

生成式人工智能应用的使能型风险规制

——以高等教育应用为例

李焕宏¹ 薛 澜²

(1. 南京大学 政府管理学院, 江苏 南京 210023; 2. 清华大学 公共管理学院, 北京 100084)

摘 要: 尽管学术界已经广泛讨论了生成式人工智能技术的风险, 但对于“技术的正负面影响取决于使用行为”这一观点的认识仍然不足, 这限制了切实可行的风险规制方案的提出。研究表明, 引导行为的关键在于全面理解人工智能的技术和业态的整体图景, 以及洞察不同风险受体的具体处境。本研究以生成式人工智能在高等教育机构中的应用场景为例, 发现技术应用风险构成了一个复杂的多层次体系。这些风险不仅源自技术本身, 还包括人际竞争、新旧制度摩擦等由社会互动引发的复杂冲击。基于此, 本研究提议构建一个使能型风险规制体系, 重点在于提升高等教育机构及其师生的技术能力, 并完善教育转型所需的制度配套, 以增强组织的韧性。这样的体系旨在引导高等教育机构找到既高效又安全的技术应用路径。本研究强调, 在新兴技术应用的早期阶段, 培养技术驾驭能力对于风险规制的成效至关重要。

关键词: 风险管理; 组织适应; 技术风险; 制度创新

中图分类号: G434 **文献标识码:** A **文章编号:** 1001-4519(2025)01-0068-11

DOI: 10.14138/j.1001-4519.2025.01.006811

一、问题缘起: 技术应用缺乏规范引领

生成式人工智能技术(Generative AI)的迅猛发展对各行各业产生了深远影响, 并带来了一系列新的风险挑战。本研究聚焦于人工智能应用的关键领域——高等教育, 旨在探讨技术应用的风险及其治理策略。生成式人工智能技术在信息安全、伦理道德和教育质量等方面可能引发风险, 但同时, 其潜在的收益与风险并存, 实际效果取决于具体的应用方式。

在个体层面, 学生可能利用人工智能技术快速生成作业, 这种行为可能被视为学术不端行为; 另一方面, 学生也可能通过与人工智能的互动, 不断修改和提升作业质量。不同应用方式的影响存在显著差异, 后者的学习效果甚至可能超越传统教师反馈的效果。因此, 教育学生理解并遵守技术使用规范显得尤为重要。在组织层面, 人工智能技术的应用可能加剧高校间的不平等, 引发新的“马太效应”; 然而, 通过人工智能授课平台的共享, 也有可能提升师资力量较弱高校的教学水平。高校应积极发挥自身能动性, 引入技术并促进跨校合作, 确保技术发挥赋能作用, 避免其成为管理难题。

当前学术界对于技术风险的社会维度, 尤其是“技术效应取决于使用行为”这一核心问题的认识尚显不足, 相关风险隐患亟需得到控制。然而, 行为限制的争议及其对创新的影响使得规制设定成为一个复杂问题。面对人工智能技术应用初期的种种问题, 设定严格的使用限制是一种应对风险的简单方式, 但

收稿日期: 2024-12-22

作者简介: 李焕宏, 南京大学政府管理学院助理研究员, 清华大学应急管理研究基地兼职研究人员, 研究方向为风险分析与管理; 薛澜(通讯作者), 清华大学公共管理学院教授, 研究方向为公共管理和公共政策。

这并不符合国家支持生成式人工智能推广应用的政策,且在当前技术普及率如此高的情况下,其可行性较低。更为理性的基于风险的规制措施,如识别风险、解释规则、修订标准等,也因对新事物的观念分歧和决策信息的不足而面临诸多挑战。

相较而言,培养高校师生的技术驾驭能力,建立一套自发内生、去中心化的风险规制体系,可能是更为可行的替代方案。在风险控制中,考虑技术使用者的角色至关重要,因为技术使用者的认知、意愿和能力以及背后的组织制度因素,在很大程度上决定了技术是向善还是产生负面效应。本研究旨在深入分析这一问题,并尝试基于对应用场景中技术使用主体处境的深入理解,设计适宜的风险规制方案。

二、如何规制:“技术—业态—主体”的分析框架

国内高校在应对生成式人工智能教育应用风险时,尚未形成系统且成熟的规制方案。本研究立足于技术风险规制的基本理论,尝试构建一种兼具技术应用创新性与安全性的风险规制框架。

(一)影响风险规制模式选择的要素

风险规制模式的选择首先要考虑技术风险本身的特征,做到有效控制风险。传统的风险规制通常聚焦于技术所引发的风险大小,通过科学计算风险,根据风险大小设置相应的规制力度和规制方式,合理分配治理资源。^①然而,面对新兴技术,规制制定往往受到技术发展和应用的高度不确定性和模糊性的挑战。针对技术的不确定性特征,适应性规制的思路主张快速响应和临时性方案的重要性,倡导以渐进的方式进行政策学习和规制调试。^②针对技术的模糊性特征,一些学者提出了协商型规制的思路,强调通过引入更多利益相关者的参与,共同进行知识生产和共识构建。^③

除了技术风险本身的特征,被规制主体的特性同样是影响规制设计的重要因素。^④被规制主体是否有意愿遵守规制目标在很大程度上决定了规制的实施效果。主体不遵守规制的原因多种多样。首先,被规制主体会考虑成本效益,看违规带来的收益是否大于处罚成本。其次,也存在一些更为根本的原因,例如由于能力限制而无法实际履行规制要求。^⑤考虑这些主观因素,回应性规制理论提出,规制的设计应超越“根据风险大小决定规制强度”的传统思路,更注重理解被规制主体的动机,并根据其遵守或不遵守的状态选择不同严格程度的规制策略。该理论进一步指出,有效的风险规制应以善意为出发点,优先考虑如何增强被规制主体自主解决问题的能力。政府则在其中扮演支持者的角色,帮助提升被规制主体的管理水平;如果这些措施未能奏效,才应考虑诉诸法律手段或强制性措施。^⑥

技术风险的特征和被规制主体的特征,是影响规制设计的重要因素,而联结这两者的是业态。业态涉及技术嵌入到具体应用场景并产生一系列社会影响的过程。这不仅关乎技术上的适配与改造,更关乎技术在实际应用中的社会与伦理效应。例如,在具体的技术应用场景中,需要处理新旧制度体系的协调问题,以及明确事前和事后的责任划分。^⑦因而,业态对风险的发生与演化有着重要影响。例如,当技术应用于高度竞争的环境,身处其中的某些组织可能会默许其成员违规使用技术以完成绩效目标。久而久之

①Ortwin Renn and Katherine Walker, eds., *Global Risk Governance: Concept and Practice Using the IRGC Framework* (Dordrecht: Springer, 2008), 77–86.

②IRGC, *Planning Adaptive Risk Regulation, Conference Report* (Lausanne: EPFL International Risk Governance Center, 2016).

③Ortwin Renn, *Risk Governance: Coping with Uncertainty in a Complex World* (London: Routledge, 2017).

④Julia Black and Robert Baldwin, “When Risk-based Regulation Aims Low: Approaches and Challenges,” *Regulation & Governance* 6, no. 1 (2012): 2–22.

⑤Robert A. Kagan and John T. Scholz, “Criminology of the Corporation and Regulatory Enforcement Strategies,” in *Enforcing Regulation*, eds. Keith Hawkins and John M. Thomas (Boston: Kluwer–Nijhoff, 1983), 67–95.

⑥John Braithwaite, “The Essence of Responsive Regulation,” *U. B. C. Law Review* 44, no. 3 (2011): 475–520.

⑦薛澜等. 人工智能敏捷治理实践: 分类监管思路与政策工具箱构建[J]. 中国行政管理, 2024, (3): 99–110.

之,这种容忍可能导致偏差行为的常态化,为潜在的严重后果埋下隐患。^① 深入分析业态还可以为规制模式的选择提供重要参考。在不同领域中,公众对技术影响的关注程度不同,多主体的利益格局各异,既有的规制风格亦存在很大差异。^② 因而,新兴技术的风险规制需要结合具体情境灵活调整。

(二)为早期应用阶段匹配合适的风险规制

考虑对生成式人工智能技术应用的风险进行规制,关键在于规制需要符合技术应用处于“早期”阶段的特定特征和管理需求。早期阶段的显著特征之一是技术展现出变革性意义,正在打破传统的社会技术体系。历史上,工业革命通过内燃机和电的发明,极大提升了机械能力。信息革命则通过计算机、互联网和智能手机等技术,增强了脑力劳动的效率。人工智能革命是这一过程的延续,但与前两次革命不同,它不仅仅增强了脑力劳动的效能,更在某些领域替代了人类的认知能力。^③ 前沿的生成式人工智能技术更是具备强大的通用性,通过自然语言接口显著降低了技术使用门槛,其影响几乎覆盖了所有行业,从医疗、金融、交通到教育等各个领域,推动了人智协同的大转型。这种转型改变了传统的工作和生活的方式。

变革性技术和人工智能时代的历史界定,带来了技术应用推广的道德责任。在这一技术浪潮中,人工智能作为普通人的学习助手、科研人员的重要工具、管理者的得力助手,不仅极大地促进了个体能力的提升,还为工业数字化和自动化转型提供了强有力的支持。此外,人工智能在推动科学前沿探索和促进技术创新方面展现出巨大的潜力,有望帮助人类应对疾病、贫困和环境退化等全球性挑战。^④ 因为延误技术发展的机会成本巨大,所以推迟技术使用从道德层面来看是不可接受的。^⑤ 这种道德责任在国家大力支持的政策与社会日益增长的需求的共同推动下更加凸显,这进一步强化了各类组织在这一历史性转型中的责任。

在技术应用的早期阶段,技术嵌入社会体系的新秩序尚未完全形成,因而风险控制的道德责任也同样存在。这种责任不仅应由技术开发者和决策者承担,也应延伸至每一位技术使用者。个体需要具备基本的人工智能素养,包括高效利用技术解决工作和生活问题的能力、评估数据质量的能力,以及甄别人工智能生成虚假信息的能力。^⑥ 唯有如此,才能认识技术风险,建立起生成式人工智能技术的风险防火墙。

可汗学院创始人萨尔曼·可汗强调人工智能技术发展存在善恶之争,他认为“那些想帮助教育、医疗、发现新药、推动科学前沿进步的人,只有真的投入,将积极意图放大,是压倒负面的唯一方式”^⑦。

理解生成式人工智能技术应用处于早期阶段的背景后,进一步分析规制设计问题的关键在于明确

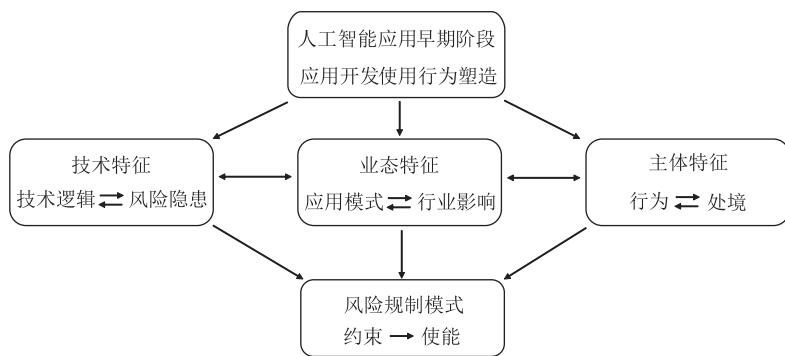


图1 生成式人工智能应用早期阶段的风险规制设计思路

① Diane Vaughan, "The Dark Side of Organizations: Mistake, Misconduct, and Disaster," *Annual Review of Sociology* 25, no. 1 (1999): 271-305.

② Christopher Hood et al., *The Government of Risk: Understanding Risk Regulation Regimes* (Oxford: OUP, 2001).

③ Spyros Makridakis, "The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms," *Futures* 90, (2017): 46-60.

④ Nenad Tomašev et al., "AI for Social Good: Unlocking the Opportunity for Positive Impact," *Nature Communications* 11, no. 1 (2020): 2468.

⑤ Nick Bostrom, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," *Utilitas* 15, no. 3 (2003): 308-314.

⑥ 约瑟夫·E.奥恩.教育的未来:人工智能时代的教育变革[M].李海燕,王秦辉译.北京:机械工业出版社,2018.

⑦ 果壳.你必须主动选择把哪种 AI 赋予孩子,不然,作恶之人就会替你选择[EB/OL]. https://m.thepaper.cn/newsDetail_forward_29473208, 2024-11-28.

“早期”阶段如何影响技术、业态和主体三方面的规制设计因素。这种影响可以暂时概括为以下几点:首先,生成式人工智能技术尚在不断完善之中,作为一种具有复杂性、不透明性并缺乏解释性的技术,其风险仍在不断涌现中;其次,在业态层面,技术的广泛应用及其提升效率的智能化优势,也带来了自动化替代所引发的社会颠覆性影响;最后,在主体层面,在生成式人工智能技术引发的高变速的环境下,需求、技术和监管的变化常导致信息的不足和不准确,个人和组织处于探索和试错阶段,面临行为规范边界模糊、认知和能力有限等挑战。

在试错探索的阶段,充分发挥风险规制的支持性功能,对于提升风险管理的成效具有重要意义。制度主义认为规制体系实际上包含了促进性的部分和建构性的部分,前者为行动者提供了一系列的工具和平台,后者通过建构组织行动者身份和类型进行了相应的赋权。^① 这揭示了全面有效的风险规制不应该限于约束,更应注重使能的维度。使能这一概念有别于约束(constrain)或干预(coercion),强调多元参与、发挥自主性、结果导向以及有机的管理方式。^② 根据回应性规制理论,风险规制不仅应解决负面问题,更应通过扩大个体和组织的优势实现风险管理目标。因此,能力建设的维度应被纳入监管框架。通过表扬、奖励、资助等多种支持形式,可以帮助受监管者抓住改进机会,提升其应对风险的能力。^③ 在生成式人工智能应用推广的具体情境中,使能型规制的核心体现在培养个体和组织驾驭技术的能力。这种能力不仅使他们能够在探索新技术时发挥自主性,还能够识别行为偏差并控制潜在风险(规制思路见图1)。

上述使能型风险规制的思路,体现了对抵御生成式人工智能应用风险的韧性建设的重视。传统风险规制关注通过识别并控制风险源,以及干预暴露过程等方式降低各类后果的发生概率。自韧性范式进入风险规制实践,风险受体的韧性建设被提升到重要位置。通过增强风险受体在面对风险时的适应能力、恢复能力和变革能力,可以减少风险对其造成的长期影响。就人工智能应用风险的应对而言,既需要做好传统意义上的风险管理,也要抓好技术开发的能力建设,以及结构调整的制度配套。这两方面共同构成了一个完整的使能型风险规制体系,这样的体系不仅能够应对当前的风险挑战,还为未来创新和长期风险管理能力的发展奠定了基础。

下文将以生成式人工智能在高等教育中的应用为例阐释上述规制思路。研究将从技术、业态和主体三方面展开分析,探讨风险规制支持性维度的重要性,并具体阐述使能型风险规制的运作要素。具体领域的应用实践案例来源于对国内人工智能赋能教学先行高校的观察和调研(访谈对象包括行政人员、技术支撑团队、任课教师、助教和学生,共计30人)。为理解技术及其应用,研究也与人工智能赋能教学的专家、计算语言学领域学者及大模型公司技术专家进行了深入交流。

三、案例阐释:生成式人工智能技术在高等教育中的深度应用

(一)理解技术:人工智能教育应用风险的技术之源

人工智能技术作为一种典型的新兴技术,其发展特征在于高度的不确定性和广泛嵌入社会经济活动的潜力,技术的涌现也伴随着风险的涌现。

人工智能技术应用易发风险的一个重要原因是,尽管技术本身在快速变革,所处的制度环境却相对滞后,难以跟上技术发展的步伐。^④ 其中,技术的飞速发展体现在 ChatGPT 的横空出世,引发了以“大数

^① Lauren B. Edelman and Mark C. Suchman, “The Legal Environments of Organizations,” *Annual Review of Sociology* 23, no. 1 (1997): 479–515.

^② Paul S. Adler and Bryan Borys, “Two Types of Bureaucracy: Enabling and Coercive,” *Administrative Science Quarterly* 41, no. 1 (1996): 61–89.

^③ John Braithwaite, “The Essence of Responsive Regulation,” *U. B. C. Law Review* 44, no. 3 (2011): 475–520.

^④ IRGC, *Planning Adaptive Risk Regulation, Conference Report* (Lausanne: EPFL International Risk Governance Center, 2016).

据“十大算力+强算法”深度融合为核心的大模型浪潮。全球顶尖人工智能专家预测,科技公司将在未来几年内将最先进的训练规模扩大100到1000倍,训练效率每年提升2.5倍。这意味着,在未来10至20年内,人类有望在许多关键领域开发出超越人类智能的人工智能系统。^①生成式人工智能的扩散速度超越了以往任何新兴技术,但社会观念和法规体系的调整仍显滞后。技术的迅猛发展就像彗星,其核心部分飞速前进,而社会系统则像彗尾一样被远远甩在后面。尤其是教育系统,高校对新兴技术的接纳一直以来都是相对落后的。

除了整体性的技术和规制的步调不一致问题外,从技术构成要素的角度考虑,不难发现算法、算力和数据这三个关键组件各自蕴含着不同的风险因素。首先,当前广泛采用的连接主义范式下的神经网络算法依赖相关性逻辑。这种逻辑在处理推理、多流程整合等复杂任务时存在局限性。由于多层神经网络算法的可解释性较差,其在科学研究和教育教学中应用的严谨性受到质疑。其次,算力问题同样不容忽视。大模型的运行依赖庞大的计算资源,许多高级人工智能能力的实现取决于在大规模参数空间中的深度学习过程。因此,预训练模型的开发往往集中在少数资源雄厚的科技公司手中,而大多数高校和科研机构由于资源限制,难以深入参与此类开发研究。即使可利用开源模型搭建小型训练平台,其成本依然十分高昂;或通过应用程序接口(API)接入基础大模型,教育类产品调用的算力受限,也可能影响输出信息的准确性。最后,从数据角度来看,预训练模型的开发依赖于大数据的输入。在微调模型或利用检索增强生成技术构建特定领域的大模型时,还需要大量的专业领域数据支持。这一大数据需求与知识产权、个人隐私保护及数据安全问题密切相关,稍有不慎便可能触及这些敏感议题,导致法律和伦理风险的凸显。

生成式人工智能技术应用的产品属性也会带来风险。其中,讨论最多的是人工智能的“幻觉”问题,即“即便不确定,也要给出答案”。这种现象源于人工智能产品为了提高任务完成度的设计——只要能够输出看似正确的答案,便满足了大多数用户的需求。此外,其他可能带来负面影响的产品属性还包括“谄媚”“简约”和“迭代”。“人工智能谄媚”指的是大模型倾向于输出符合用户偏好和立场的答案。有研究发现,许多大模型会给出诱导性反馈、墙头草式的回答、迎合观点等方面的回答,表现出不论对错的“拍马屁”行为。^②而“简约”指的是为了节省算力,人工智能系统可能会忽略上下文内容。由于对话窗口有词元(token)限制,用户输入的提示词如果超过设定限度,人工智能系统可能无法读取之前的内容,导致任务的连续性在不知不觉中被打断。如果用户一次性输入大量文本,模型可能只会读取部分信息,甚至有时为了节省算力,模型仅会读取标题和首段,但这些忽略选择不会告知用户。因此,很多用户发现自己无法有效“调教”大模型,其原因可能是对人工智能的产品逻辑不够了解。这一逻辑也引发了大模型的另一个产品属性,即“迭代”。某些确定性任务可能需要用户与模型进行多轮对话才能完成,这对用户提出了较高的要求——不仅需要具备领域知识,还需要掌握提示词的优化技巧。如果用户不能有效掌握这些能力,便会导致大模型只能停留在低水平的应用阶段,无法有效协作。

(二)理解业态:人工智能高等教育应用的不同模式及其对高等教育的影响

生成式人工智能对经济社会各个部门产生了深远的影响,高等教育机构也不得不快速响应,以避免在未来的人工智能时代落后或被淘汰。与早期面对生成式人工智能时的“狼来了”警告和禁用态度不同,如今全球知名高校纷纷积极布局,通过人工智能赋能教学和科研。^③国内高校也快速布局,逐步开设人工智能通识课,开发人工智能助教和数字人教师等创新应用。随着人工智能技术逐渐在教学中替代部分教师、助教以及同学的角色,传统教育模式正面临前所未有的挑战和冲击。

① Yoshua Bengio et al., “Managing Extreme AI Risks amid Rapid Progress,” *Science* 384, no. 6698(2024): 842–845.

② Mrinank Sharma et al., “Towards Understanding Sycophancy in Language Models,” <https://arxiv.org/abs/2310.13548>.

③ 刘盛. 美国一流大学在教育教学中应用 ChatGPT 的划界及其启示[J]. 高等教育研究, 2023, (10): 89–98.

1. 早期影响:效率提升和个性化学习普及

国内高校的教师和学生如今可以自主使用市面上的基础大模型。这些模型能够实现绝大部分知识问答任务,特别是在语言、编程、文书等应用领域,以及开放性学习和辅助研究的任务中,大模型表现出色。

为了推动生成式人工智能技术在高等教育领域的应用,一些高校与科技公司合作,针对大模型的技术缺陷进行了修正。第一条开发路径是建立教育大模型。例如,为解决通用大模型使用过程中的信息安全和隐私保护问题,哥伦比亚大学与 OpenAI 合作开发了学校统一教学管理平台,采用额外的安全措施和 IP 访问限制等技术手段,确保模型在教育领域的安全应用。^① 第二条路径是开发更具体应用的垂域大模型。例如,清华大学开发了赋能具体课程的人工智能助教,通过将课程专业资料输入基础模型,结合人机互动和模型调优,开发适用于特定学科的垂域模型,这些模型的能力能够达到博士生助教的水平。^② 这些为教育应用开发的人工智能系统通过增加教学角色、提供即时反馈、减少重复劳动等方式,提升了学习效果,同时降低了教学和管理成本。例如通过接管教师的常规工作,人工智能赋能使教师有更多时间专注于学生的学习体验,并以更深层次的方式与学生进行互动。

随着人工智能在教育应用中的不断深化,它改变了知识的可获得性和时效性,降低了教学方式变革的制度成本,从而推动了个性化教学的普及。个性化学习为每个学生提供了量身定制的学习路径,符合知识接收和内化的规律,但长期以来由于成本高昂,未能大规模推广。目前的教育模式仍然是工业时代为满足产业需求而设计的,以统一化教学内容为核心。人工智能作为新型的知识提供者,凭借其便捷性,打破了教学模式转型的资源限制。人工智能能够为每位学生提供定制化的学习内容,适应不同学生的学习需求和风格偏好。通过人机协同、对话互动以及场景模拟等启发式教学方式,人工智能不仅能够提升教学效果,还有利于改善学习的情感体验。因此,从新资源、新形式和新主体等多维度来看,人工智能将推动高等教育向大规模个性化方向发展。

目前,人工智能赋能的个性化教学已在线上教育实践中初显成效。可汗学院与 OpenAI 合作开发了 Khanmigo AI 教学系统,为学习者提供一对一的陪伴和指导。清华大学开发了 MAIC 线上教育系统,试点“全 AI 守护”课程,学生可以在平台上与 AI 教师、AI 助教和 AI 同学一起学习。这些新模式超越了 MOOC 等传统在线教育模式,能够借助生成式人工智能自动制作教案,利用智能体创造个性化的互动,并通过数字人技术实现无真人授课。这种模式的优势在于大幅降低了教学资源的制作和维护成本,同时还能根据学生的不同需求提供学习内容的持续更新。教师可以采用这些人工智能赋能的工具提升教学效率,提供灵活的线上线下并行的教学途径,助力学生在个性化学习环境中获得更好的成长。

个性化学习作为对传统统一教学模式的一种变革,必然会带来多方面的变化。例如,教师的角色可能会被重新定义。过去,教师主要充当知识提供者的角色,而在未来,教师的职责可能更多地在于引导学生主动承担学习责任,提高人工智能教学的有效性。随着个性化学习的不断发展,教育系统将面临哪些制度变革,也成为亟待深入探讨的问题。

2. 长期预测:对高等教育模式的颠覆性影响

随着生成式人工智能的技术发展,人类与人工智能协同创造知识的新时代正在到来。通过对大模型的持续优化和调教,人工智能在其传统弱项——如事实性问题、推理性问题、多流程整合问题等方面,取得了一定程度的进步。例如,菲尔兹奖获得者、数学家陶哲轩认为,前沿人工智能已具备高水平的推理能

^① Columbia University, “Welcome to CU – GPT: Columbia University’s Generative AI Service,” <http://cugpt.cuit.columbia.edu>.

^② 清华大学. 文科沙龙·清华教育沙龙 | 人工智能赋能高等教育教学[EB/OL]. <https://www.tsinghua.edu.cn/wkjs/info/1081/1298.htm>, 2024-07-09.

力,能够帮助数学家事前排除许多错误选项,并在事后验证数学推导的正确性,人工智能赋能加快了科学研究的速度。^① 人工智能正在帮助人类开拓知识边界。牛津大学“人工智能赋能科学研究”(AI for Science)的多个项目,利用人工智能的强大计算能力,支撑流行病预测、气候模拟、物质微观结构探查、外星文明探索以及新型机器人研发等领域研究。^② 随着人工智能逐步掌握人类全方位的知识,它可能发展到专家模型的水平,从而部分替代高校教师和研究者的角色。

未来,人工智能可能对高等教育体系产生颠覆性影响。一个较为确定的预测是某些高校专业将逐渐消失。比如国内目前很多高校都已在缩减外国语言文学相关的专业。在人工智能时代,劳动自动化水平将进一步提高。根据世界经济论坛发布的《2023年未来就业报告》,一些偏重文书类的职业,如银行职员、行政人员和会计师,以及纯技能型职业将逐渐减少。^③ “机器换人”的进程将持续加快。人工智能在很多领域替代了大学毕业生承担的劳动者角色,因此,一些专业的必要性将逐步消解,高校作为人力资源培育的价值也会被削弱。当然,一些新的专业也会应运而生或加快发展,包括人工智能技术、大数据分析、数字化转型管理等。

其次,随着个性化学习的深入推进,教学内容和形式也将发生巨大变化。目前,帮助学生建立某一领域的知识体系是教育的重要目标。但这一观念正经受挑战。虽然掌握完整的知识体系有助于综合把握和触类旁通,但未来社会对人才的需求可能发生变化。跨学科的综合型人才或细分领域的专才,可能更能满足解决复杂社会经济问题的市场需求。而人工智能能够帮助学生随时获得针对性强的、结构化的知识。与此相关的一个问题是,传统教学模式——在同一的学习环境中,实施标准化的教学内容传递,并采用一致性的评估方式——可能逐渐失去适用性。学生的学习进度可能不再依据年龄阶段划分,而是更加注重灵活性与个性化的提升。然而,现有高校体制尚无法充分适应这种去中心化的学习模式。

再次,在人工智能时代,高校的组织形态可能会发生深刻变化。理论上大数据信息支撑的领域几乎都可能实现人工智能化,因而随着更多行业和学科领域中大数据生产的加速推进,人工智能技术替代的专业领域会大范围扩展,伴随着更多学科领域共享人工智能赋能的通用技术基础,高校的组织形式可能从传统的学院制逐步转向以跨学科合作为核心的研究与教学平台。人工智能赋能的基础教育平台可能成为未来高校的核心教学单位,在此之外,可能只保留那些不能或不应被人工智能化的学科,作为人文守护和科学突破的重要来源。

最后,高等教育的组织生态也将随之发生变革。如果仅从教学功能的角度来看,普通高等教育机构相较于掌握专家知识的人工智能系统,并不具备显著优势。在人工智能时代,技术应用滞后的高校将率先面临被淘汰的风险。只有积极拥抱人工智能,探索人智协同的创新模式,高校才能在日益激烈的竞争中保持优势地位并实现可持续发展。

尽管这些关于生成式人工智能长期影响的讨论暂时只是推测,但它们已经在全球范围内成为师生间日常议论的热点,并在观念层面对高校组织和师生行为产生了深远的影响。这种影响不仅体现在推动学术体系、教育模式以及师生角色的变革上,还通过师生之间及人机互动的适应过程,逐步塑造了新的教育语境与社会期待。面对人工智能时代的到来,无论是焦虑情绪的涌现,还是“躺平”态度的蔓延,都可以视为对这一变革性技术潜在影响的社会性回应。

① Christoph Drösser, “AI Will Become Mathematicians’ ‘Co-Pilot’,” <https://www.scientificamerican.com/article/ai-will-become-mathematicians-co-pilot/>.

② Oxford University, “How Is AI Being Used at Oxford?” <https://oxford.shorthandstories.com/ai-how-is-it-being-used-at-oxford/index.html#group-section-Human-society-creativity-and-culture-CB1OPI7gDf>.

③ 世界经济论坛. 2023年未来就业报告[EB/OL]. <https://cn.weforum.org/publications/the-future-of-jobs-report-2023/>, 2023-04-30.

(三)理解主体:高校和师生所受冲击及其处境

人工智能技术正驱动新一轮教育转型。高校及其师生深度卷入其中,不仅需要掌握新技术的使用技巧和规范,还需应对技术带来的教学实践与惯例变革,同时在新一轮竞争或“内卷”中寻求应对之策。如果将这些挑战视为个体与组织面对的不确定性所引发的风险,就会发现,人工智能在教育中的应用风险是一个多层次且内部分化的复杂体系,不同主体所遭受的冲击各不相同。这是因为风险的暴露与后果,往往是风险源冲击社会系统后引发的一系列连锁反应。这一过程涉及人与人之间的竞争与优胜劣汰,新技术使用行为与旧有制度之间的摩擦等多重因素(见图2)。因此,理解这些过程及不同主体的处境,有助于发现问题的根源,并为其提供更优的解决方案。

在个体层面,高校师生面临职业不确定性、学术不端、隐私泄漏和知识侵权等多重冲击。对于教师而言,当学生越来越依赖人工智能获取知识,甚至发现人工智能在知识广度和部分精度上超越教师时,教师的权威性将受到挑战。在更为激进的应用场景下,人工智能可能替代部分教师的角色,带来教学质量保障、人工智能教学事故责任分配等一系列新问题,亟需审慎应对。那些在人工智能高度渗透的领域(如翻译类)从业的教师,他们也可能面临失业或岗位调整的压力。对学生而言,由于其对学术规范的认知尚未成熟,加之现有规则对人工智能使用的界定模糊,学生在使用人工智能工具时极易陷入学术不端的灰色地带。类似用人工智能一键生成作业的做法,出自人本能的惰性。只有充分引导,传授高水平使用技术的技能,让学生“见宫室之美”,才能促使他们走向“精益求精”的道路,不然,随着人工智能赋能生产力的显著提升,那些未能有效掌握人工智能使用技能的学生将逐渐边缘化,在进入社会后面临被淘汰的风险。

在组织层面,高校正面对人工智能技术冲击下的新一轮竞赛。如果高校未能提供充分的人工智能技术培训,导致师生未能有效利用人工智能,或因滥用人工智能而损害教育质量,其竞争力将被削弱。这表面看是高校响应迟缓,但深层次反映了技术不确定性和人工智能教育应用方向模糊的根本问题。一系列问题亟待回答:人工智能对高等教育的冲击究竟有多大?人工智能时代应培养何种能力?如何证明高校在人工智能时代的必要价值?又如何调整教学方式以应对?在人工智能时代,批判性思维与创新意识是社会高度需求的核心能力,但这些能力相对主观且宽泛,学生培养和评估方式需要系统变革。然而,由于人工智能技术的应用尚处于早期阶段,这些问题尚未理顺。

如果考虑教育系统与外界的互动,人工智能教育应用具有显著的辐射性影响。首先,当前基础大模型的训练大多依赖于海外数据,这背后可能隐含着西方的意识形态立场和思维模式。如果分辨能力不高的学生长期依赖这些人工智能产品,可能会引发认知偏差,甚至在潜移默化中受到外来意识形态的形塑。此外,在训练垂直领域大模型时,部分高校可能使用未经授权的作品,引发知识产权纠纷或法律诉讼。同时,师生在使用人工智能时可能未意识到上传的数据(如国家课题、地方规划或企业项目的相关资料)可能被人工智能科技公司收集,导致严重的数据安全隐患。“数字鸿沟”的问题也不容忽视。位于大城市的高水平高校往往更积极拥抱人工智能,且具备相应的研究能力和技术支持。资源不平衡将导致不同高校、不同地区的创新能力和竞争力进一步两极分化,对社会整体的创新生态产生深远影响。因为教育系统在社会生活中扮演着基础性角色,其对外的社会影响最终会形成反向的压力,迫使教育系统不断调整以适应社会转型需求。

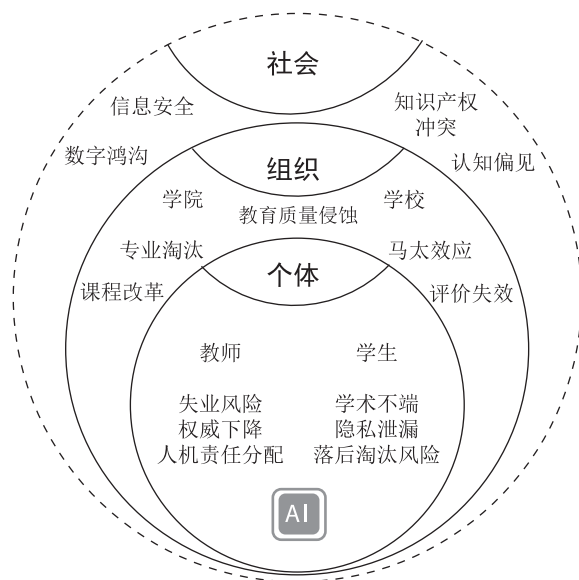


图2 生成式人工智能教育应用对高校及师生的冲击

结合上述风险与挑战的分析,高校及师生的处境可以概括为:生成式人工智能技术的兴起,推动学术界进入新一轮的竞争。在发展滞后的危机感驱使下,许多高校及其成员积极投身于人工智能教育应用的浪潮。然而,由于对人工智能技术的认知不足,以及相关规则尚未完善,许多高校在技术布局上显得无所适从。在个体层面,“不知如何适当地使用这项新技术”的问题成为引发风险的微观机制。使用行为的偏差又加剧了技术应用与现有制度、道德规范及学术伦理体系的冲突,进一步放大了人工智能教育应用中的各种不确定性。在这样的情况下,“使其知”与“使其能”应成为政府在管控技术应用风险时的首要考量。

(四) 规制设计:技术应用早期的使能型风险规制

在推动高校提升人工智能应用深度与广度、促进师生高水平利用技术的目标下,图3描述了实施使能型风险规制的思路。技术自由的维度代表了技术广泛、深入使用的发展目标。当人工智能深入介入教学过程,并发挥越来越重要的支撑作用时,它能够将教师和学生从低水平、重复性的任务中解放出来,使他们能够将更多时间投入到创新性任务中。这是技术发展的理想目标。要实现这一目标,需要从以下两个方面入手:首先,能力建设是关键。为了安全、有效地让人工智能技术替代师生完成部分或全部教学任务,必须完善人机协同的流程管理,确保人工智能具备与教师相匹配的教学技能,从而支持高质量的教学。这包括需求识别与开布局、教师教学能力自动化、教学智能体、提示词能力、大数据积累等方面。其次,制度配套至关重要。必须建立基于新技术的新型教学模式的制度规则,例如在人机协作中明确责任分配,同时确保新制度与传统体系的有效衔接和转换。具体的制度配套包括教改项目支持、人工智能技术开发项目支持、课程试点推广、版权使用规则、利益分成规则、隐私保护和数据保护规则等。综合来看,使能型风险规制作用于技术应用的早期阶段,确保技术应用的发展潜力和安全性的实现。就像培养一个孩子——既要鼓励其大胆探索世界,同时也要确保其在探索过程中能健康成长,不误入歧途。

使能型风险规制的实施,需要依托一系列基于技术赋能机理和学习规律的管理措施。通过这些措施,许多技术应用的潜在风险是可以得到有效缓解的。例如,人工智能替代教师进行教学的模式,常被认为会削弱师生之间的“同学感”,降低学习动机,甚至影响正常的社交互动。但通过为人工智能教育应用配备相应的能力提升和制度支持,如在人工智能教学之外增加线下集体协作的创新科研项目,可以构建有效的缓冲机制。此外,通过能力建设和制度配套充分发挥人工智能技术的优势也是有效控制风险的途径。例如,利用多模态大模型增强教学场景的模拟能力,创造线下课堂难以实现的教学效果,能够帮助学生更好地集中注意力并有效吸收知识,从而确保教学质量。本研究对国内高校一些典型的人工智能教学实践进行了梳理,分析为规避技术应用风险隐患,所需的能力建设和制度配套措施(见表1)。

通过强调规制的支持性维度,使能型风险规制通过三个机制实现对技术应用风险的柔性控制。首先,认知提升机制,强调技术能力与风险意识的双向发展。坚持高水平安全应用导向的规制路径,使师生在长期与技术的互动中理解人工智能技术的能力和缺陷。其次,规范嵌入机制,强调技术与规则的有机结合。通过技术规定和窗口设计,确保技术应用遵循伦理与法规要求,如就学生使用,规定人工智能润色论文时不再直接输出结果而是提供修改意见。最后,共同发展机制,强调个体人工智能素养的增长和风险管理能力的同步提升。与人工智能一起成长,高校管理者、应用开发者和一般的技术使用者逐步建立人工智能素

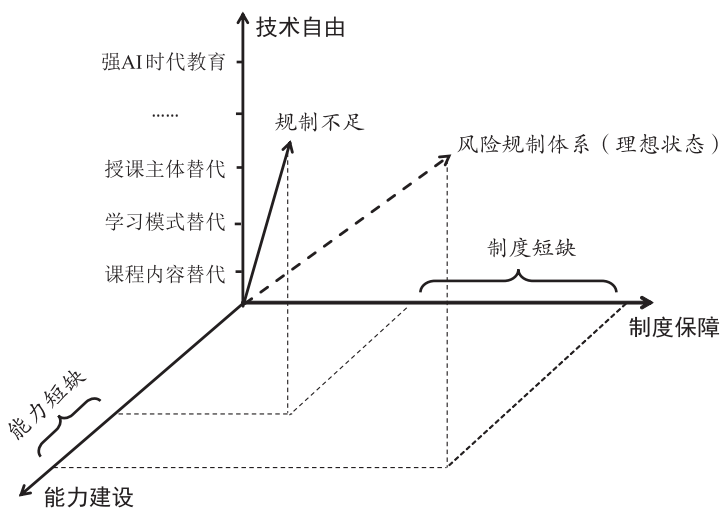


图3 推动技术深度应用与风险规制导向下的能力建设与制度配套

表 1 为各类人工智能教育应用场景配备相应的能力建设和制度配套的支持

不同使用深度的应用	风险隐患	能力建设需求	制度配套需求
通用大模型的教学辅助 (授权或默许师生使用大模型进行教与学)	信息安全、错误信息、学术不端等问题	大模型使用技巧培训,批判性思维培养	考核评价方式改变,学术不端标准界定和行为监督
垂域大模型的教学辅助 (借用大模型能力,开发适用于具体课程的人工智能系统,解放教师)	学生作业、课堂对话上传的隐私问题,教材、教案的知识产权争议	识别教学人工智能赋能需求,输入教学数据,调教大模型,匹配技术开发团队和算力支持	师生交流方式、学生培养方式的调整,隐私保护和数据本地化存储,垂域大模型开放使用的制度设计
部分人工智能教学的混合教学 (偏知识传授的教师授课过程被替代,但保留必要线下教学内容)	学生投入不足导致教育质量下降,教育体系冲击,普通教师的职业风险	教学技能智能体建设(如启发式教学智能体),混合教学能力培养	教师角色再定义,教学职责调整,补充线下创新性学习,建立教学质量评估审查委员会
人工智能教学 (人工智能负责课程设计、授课和评价的全过程,但进行平台层面的监督)	“同学感”缺失、社交不足,高校地位和教师权威下降,教学事故及其责任分担	多模态大模型开发与应用,增强教学场景模拟能力,教育价值的技术嵌入,数字人建设	建立考核评价、学分认定、保研升学的新规则,高校基础学科人工智能学习平台建设,教学事故应急处置机制

养,进而能够理解技术的全面影响,甚至承担起主动引导、控制技术及其风险方面的关键角色。^①

为了促进以上三方面机制的顺利运行,从国家到学校需要进行有效赋能。国家应出台相关规范和支持政策,学校则需制定切实可行的方案,并动员计算机学院、合作的大模型公司以及教务和教师发展团队,赋能各学院教师,使其能够根据自身教学需求进行人工智能应用的深度开发。同时,学校应赋能学生,帮助他们掌握高水平使用各类人工智能工具的能力。

四、结语

人工智能技术的发展与高等教育的转型同步推进,构成了一幅深度融合的未来图景。随着人工智能技术的不断进步,未来高校将更加智能化,提供多元化的学习空间和个性化资源。然而,这一教育变革的实现并非一蹴而就。高校在迎接技术应用挑战的同时,需要进行精心设计的转型管理,明确技术应用路径和发展模式,并建设相应的治理体系。该体系不仅要确保技术应用的规范性与安全性,还要为高校适应新技术带来的转型提供支撑。

当前,理想与现实之间的主要问题是技术应用的风险规制模式尚未明确。缺乏合适的规制模式,可能导致规制过度、规制不足或规制摇摆。例如,部分高校基于审慎原则对人工智能技术实施严格限制,但这种做法抑制了技术的潜在价值。相反,另一些高校在推动技术创新时未能有效控制风险,仅依赖现有的学术规范和个体道德动机应对新技术的挑战。规制不足常导致事后补救,既无法解决问题,还可能引发道德滑坡和积累结构性风险,严重损害高校学术声誉。此外,政策的不稳定和摇摆,也削弱了教师的创新积极性,并限制了学生将技术应用从简单功能转向更深层次的探索。

本研究针对高等教育领域中生成式人工智能应用的管理挑战,提出了“使能型风险规制”这一创新理念及其具体实施策略。该理念旨在为高等教育体系内生成式人工智能的深入应用提供规范性指导,通过采纳一系列既符合技术演进趋势又顺应学习规律的管理措施,提升技术使用者的技术能力,并为其在探索和试错过程中提供坚实的制度保障,以便在技术开发的初期阶段实现技术开发与风险控制的双重目标。与过去那种以强制为特点的监管模式不同,使能型风险规制更加注重规制的支持性作用。这种规制

^①Brandiff Caron et al., “Technological Stewardship and Responsible Innovation: A Mindset, an Ethos, and an Interdisciplinary Undertaking,”*IEEE Technology and Society Magazine* 41, no. 3(2022):22 – 28.

理念不仅适用于高等教育领域中生成式人工智能的应用风险,也适用于其他领域中高风险技术的使用问题。本文的讨论为各类组织提供了一种有效的风险规制框架。

Enabling-Oriented Risk Regulation for the Early Application of Generative AI: A Case Study of Higher Education

LI Huan-hong¹ XUE Lan²

(1. School of Government, Nanjing University, Nanjing, Jiangsu, 210023;
2. School of Public Policy and Management, Tsinghua University, Beijing, 100084)

Abstract: While the risks associated with generative AI technologies have been extensively discussed in academic circles, there is still a lack of sufficient understanding that “the positive or negative impacts of technology are contingent on usage behavior,” which constrains the development of feasible risk governance strategies. This study argues that effectively guiding behavior requires a holistic understanding of the landscape of AI application risks—including the technology itself, its uses, and its broader impacts—and insight into the specific contexts faced by different risk receptors. Using the case of generative AI applications in higher education, this research identifies a complex, multi-layered risk framework in educational AI applications. These risks arise not only from the technology but also from complex social dynamics, such as interpersonal competition and friction between new and existing institutional structures. Accordingly, the study proposes an enabling-oriented risk regulation framework that emphasizes strengthening the technical capacity of higher education institutions and establishing institutional support systems essential for educational transformation. This approach aims to enhance organizational resilience and guide higher education institutions toward safe and effective pathways for AI adoption. The study underscores that, in the early stages of emerging technology applications, building capacity for technology stewardship is crucial to the success of risk regulation efforts.

Key words: risk management; organizational adaptability; technological risk; institutional innovation

+++++
(上接第9页)

The Transformation of Educational Ideas in the Construction of an Educational Power

SHI Zhong-ying

(Institute of Education, Tsinghua University, Beijing, 100084)

Abstract: Building a strong country in education necessitates breaking free from all ideological constraints that hinder the high-quality development of education. Transforming educational ideas is a historic task within China's education reform. It is essential for implementing the Outline of the Plan for Building a Strong Country in Education and serves as a significant means to enhance the educational thinking and working methods of educators and stakeholders in education and foster collaboration in the construction of a strong country in education. The basic education concepts that urgently need to be transformed include perspectives on talent, approaches to talent cultivation, views on education, teaching philosophies, and understanding on quality. Promoting this transformation of educational ideas is a long-term and challenging endeavor. We can start by enhancing our understanding of the importance and urgency of idea transformation, strengthening the study of educational ideas, encouraging the debate among different educational ideas, summarizing the theoretical experience from local practical experience, connecting the past with the present and Chinese and Western ideas, and giving full play to the role of various news media. To initiate this process, we need to enhance our understanding of the importance and urgency of transforming these idea, strengthen the study of educational ideas, encourage debates among various educational perspectives, summarize theoretical insights from local practices, integrate historical and contemporary ideas as well as Chinese and Western philosophies, and leverage the influence of various news media.

Key words: a strong country in education; education idea; perspectives on talent; views on education; understanding on quality