 9, 2026; published: September 17, 2026

Abstract

The widespread application of adaptive learning systems in K-12 education has created a complex

interaction between technological ethics and educational ethics. This paper takes adaptive learning systems as its research object, first explaining the basic framework of its technological ethics from the two dimensions of data ethics and algorithm ethics, and then revealing the core concerns of K-12 educational ethics from four dimensions: subject autonomy, teacher-student relationship, educational equity, and holistic development. Based on this, this paper proposes four interactive mechanisms between technological ethics and educational ethics: value intrusion and ethical dilution, risk transfer and responsibility ambiguity, agency relationship and role reconstruction, and scale expansion and effect amplification. Furthermore, it analyzes the three-stage generation path from technological ethical issues to educational ethical risks: embedding, transformation, and amplification. Research shows that the ethical risks of adaptive learning systems essentially stem from the structural tension between the detached technological logic and the embodied educational practice. The path to alleviating these risks lies in promoting a dynamic balance between the two, achieving the benevolent application of technology while upholding the educational essence of K-12 education.

Keywords

Technology Ethics, Educational Ethics, Interaction Mechanisms, Risk Generation, Educational Artificial Intelligence, K-12 Education

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自适应学习系统正迅速重塑 K-12 教育的决策机制，推动其由传统的经验驱动转向数据驱动。智能辅导系统、学习分析工具和自适应学习平台日益普及，为 K-12 学生提供个性化学习体验。然而，这些技术工具并非价值中立。它们往往嵌入并再现了长期存在的社会结构性不平等。数据驱动算法决策带来了个性化学习的黑箱机制与伦理模糊。

在倡导教育包容与公平的时代背景下，教育面临着以个性化、自适应为特征的智能技术所带来的偏见风险。在智能教育产品的研发过程中，潜在的设计偏见、数据偏见和算法偏见使教育面临多重风险[1]。与此同时，人工智能技术对 K-12 学生主体性的削弱亦不容忽视。学生长期依赖智能工具可能产生思维惰性，陷入信息茧房，认知趋于单一[2]。

然而，既有研究多聚焦于技术本身的伦理缺陷或教育实践的价值偏离，未能系统回答一个更为根本的问题：自适应学习系统的技术伦理问题如何转化为 K-12 教育中的教育伦理风险？二者之间通过何种机制发生互动？

本文尝试回应这一问题。首先，阐释自适应学习系统技术伦理的基本维度；其次，揭示 K-12 教育伦理的核心关切；再次，分析二者之间的互动机制与风险生成路径；最后，以典型案例验证理论框架并提出治理建议。

2. 理论基础

为系统分析自适应学习系统在 K-12 教育中的伦理风险互动机制，本文整合了技术社会建构理论、行动者网络理论、风险社会理论与具身认知理论的思想资源，构建了一个以“具身-离身”张力为核心的分析框架。以下依次阐述各理论的核心主张及其与本文框架的关系。

2.1. 技术社会建构理论：技术何以承载价值

技术社会建构理论(Social Construction of Technology, SCOT)由 Bijker 等学者系统提出[3]。该理论主张,技术的发展并非由技术逻辑单向决定,而是不同社会群体对技术进行解释、协商和建构的结果。SCOT 的核心概念包括相关社会群体(relevant social groups)、技术框架(technological frame)和解释弹性(interpretive flexibility)。同一技术在不同群体眼中具有不同的意义和用途,技术的最终形态是社会协商的产物而非技术自身的必然展开。

在教育 AI 领域, SCOT 有助于理解一个关键事实:自适应学习系统的算法偏见、数据滥用等问题,并非技术本身的必然属性,而是技术在被设计、部署和使用过程中被特定价值取向和利益格局所塑造的结果。有研究以 SCOT 为框架考察学术界、产业界与政策群体对 AI 的不同建构,发现商业部门的实践行使着最大的权力,揭示了商业逻辑对教育 AI 技术设计的深刻影响。另有研究以 SCOT 为理论基础分析 AI 聊天机器人在教育中的滥用与伦理风险,强调技术采纳并非单纯的技术问题,而是受社会文化与制度因素塑造的过程。

2.2. 行动者网络理论：非人行动者如何参与教育

行动者网络理论(Actor-Network Theory, ANT)由拉图尔等学者提出[4],主张人类与非人类行动者在网络中处于对称地位,能动性不再专属于人类,而是分布在去中心的、多中心的网络之中。ANT 的核心贡献在于打破了技术是中性工具的传统认知,将技术视为网络中具有能动性的行动者。

在教育场景中,自适应学习系统作为非人行动者进入课堂,与教师、学生、学校管理者、政策制定者等共同构成动态网络。有研究以 ANT 为框架,发现教师在与技术交互的过程中感知 AI 具有类人性、物质性和工具性,识别出人机互动网络中的多元主体。另有研究指出,ANT 视角下 AI 驱动的工具增强了个性化但可能限制学习者自主性,伦理关切涉及算法偏见、去个性化与数据隐私。ANT 还揭示了一个关键发现,伦理责任不再由任何单一主体承担,而是在多主体互动中生成。

2.3. 风险社会理论：系统性风险与责任的消解

德国社会学家贝克的风险社会理论指出[5],现代社会中的风险是人造风险,其分配逻辑不同于财富分配。在风险社会中,风险的制造者(企业、政府、专家)系统性地逃避责任,形成了责任分摊的格局。

在教育 AI 领域,当自适应学习系统的算法偏见造成损害时,技术供应商、学校管理者、一线教师之间形成难以追溯的责任链条。这正是风险社会理论所描述的责任消解的典型表现。人工智能决策的深度介入使教育治理出现判断悬置、责任漂移等风险,其深层机理在于,教育问题被持续转写为可识别、可比较、可计算的技术对象,责任在代理链条中层层切分,为治理增加了难度。

2.4. 具身认知理论：教育伦理的具身性根基

具身认知理论主张认知不是脱离身体的抽象运算,而是根植于身体与环境持续耦合的动态过程[6]。在教育领域,这一理论揭示了教育伦理的独特根基——师生关系的身体共在、教学实践的情境嵌入、教育意义的交互生成,共同构成了教育不可让渡的具身特征。

K-12 阶段学生正处于认知发展的关键期,其主体自主性尚未完全形成,对技术依赖的敏感性更高,因此教育伦理的具身性特征在该阶段尤为突出。当自适应学习系统将学习者的身体经验简化为可计算的数据指标时,教育便从具身的实践退化为离身的计算。

2.5. 本文的理论定位

基于以上四组理论梳理,本文分析的理论出发点为:

整合多理论资源构建了一个贯通的分析框架。SCOT 解释了技术何以携带价值, ANT 解释了非人行动者如何参与教育网络, 风险社会理论解释了系统性风险何以逃避责任, 具身认知理论揭示了教育伦理的特殊性何在。但四者均未能完整解释技术伦理风险如何经由教育场景转化为教育伦理风险、又如何上升为社会伦理风险这一完整的转换过程。本文的嵌入、转化、放大三阶段模型与四种互动机制, 正是对这一完整链条的理论建构。

引入了具身性作为教育 AI 伦理分析的核心维度。既有研究多从数据、算法、应用的技术逻辑出发分析教育 AI 伦理风险, 本文则从具身 - 离身的张力结构出发, 揭示了教育伦理风险区别于一般技术伦理风险的特殊性。K-12 教育的伦理根基在于师生之间的身体共在, 而自适应学习系统的技术逻辑本质上是离身的。这一结构性张力是教育 AI 伦理风险持续生成的深层根源。

3. 自适应学习系统技术伦理的基本维度

自适应学习系统通过算法分析学生的学习行为数据, 动态调整学习内容、路径与难度。其技术伦理风险可以从数据伦理与算法伦理两个维度加以把握[7]。

3.1. 数据伦理维度

3.1.1. 数据隐私与保护

数据是人工智能系统的燃料, 数据伦理关注的是数据在采集、存储、处理与使用全流程中的伦理规范。人工智能技术给教育教学带来智能化、便利性和高效率的同时, 其发展的不确定性和应用的广泛性, 也可能带来技术滥用、数据泄密、侵犯个人隐私等伦理风险问题[8]。自适应学习系统需要持续采集学生的学习行为数据——答题记录、停留时间、点击序列、错误类型乃至情绪状态。K-12 学生属于未成年人, 其认知判断能力尚未完全发展, 隐私保护意识相对薄弱。数据一旦泄露或被滥用, 后果可能是长期的和不可逆的。同时, 数据隐私风险的特殊性在于其隐蔽性与累积性。与显性的物理伤害不同, 数据隐私的侵害往往不易被即时感知, 但其后果可能是长期的和不可逆的。学生的数据一旦泄露, 不仅可能影响其当下的学习生活, 还可能对其未来的升学、就业乃至社会信用产生深远影响。

3.1.2. 数据采集与知情同意

数据采集的伦理问题涉及数据主体是否对数据被采集的范围、方式与用途具有充分的知情权与自主选择权。在 K-12 场景中, 知情同意面临特殊困难。尽管我国法律规定处理未成年人信息需监护人同意, 但多数学校仍采用“全选式”授权协议, 家长和学生难以真正理解数据被采集的范围与用途。学生在使用教育平台时, 往往被要求勾选冗长晦涩的用户协议, 所谓的同意更多是形式上的, 而非实质性的。

此外, 数据采集的最小化原则在实践中常被违反。许多教育 AI 平台倾向于能采尽采, 将数据视为未来的资产加以囤积, 而非服务于明确的教育目的。超出教育目的范围的数据采集——如采集学生的地理位置、社交关系、生物特征等与学习目的无明确关联的信息——构成了对数据主体的不当侵扰。

3.1.3. 数据使用与目的限定

数据使用层面的伦理问题涉及数据是否被用于其采集时声明的目的之外的范围。目的限定原则要求数据采集必须服务于明确、具体、合法的目的, 后续处理不得超出该目的范围[9]。然而, 在教育人工智能实践中, 这一原则常被突破: 学生数据被用于精准营销、被出售给第三方数据中介、在未经同意的情况下被用于训练商业 AI 模型等非教育目的。当数据从改善教育的工具变为可交易的商品时, 学生便从教育主体沦为数据来源, 教育的根本价值被技术逻辑所侵蚀。

3.1.4. 数据生命周期与安全管理

数据安全贯穿数据的全生命周期——采集、传输、存储、处理、共享与销毁。技术漏洞和黑箱特性

可能被不法分子利用，泄露学生个人信息，造成教育机构名誉损害。教育数据的高敏感性使其成为网络攻击的重点目标，但许多教育机构在数据安全方面的投入与能力严重不足。

数据生命周期的另一个关键环节是数据的“遗忘权”——学生是否有权要求删除与其相关的数据，尤其是当这些数据可能对其产生持续性的负面影响时。当前的教育 AI 系统往往缺乏明确的数据过期与删除机制，学生的数据被无限期存储，这种数字记忆可能成为其未来发展的隐患。

3.2. 算法伦理维度

算法的设计逻辑与运行机制直接影响着自适应学习系统的输出，也因此成为技术伦理的核心关切。算法伦理维度关注的正是算法在运算与决策过程中应当遵循的道德规范。具体而言，算法伦理维度可以从以下三个层面加以把握。

3.2.1. 算法公平

算法的客观性常常被当作技术中立的依据，但事实上，算法离不开人的主观判断、知识结构等因素的影响，其背后往往隐藏着数据采集者和算法设计者的价值认同与思想意图。面向 K-12 群体的自适应学习系统的算法公平问题尤为敏感。训练数据若隐含地域、阶层等偏差，输出也必然带有对应偏见。城市样本训练出的算法可能无法准确识别农村学生的学习需求。算法同质推荐以协同过滤为机制，将用户选择归为某一标签后过滤其他内容，可能为 K-12 学生织就信息茧房，影响其认知正常发展。算法偏见可能导致对特定群体学生的系统性误判。更令人担忧的是，算法的偏见往往被客观数据、科学算法的外衣所遮蔽，使其比人类的显性偏见更隐蔽、更难质疑。

3.2.2. 算法透明

算法透明关注的是人工智能系统决策过程的可理解性与可解释性。人们常用“黑箱”隐喻算法的不透明性，即无法从用户视角直接观测到算法的运算逻辑。由于企业或国家的保密、个人技术素养的不足，逐渐使生成式人工智能的任何决策都缺乏解释，最终形成难以捉摸的黑箱形态。

在 K-12 教育中，算法黑箱的危害尤为突出。当教师无法理解智能系统为何给某个学生推荐某一学习路径时，其专业判断权便被架空——教师不再能够基于对学生的具身感知做出教学决策，而是不得不依赖一个无法理解的黑箱输出。当学生无法理解智能系统为何给出某一评分时，其申诉与改进的基础便不复存在——学生无法知道自己“错在哪里”，也就无法有针对性地改进。

算法透明并非追求“完全透明”——这在技术上既不可能也不必要。教育场景对透明度的要求是：算法决策的关键依据与逻辑应能够以不同层次的用户可理解的方式进行解释。对技术人员的解释可以详细到代码层面，对教师的解释应聚焦于决策的关键因素，对学生的解释则应简明直观。不同层次的可解释性，是算法透明在教育场景中的合理实现方式。

3.2.3. 算法问责

算法问责关注的是当人工智能系统造成损害时，责任归属的清晰性与可追溯性。当前，人工智能技术的迭代速度远超制度演进速度，现有治理体系在应对新型风险时存在滞后性。人工智能决策的深度介入使教育治理出现判断悬置、责任漂移等问题。

算法问责的困境源于多重因素的叠加。在技术层面，算法的复杂性与不透明性使因果追溯变得困难。当算法的决策涉及上亿个参数时，即使开发者也无法完整还原某一特定决策的形成过程。在制度层面，缺乏明确的法律法规界定智能系统的责任归属。当算法犯错时，究竟应当归咎于开发者、部署者还是使用者，缺乏清晰的规范框架和指引。在实践层面，使用者、学校与技术供应商之间的责任划分无法清晰界定。有研究深刻指出，教育问题被持续转写为可识别、可比较、可计算的技术对象，责任在代理链条

中层层切分，治理主体则在程序依赖中逐步后撤。

4. K-12 教育伦理的基本维度

教育伦理关注的是教育活动中人与人、人与知识、人与社会之间关系的道德规范与价值准则。与人工智能技术伦理不同，教育伦理的核心关怀不是技术如何运作，而是人如何成长。K-12 阶段学生正处于认知发展的关键期，其主体自主性尚未完全形成。K-12 教育伦理的核心关切可以从四个维度加以把握。

4.1. 主体自主维度

主体自主关注的是教育活动中人的自主性是否得到尊重与保障[10]。主体自主是教育伦理的基石。K-12 阶段的教育根本目的在于培养具有独立人格、自主判断能力和自我发展能力的人。然而，自适应学习系统的深度介入正在对这一基石构成挑战。

技术凝视可能导致学习者因过度依赖技术支持而丧失自主选择能力，因过度数据化而抑制主体性发展。当系统接管学习路径的规划和学习内容的选择时，学生的自主选择就可能被算法引导所替代。学生可能不再关注“我学到了什么”，而是关注“系统认为我学到了什么”。K-12 学生由于认知判断能力尚未完全发展，更容易产生技术依赖。学生长期依赖智能工具，遇到问题直接问 AI，不愿自己动脑。学习自主学习能力可能会产生退化。

4.2. 师生关系维度

K-12 教育从根本上说是一种关系性实践。师生之间的信任、关怀与双向成长构成了教育的伦理空间。然而，自适应学习系统的介入正在改变这一关系结构。

教师从教育主导者滑向技术辅助员。教师不再需要通过与学生面对面的互动来感知学情，只需阅读系统生成的报告即可。教师面临责任边界模糊、价值目标窄化、师生关系疏离及专业认同受限等伦理失序风险。学生层面，智能体正逐渐取代教师原有的倾听和引导职责，甚至会导致师生课外情感互动的淡化。部分学生将智能体作为更加高效和安全的情感交流或互动对象。师生情感互动由“人-人”关系转向“人-技术”关系，教育的温度被技术界面所中介。

4.3. 教育公平维度

教育公平是社会公平的基石，也是教育伦理的核心关切之一。教育公平不仅指向教育机会的均等分配，更指向教育过程中的公正对待——教师理应一视同仁，尊重所有学生的教育权利，保护所有学生的人格尊严，公平地向所有学生敞开教学资源[11]。

自适应学习系统在 K-12 教育中的广泛应用，使教育公平面临新的挑战。算法偏见可能导致对特定群体学生的系统性误判。自适应学习系统可能无法准确识别学习基础薄弱学生的学习需求。数字鸿沟使技术弱势群体被进一步边缘化——城市样本训练出的算法可能对农村学生产生系统性偏差。自适应学习平台在乡村地区的使用被描述为数字分层。初始资源不平等抑制了学习者的感知有用性，导致与城市地区的学习差距持续扩大。付费壁垒可能使经济条件优越的学生获得更优质的 AI 教育资源，形成付费-免费两层教育市场。

4.4. 全面发展维度

全面发展关注的是教育是否促进了人的完整成长。教育的本质并不会因数字化技术的介入而发生改变，教育传承文化和培育人才的本质不会变，立德树人的根本任务不会变。智能技术的工具理性单向度扩张可能消解教育的价值理性根基[12]。

当自适应学习系统以效率和精准为优先目标时，那些无法被量化的教育价值——好奇心、创造力、审美体验、道德敏感——有可能被系统性地边缘化。K-12 阶段是学生认知、情感与价值观发展的关键期，这一阶段的教育不能简化为数据优化。

5. 技术伦理与教育伦理的互动机制

自适应学习系统的技术伦理与 K-12 教育伦理之间并非彼此独立，而是通过四种核心机制持续互动。

5.1. 价值侵入与伦理稀释

技术并非价值中立。自适应学习系统内在地承载着工具理性逻辑——追求效率、标准化、可量化、可预测、可优化。当这种逻辑被植入教育这一以人的全面发展为终极关怀的领域时，技术工具理性便作为一种强势的潜在价值观，对教育的价值理性产生持续的侵蚀和稀释。那种关注过程性成长、非预期性价值、情感交流与人格陶冶的人性教育可能被排斥在技术系统之外。

这一机制的运作可以分为三个步骤。第一步是简化与还原。智能系统为实现其功能，必须将复杂、多维的教育实践转译为可量化、可计算的数据指标。学习被简化为点击率与正确率，教学被简化为知识点的覆盖度，情感被简化为情绪标签。在这一还原过程中，那些无法被量化却被教育视为核心价值的维度，如好奇心、创造力、批判性思维、审美体验、道德敏感等被系统性地遗漏。第二步是优化与替代。当智能系统被赋予优化学习的目标时，它会按照被量化的指标来调整教学策略。原本应由教师基于专业判断做出的教学决策，被算法的优化函数所替代。第三步是合法化与固化。经过一段时间的成功运行，这种被简化的、被量化的教育逻辑逐渐被视为理所当然。教师开始习惯依赖智能提供的学情分析，学生开始习惯接受算法推荐的学习路径。教育伦理的具身根基在这一合法化过程中被悄然置换[13]。

5.2. 风险转移与责任模糊

当自适应学习系统被引入 K-12 学校时，其固有的技术风险并不会停留在技术层面，而是会转移给教育系统、教师和学生。然而，责任归属却在此过程中被系统性地模糊化，形成责任漂移问题。

这一机制包含三个相互关联的环节。第一，风险的技术化包装。技术供应商将智能系统呈现为客观的、科学的解决方案，算法决策被赋予一种技术权威，将教师的专业判断贬抑为主观的、不可靠的。第二，风险的场景化转移。算法偏见在一般场景中可能只是一个统计学问题，但在教育决策中，转化为对具体学生发展机会的实质性剥夺。教师的具身判断原本是抵御这种不公的第一道防线，在算法权威面前被边缘化。第三，责任的链条化断裂。当风险事件爆发时，技术供应商、学校管理者、一线教师之间形成了一条无法追溯的责任链条，各自均可声称自己只是系统中的一部分。教师面临责任边界模糊、价值目标窄化、师生关系疏离及专业认同受限等伦理失序风险。

5.3. 代理关系与角色重构

自适应学习系统从工具演变为代理者。这一角色配置深刻地重构了传统教育中教师 - 学生这一核心伦理关系，导致教育主体的角色异化与功能位移。

在教师端，当智能系统承担了部分知识传递、作业批改、学习分析的功能后，教师从教育的主导者滑向技术系统的辅助员。教师不再需要通过与学生面对面的互动来感知学情，只需阅读系统生成的报告即可。扎根于身体经验的教学直觉被离身计算的算法建议所取代[14]。在学生端，从智能系统的视角来看，学生不是具有无限发展潜力的独特个体，而是一系列可供计算的数据。学生可能开始为数据而学习。为了在系统上获得更好的画像而调整行为，而非为了真正的理解与成长。

5.4. 尺度扩展与效应放大

技术的可扩展性和自动化能力,使原本局部的、微观的伦理失范能够以惊人的速度和规模被系统化、制度化地放大。这一机制是技术伦理问题转化为教育伦理危机的关键环节。

这一机制的运作体现在三个层面。第一,空间维度的扩展。智能产品一旦开发完成,可以以极低的边际成本复制和分发,一个带有设计缺陷的智能产品可以同时影响无数学校与学生。一个有缺陷的算法可以一夜之间影响成千上万的学生。第二,时间维度的深化。一次算法对学生的误判可能被记录在学生档案中,影响后续的教育决策;学生早期被算法贴上带有偏见的标签,可能在其整个学习生涯中被反复强化。第三,结构维度的固化。当技术系统被嵌入教育制度时,其伦理风险便不再是个体层面的偶发事件,而是制度层面的结构性特征。教育人工智能伦理风险治理过程中存在着政府、研发企业、学校、教师及学生等多个利益相关主体,分别基于各自利益作出策略选择,治理困境正是利益相关主体间进行博弈的结果。

6. 从技术伦理到教育伦理的风险生成路径

基于上述机制的分析,自适应学习系统的技术伦理问题并非直接等同于教育伦理风险,而是经历了一个复杂的生成过程。这一过程可以从嵌入、转化、放大三个阶段加以把握。

6.1. 技术逻辑嵌入教育场域

风险生成的第一步是技术逻辑嵌入教育场域。当自适应学习系统被部署到学校时,其内嵌的算法逻辑、数据逻辑与效率逻辑便一同进入了教育实践。在这一阶段,技术伦理问题(如算法偏见、数据隐私风险)尚处于潜伏状态,尚未与具体教育情境发生深度耦合[15]。

然而,技术逻辑已经开始对教育的运行方式产生潜移默化的影响。教学决策的依据从教师和管理者的专业判断逐渐转向算法的数据输出,学习路径的规划从师生的共同协商逐渐转向系统的自动推荐。算法的不透明性侵蚀了教育评价的公正基础,削弱了教师的专业判断权。在这一阶段,技术伦理问题尚未显现为显性的教育伦理风险,但其潜伏本身即构成了风险的初始条件。

6.2. 技术问题转化为教育问题

嵌入阶段之后,技术伦理问题开始与 K-12 教育实践的具体情境相结合,转化为教育伦理问题。这一转化通过以下途径实现:

价值冲突型转化。当技术的效率逻辑与教育的育人逻辑发生冲突时,技术伦理问题便转化为教育价值偏离问题。技术效率与教育规律之间存在结构性冲突。例如,算法推荐为了最大化参与度而不断推送学生熟悉的内容,在技术层面是优化,在教育层面却是对学生接触异质知识、发展批判性思维的剥夺。

关系重构型转化。当智能系统承担了部分教学功能时,师生之间的互动模式发生变化,技术伦理问题便转化为师生关系异化问题。教师从引导者变为监督员,学生从学习者变为数据源。这种角色重构本身就是教育伦理风险的直接体现。

权力让渡型转化。当教育决策权从教师让渡给算法时,技术伦理问题便转化为教育主体性危机。算法逻辑与育人目标之间存在内在冲突。教师不再拥有如何教的自主选择,而是必须按照系统设定的参数和标准来组织教学。

6.3. 局部风险演化为系统危机

在转化阶段之后,局部的教育伦理问题通过技术的扩展效应被系统性地放大。资源与技术的可及性

差异加剧了教育资源配置的结构失衡，进而危及教育公平。一个有偏见的算法在单一学校的影响是局部的，但当同一算法被部署到成百上千所学校时，其影响便被指数级放大[16]。

更为关键的是，这种放大往往不被个体感知。没有一个教师能够亲眼看到数百万学生同时被算法影响的图景，因此，针对这种大规模风险的伦理感知与问责机制便难以启动。教育人工智能伦理风险治理需要加快前置式教育人工智能伦理规则研制，预防潜在伦理风险；提升博弈主体对教育人工智能伦理风险的认知，促进伦理价值认同。

7. 结论

本文从技术伦理与教育伦理的维度界定出发，系统分析了二者之间的互动关系与风险生成路径。研究表明，技术伦理风险与教育伦理风险之间并非简单的技术应用引发伦理问题的线性关系，而是通过价值侵入与伦理稀释、风险转移与责任模糊、代理关系与角色重构、尺度扩展与效应放大的机制发生深度互动。

从技术伦理问题到教育伦理风险的生成，经历了嵌入、转化、放大三阶段演进过程。技术逻辑嵌入教育场域后，经由价值冲突、关系重构与权力让渡等途径转化为教育伦理问题，再通过技术的扩展效应被系统放大，最终形成教育伦理风险。

教育人工智能伦理风险的治理不能仅仅依靠技术修补或伦理规范的单一路径，而需要在技术伦理与教育伦理的互动中寻求动态平衡。既要承认技术逻辑的正当性，也要坚守教育伦理的底线，使技术服务于教育的本质而非替代或扭曲它。对教育人工智能的应用风险规制，应以技术进步推动制度改革，以制度改革促使共识构建，以观念共识反哺技术发展，形成互补的反馈循环。唯有如此，才能在智能时代实现技术向善与教育本真的协同共生。

上述分析揭示了教育人工智能伦理风险治理的根本方向：治理不是对风险的消除，而是对风险生成过程的管理。在技术逻辑与教育逻辑的持续互动中推动系统从冲突向平衡。然而，原则的落地需要转化为不同主体可理解、可执行的具体行动。未来，教育行政部门、学校管理者、一线教师与 AI 开发者可尝试从以下建议出发，完善和落地伦理风险的治理。

7.1. 对教育行政部门的建议

建立 K-12 阶段自适应学习系统的准入评估制度。对涉及学生认知发展关键期的高风险应用实施强制伦理影响评估，将对师生具身互动空间的压缩程度作为核心评估指标。

制定算法审计的实践流程。要求进入 K-12 学校的自适应学习系统定期提交算法审计报告，审计内容应包括训练数据的代表性、推荐内容的多样性、不同群体学生接收内容的差异分析。

完善教育数据治理的政策工具。明确“高具身敏感数据”（涉及学生身体体验、情绪状态、行为姿态的数据）的保护标准。

7.2. 对学校管理者的建议

确立教师最终决定权机制。在涉及学生评价、分班等关键决策节点，系统建议仅为参考，教师基于具身感知的判断具有最终决定权。

为师生具身互动保留制度空间。在课时安排中保障师生面对面互动的的时间，防止自适应学习系统过度压缩师生身体共在的空间。

建立校级 AI 伦理委员会。由一线教师、学生代表、家长代表共同参与，对学校引入的自适应学习系统进行日常监督。

7.3. 对一线教师的建议

培养批判性使用 AI 的专业能力。教师培训应从如何使用系统转向如何批判性地评估系统输出，避免将专业自信从亲身感知转移到数据分析。

在 AI 辅助教学中保持具身判断。将系统生成的学情报告作为参考而非判决，坚持基于日常课堂观察与师生互动所获得的具身感知。

7.4. 对 AI 开发者的建议

在产品设计中嵌入具身兼容约束。算法的优化函数不应仅包含效率指标，还应包含对学生认知多样性的保护约束。

主动标识算法建议的不确定性。避免以确定性输出的形式遮蔽可能的错误，为教师的具身判断保留介入空间。

建立人类在环境的强制性通道。在关键教学决策节点(如学情评估)，确保教师能够介入、审核并否决算法建议。

在更宏阔的视野中，自适应学习系统在 K-12 教育中的伦理风险治理，关乎的不只是技术如何更好地服务于教育。它更关乎在技术加速迭代的时代，K-12 教育如何保持其作为人与人相遇的本质，特别是如何守护师生之间面对面相遇的具身空间。当算法的智能日益强大时，我们需要清醒地认识到，真正的教育，从来不是数据的优化，而是人的唤醒；而人的唤醒，需要身体的在场。

基金项目

研究为教育部人文社会科学研究一般项目“教育人工智能伦理风险感知的影响因素及其作用机制研究”(编号: 23YJA880007)阶段性成果之一。

参考文献

- [1] 曾天山, 陆宇正. 人工智能在教育应用中的伦理边界与规范[EB/OL]. 2025-04-28. https://cssn.cn/skgz/bwyc/202504/t20250428_5871389.shtml, 2026-02-05.
- [2] 冯雨兔, 谢心怡. 超越符号表征: 教育人工智能价值对齐的困境审视与范式转向[J]. 西南民族大学学报(人文社会科学版), 2025, 46(9): 232-240.
- [3] Bijker, W.E., Hughes, T.P. and Pinch, T.J. (1987) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. MIT Press.
- [4] 拉图尔. 科学在行动: 怎样在社会中跟随科学家和工程师[M]. 刘文旋, 郑开, 译. 北京: 东方出版社, 2005.
- [5] 贝克. 风险社会[M]. 何博闻, 译. 南京: 译林出版社, 2004.
- [6] Varela, F.J., Rosch, E. and Thompson, E. (1991) *The Embodied Mind*. The MIT Press. <https://doi.org/10.7551/mitpress/6730.001.0001>
- [7] 刘骥, 丘霖. 生成式人工智能嵌入教育应用的风险生成及其规制[J]. 现代远距离教育, 2024(4): 12-19.
- [8] 陈翠荣, 崔红岩. 博弈论视角下教育人工智能伦理风险治理困境及化解策略研究[J]. 中国地质大学学报(社会科学版), 2025, 25(3): 131-144.
- [9] Wieczorek, M. and Tollon, F. (2026) Teachers' Busywork, Moral Entanglement, and the Automation of Responsibility. *Ethical Theory and Moral Practice*, 29, 565-582. <https://doi.org/10.1007/s10677-026-10557-9>
- [10] 胡小勇, 黄婕, 林梓柔, 黄漫婷. 教育人工智能伦理: 内涵框架、认知现状与风险规避[J]. 现代远程教育研究, 2022, 34(2): 21-28.
- [11] 沈苑, 汪琮等. 人工智能在教育中应用的伦理考量——从教育视角解读欧盟《可信赖的人工智能伦理准则》[J]. 北京大学教育评论, 2019, 17(4): 18-34+184.
- [12] 刘培, 池忠军. 算法歧视的伦理反思[J]. 自然辩证法通讯, 2019, 41(10): 16-23.

-
- [13] 白钧溢, 于伟. 教育人工智能伦理原则的有限性及其补正——兼论“原则-美德-法律”框架[J]. 中国电化教育, 2024(2): 23-31.
- [14] 郭庆, 吴砥. 国际视野下人工智能教育应用伦理风险与治理策略[J]. 比较教育研究, 2025, 47(1): 56-64.
- [15] 杨俊锋, 沈中奇, 陈睿宁. 生成式人工智能的教育应用及伦理风险探析[J]. 湖州师范学院学报, 2023, 45(12): 1-12.
- [16] 谢娟. 智能教育“人在回路”中教师角色转型的伦理审视[J]. 西南民族大学学报(人文社会科学版), 2026, 47(6): 224-231.