

敏捷治理:教育人工智能伦理治理新模式

王佑镁, 王 旦, 梁炜怡, 柳晨晨

(温州大学 大数据与智慧教育研究中心, 浙江 温州 325035)

[摘 要] 进入 21 世纪以来,人类迈入数据规模呈指数级上升的人工智能时代,人工智能教育成了热切的必然趋势,也出现了不可避免的风险挑战,引发一系列伦理问题,而教育人工智能伦理治理研究始终停留在单一化、被动化的层面,治理效能不显。文章将教育人工智能的伦理风险归纳为“PBAS”四类问题,剖析伦理风险治理现存的五个短板,提出借鉴使用一种全新的治理模式——敏捷治理,来创新应对教育人工智能伦理治理问题。研究以敏捷治理的演绎趋势与内涵特征为切入点,结合敏捷治理的原则与能力,构建教育人工智能伦理敏捷治理五步循环模型,设计五层实践路径,并以“PBAS”四类问题为例,详细阐述敏捷治理解决方案。以期为教育人工智能伦理问题提供一种新的治理范式,为伦理风险的可持续治理与健康发展提供参与与借鉴。

[关键词] 教育人工智能; 伦理问题; 敏捷治理; 治理模式

[中图分类号] G434 **[文献标志码]** A

[作者简介] 王佑镁 (1974—), 男, 江西吉安人。教授, 博士, 主要从事人工智能教育、数字阅读研究。E-mail: wangyoumei@126.com。

一、引 言

随着算法与交互方式的不断更新迭代,人工智能不断推动人类社会向智能化方向迈进,对包括教育在内的社会各个领域带来了巨大的冲击。2017年7月,中华人民共和国国务院发布《新一代人工智能发展规划》聚焦人工智能技术的变革与未来发展,并指出智能时代教育改革的基本方向与根本性道路,即发展智能化教育^[1]。目前,教育人工智能发展如火如荼,在促进学生个性化学习、提升教师教学效率、实现数据精准采集等方面显示出巨大潜力和显著成效,但同时也给教育某些方面带来新的挑战,诸如师生间情感交流较少、技术存在偏见与歧视、教学过程机械化等,在教育领域逐渐被技术垄断的过程中,人工智能教育应用忽略了伦理和道德的内在意蕴,忽视了以人为本的准则,从而背离了教育的本质与初衷。社会各界开始重视教育人工智能伦理规范建设,一方面,从人工智能伦理原则中建构教育伦理规范及治理路

径^[2];另一方面,探讨人工智能与学校教育的有效融合方式以此避免伦理风险的发生^[3]。从趋势看,教育人工智能会走向强大的融合发展场域,如果仅仅依靠传统自上而下、被动的治理方式,是否能够及时应对教育人工智能引发的一系列伦理问题,是否需要改变治理方式;如果需要,又将如何开展和治理,本文试图回答上述议题。

二、教育人工智能伦理风险问题及治理现状

(一)教育人工智能主要伦理问题

人工智能是能执行通常需要人类智力或感官能力的任务或功能的技术,在科学、工程和技术等领域占据主导地位,也通过人工智能学习系统、算法产品等在教育领域蓬勃发展。教育人工智能是人工智能与学习科学相结合形成的新研究领域,通过观察、理解和掌握学习的发生过程,为学生的学习创造更好的条件,帮助教师更有效地教学^[4]。例如,人工智能可为学生提供个性化的学习系统,为教师提供自动评估系

统、辅助教学的面部识别系统等。然而,人工智能也为教育领域带来了一些伦理道德的缺陷,如算法偏见和歧视对学生造成心理伤害,放大种族、性别、地域等多种形式的 unfair 性,侵犯学生隐私和数据带来隐患等。胡小勇等将伦理问题归纳为技术设计与开发和教育实践两方面^[5],苗逢春总结了基于数据和算法的预测和决策引发的四类伦理问题^[6],邓国明等将教育人工智能伦理问题划分为利益相关者伦理问题、技术伦理问题和社会伦理问题三个层次^[7],将教育人工智能伦理问题的研究推上热潮。本文借鉴 Akgun 在 K-12 教育中提出的人工智能教育应用的伦理问题^[8],主要从“PBAS”四类来阐述教育人工智能的伦理问题,包含隐私、偏见和歧视、自治、监视,如图 1 所示。

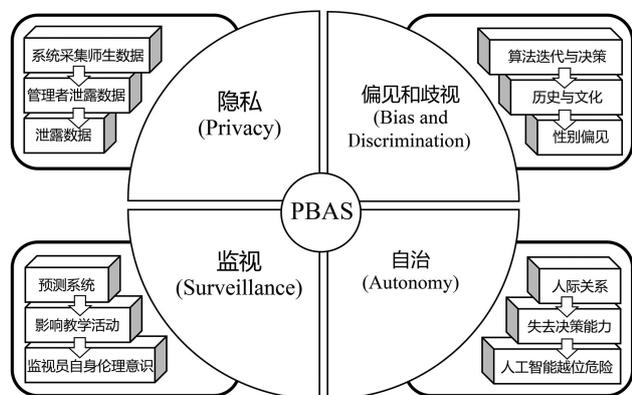


图 1 “PBAS”四类问题

1. 隐私(Privacy)

教育人工智能伦理最大的问题之一是隐私问题,隐私侵犯主要发生在学生的个人信息未被征得同意的情况下暴露过多,通过不知情分享降低学生个人能动性,与泄露隐私,隐私保护与开放共享之间的关系矛盾难以化解^[9]。例如,在人工智能系统采集学生学习行为数据、面部表情数据等情况下,学生个人隐私极有可能被传播,产生不良影响。管理者在收集数据后,也可能发生向教师、家长等随意透露的风险。数据的过度采集十分容易加剧学生隐私泄露而带来的伤害^[10]。在学校单方面使用人工智能产品来辅助教学的同时,学生和家对于隐私问题是未知和别无选择的。此外,一些在线教育平台也会在未经用户的允许下收集超出范围以外的信息并泄露给第三方。

2. 偏见和歧视(Bias and Discrimination)

教育人工智能的变革关键是算法,但算法是人类开发设计的,能操纵且不易改变。因此,经过长时间的算法迭代,它的未知性无疑给教育领域带来了一定的算法偏见或歧视,可能造成信息茧房。研发者在没有充足教育知识与经验的情况下,无法预先洞悉算法会

带来哪些问题,从而对学生造成偏见与歧视,这很有可能是无意的。例如,2020年,英国某中学利用资格等级的算法对大量学生造成了严重地分层打击,引发教育的不公平性^[11]。此外,性别偏见和歧视也是常见伦理问题,对语言学习课程学生最为明显。例如,使用人工智能翻译器时,谷歌翻译会将“她/他是护士”翻译为女性,将“她/他是医生”翻译为男性,这表明了教育人工智能的社会偏见和对性别的刻板印象。

3. 自治(Autonomy)

教育的本质是主体间发生交互的学习过程,教师与学生在享受人工智能带来便利的同时,也存在着人际关系的伦理风险。例如,教师与学生间主体地位不明,学生丧失自主学习能动性,失去自主决策的权利,教师过度以人工智能教学为主,弱化创新思维和探究性学习培养。此外,更有可能出现人工智能越位现象,人工智能不仅具有著作权、劳动权等权利与义务,甚至能精准模仿人类行为与语言。可是,人工智能在教育领域带给教师与学生学习、教学、评价和监测等方面的便利都是有益的吗?相关问题的探讨目前没有得到明晰地解决。

4. 监视(Surveillance)

在监视方面,最熟悉的教育人工智能应用则是利用智能算法设计对师生的情绪、表情、行为等进行收集与记录。将监控机制嵌入人工智能预测系统中,以衡量学生学习表现、优缺点、学习方式,甚至对师生之间的交流对话进行录音分析。这必将限制师生之间的正常交流,使其逐渐对自己想法产生质疑^[12],从而影响正常的教学活动,降低教学效率。此外,教师不恰当行为的发生,如经常性巡逻与监督,也会给学生带来未知伤害,但很有可能被人工智能产品视为对学生的正确管理,这一伦理问题的解决方式主要需要背后管理人员规范。

(二)教育人工智能伦理问题的治理状况

针对教育人工智能伦理现存的一系列问题,引起了各国广泛关注,各国开始重视对教育人工智能伦理问题的治理。目前,仅有少量的官方政策文本与报告和学者们的研究。例如,联合国教科文组织在2019年发布《北京共识——人工智能与教育》^[13]《教育中的人工智能:可持续发展的机遇和挑战》^[14],强调了人工智能需要为全民提供优质的教育和学习机会,并对其带来的影响进行预见性评测。杜静等提出“APETHICS”框架,为教育人工智能伦理问题提供解决模型^[15]。本文重点关注白金汉大学建立的教育人工智能伦理研究所在2020年发布的《中期报告:值得信赖的人工智

能伦理指南》^[6]，其中提到为实现道德化的教育人工智能而提供的机制选项，包括九条基础实现机制和七条深层实现机制。其中，基础实现机制中第五条：利益相关者的教育和意识，即提供教育相关知识提升利益相关者的道德意识。深层实现机制中第三条：保障途径，即在法律保障的范围内确保学习者使用教育人工智能的安全性。然而，现阶段对教育人工智能伦理问题治理仍存在现实难点，传统的治理方式已不能满足其风险的治理，主要问题归纳为五个方面。

1. 缺乏规范指引，系统性不足

面对教育人工智能存在的各类伦理问题，仍没有系统的教育伦理标准与管理规范。在企业不断拓宽人工智能在教育领域的应用范围下，政府政策与法律规范的更新速度已跟不上行业的迭代产出^[7]。主要原因包括：一是目前教育人工智能产品与现有规范不匹配；二是未及时出台规制教育人工智能产品带来的伦理风险制度，教育人工智能的发展长期处于一种野蛮生长的状态。例如，对于教师使用人工智能辅助教学时，产生的有关隐私、偏见等伦理问题对学生造成威胁，而学校管理者过分相信人工智能产品，这是缺乏系统地顶层规范指引而造成的。

2. 人文关怀缺漏，主动性不强

智慧教育的目标始终是以人为本，需关注人、尊重人、彰显人的主体性。在教育人工智能的伦理问题中关于师生人际关系问题值得关注与思考，教师人文关怀能给学生带来深远影响。例如，使用人工智能过程中学生对自身角色转换的感知很微妙，教师用数据来度量学习效果会造成学生学习主动性降低。现阶段教师对学生的人文关怀远远不足，不能及时且主动地预估给学生造成的伤害^[8]。此外，教育人工智能协助教学的过程中，对教师职业的稳定性存在潜移默化的影响，尽管研究者认为人工智能无法代替人类教师，但在简单重复的工作、智能批改、智能数据采集存储等方面人工智能的优势十分明显。随着时代的快速发展，给教师带来的压力是无法预估的，目前各方对教师职业的主动关怀更是甚微。

3. 治理工具缺新，智能化不够

近年来，在解决教育人工智能伦理问题时，众多国家相继发布针对人工智能伦理问题解决的政策报告，并提出教育领域人工智能的合理性边界与正确的做法，我国也同样予以高度地关注与重视。然而，有学者提出，教育人工智能伦理问题治理总体还停留在对其抽象价值的提取与共识构建的阶段中^[9]。传统治理方式并未采用智能化的治理工具对其进行及时且精

准治理，缺乏深度与创新度。例如，在教育人工智能使用中，如果没有及时记录并累积每次出现的问题，并提醒后来的使用者警惕这些风险，那么教育人工智能伦理的问题必将形成一个恶性循环，而恰巧这方面正需要智能化治理工具的协助。

4. 问责机制欠缺，反应力不快

人工智能在教育领域的顺利推行，离不开开发人员、管理人员、教师等互相协调，但在追究主体责任方面，至今欠缺良好的问责机制，治理制度的规制性和规范性不够^[20]。开发工程师是否按照可持续性、透明性等原则构建教育人工智能的系统架构，并保障人工智能的运行程序受到追踪问责，确保使用过程中公平公正对待每个学生，是值得存疑的^[21]。例如，数据的误差、操作不当、算法偏见与歧视，都可能带来相反结果。在事故发生后，受害者的权益如何得到正当的维护，追究事故背后原因，以避免下次再发生类似的情况，是迫切需要解决的问题。

5. 受众意识缺失，弥合性不高

想构筑教育领域良好的人机协同关系，需具备基础的人工智能素养知识。然而，无论是开发人员还是教师与学生，对其受众意识具有缺失性。例如，老教师现有的智能教育素养水平不足以充分发挥人工智能作用，走向高质量的数字化教育转型。将人工智能素养教育推广至学校、培训机构也没有系统地规范和精准地投入，弥合性较低。此外，教育人工智能实际上拉大了各学校与地区间的数字鸿沟，其发展没有真实辅助不发达地区、一些特殊群体的学习，群众普遍没有意识其带来的差异与影响。因此，仍需更加关注教育人工智能落后的山区学生、聋哑学生等^[22]。

三、敏捷治理：一种全新的治理导向

所谓敏捷，是面对事件或环境变化时能“简单、灵活、快速”地作出反应。2018年，世界经济论坛的召开，“敏捷治理”被赋予了新的概念和意义，强调其是一种以顾客为中心，具有适应性和包容性的政策制定过程^[23]。清华大学薛澜教授认为，敏捷治理的核心思想是针对新兴技术分享提出的一种治理理念和改革目的，两大显著特征是参与广泛度和时间灵敏度^[24-25]。此外，我国国家新一代人工智能治理专业委员会发布《新一代人工智能治理原则——发展负责任的人工智能》文件中也提出，“和谐友好、公平公正、包容共享、尊重隐私、安全可控、共担责任、开放协作和敏捷治理”八项人工智能治理的重要标度，这意味着敏捷治理在中国开始从社会治理领域拓展到人工智能领域^[26]。

基于敏捷治理的演变趋势,能发现敏捷治理与传统治理并不一致,二者之间存在一定差异,敏捷治理作为一套具有全面性、动态化及精准性的机制,创新了传统治理模式,分别从治理目标、治理形式、治理过程、治理特点、治理工具以及治理反应六个维度来阐述,见表1。

表1 敏捷治理与传统治理的区别

比较项目	敏捷治理	传统治理
治理目标	可持续性、灵活性与竞争力	合法化、遵守法律法规
治理形式	互动型	管控型
治理过程	主动型	被动型
治理特点	精准性、柔和性、快速性	模糊性、强制性、滞后性
治理工具	数字信息工具	人力资源工具
治理反应	多变型	程序型

第一,治理目标,敏捷治理关注可持续性、灵活性与竞争力,通过灵活多变的方式实现治理可持续发展,从而提升治理效率;而传统治理自上而下进行,以遵守法律法规为底线,较为稳妥^[27]。第二,治理形式和治理过程,敏捷治理强调主动与相关主体间的交流互动,征集不同意见,在变化中求变,给予良好反馈;而传统治理采用管制形式,被动接受上级安排,缺乏协作性。第三,治理特点,敏捷治理注重精准性、柔和性、快速性的要求,而传统治理却十分模糊,具有强制性和滞后性。第四,治理工具和治理反应,敏捷治理采用数字信息工具精准识别问题,节约成本和资源,由主观人为经验向客观数据驱动转变,属于多变型反应治理方式;而传统治理通过人工方式和粗糙资源手段进行治理,反应较慢,属于程序型反应治理方式^[28]。

因此,在教育人工智能的伦理问题愈发加剧、治理方式不灵活的背景下,以探讨敏捷治理发展趋势和敏捷治理与传统治理之间区别为基础,为教育人工智能伦理问题治理带来全新蜕变,健全教育人工智能共建共享的治理体系^[29]。

四、教育人工智能伦理的敏捷治理模型构建

在明确教育人工智能伦理风险的基础上,立足传统治理方式存在的问题,结合敏捷治理的演绎趋势与内涵区别,研究参照敏捷治理的基本模型,依据敏捷治理的原则和能力,提出在治理者视角下教育人工智能伦理敏捷治理的五步循环模型,从而更加全面地和系统地解决教育人工智能伦理问题。

(一)敏捷治理的基本模型

本文采用的敏捷治理基本模型来自2022年1月

28日人工智能治理准则工作组发布的《实施人工智能原则的治理指南》,其认为每个管理行为者(包括企业、政府、非政府组织等各种行为者)都需执行五步程序模型,并提出每个程序的行动目标^[30],如图2所示。

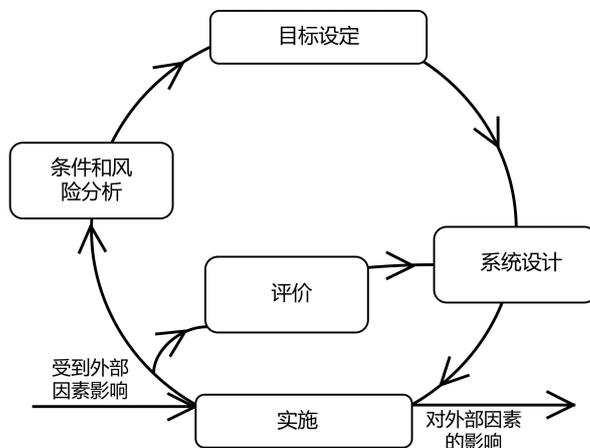


图2 敏捷治理基本模型

1. 条件和风险分析

条件和风险分析指治理者应不断分析外部条件,关注这些条件会发生哪些变化,以及由此导致的风险状况。包括:(1)了解积极和消极影响;(2)了解社会接受程度;(3)了解人工智能的熟练程度。

2. 目标设定

目标设定指治理者应设定治理目标,并根据外部条件的变化和技术影响不断审查这些目标。包括:(1)考虑人工智能治理目标;(2)设定人工智能治理目标。

3. 系统设计

系统设计指治理者根据设定的目标设计治理系统,除了技术系统外,还包括组织系统及其适用规则。在此过程中,需遵循透明和问责制、提供适当质量和数量上的选择、利益相关者参与、适当分配责任及提供补救措施的基本原则。包括:(1)分析差距和解决差距;(2)提高人工智能管理人员文化水平;(3)通过合作加强人工智能管理公司间关系;(4)减少用户事故发生率。

4. 实施

实施指实施设计的治理系统的过程。治理者根据实时数据和其他输入数据,持续监测系统的运行状况。此外,向可能受影响的利益相关者适当披露有关事项的信息,如系统产生的风险、运行设置以及补救措施等。包括:(1)确保能解释人工智能治理系统的实施状况;(2)确保能解释各人工智能系统的运行情况;(3)考虑将人工智能治理的实施情况列为非财务性信息,主动披露。在此过程中,会受到外部因素的影响,也对外部因素造成影响。

5. 评价

评价指治理者评估最初设定目标是否已完成。如果设定目标没有实现,系统就会重新设计。包括:(1)验证人工智能治理系统的设计和实施是否有效;(2)寻求利益相关者的反馈意见。

6. 对条件和风险的重新分析

对条件和风险的重新分析指治理目标因外部因素造成的影响而不得不修订,此时应分析治理系统的运行条件或风险状况是否发生变化,如果发生变化,是否需对目标重新修订,这是回到循环第一步的过程。

(二)教育人工智能伦理敏捷治理的五步循环模型

敏捷治理的原则指敏捷治理的过程中应遵守的基本原则,包括:灵活性、快速性、主动性、多元性与公众性。敏捷治理的能力指敏捷反应所具备的能力素质,包括:感知能力、反应能力、决策能力、动员能力、执行能力、协调能力。基于上述敏捷治理基本模型,依据敏捷治理的原则与能力,结合教育人工智能伦理的风险特点与传统治理依然存留的问题,本文提出教育人工智能伦理敏捷治理的五步循环模型,如图3所示,以期改善教育人工智能伦理问题的治理效果,提高治理效率。

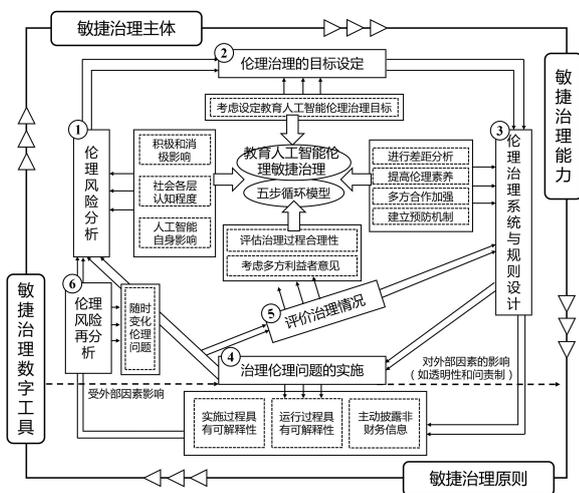


图3 五步循环模型

1. 伦理风险分析

伦理风险分析指治理者应不断总结目前发生的所有教育人工智能伦理问题,并分析这些问题是由哪些原因造成,会给教育主体带来哪些伤害。包括:(1)了解教育人工智能伦理问题并及时记录;(2)了解社会各界人士对伦理问题的认识程度;(3)了解开发教育人工智能产品的企业、研究所等对问题的认识程度。

2. 伦理治理的目标设定

伦理治理的目标设定指治理者应根据分析情况,设定治理目标,根据外部条件、影响因素、反馈情况和

人工智能技术的发展不断调整目标设定。包括:(1)分析与衡量教育人工智能伦理问题治理目标;(2)设定教育人工智能伦理问题治理目标。

3. 伦理治理系统与规则设计

伦理治理系统与规则设计指治理者根据设定的治理目标设计教育人工智能伦理问题治理系统。在治理系统中,包括每一项人工智能产品或技术存在的问题与对应治理方案,还包含系统治理的组织架构体系与各层组织应遵守的规则。包括:(1)通过数字化信息工具分析教育人工智能产品或技术等产生的伦理问题,提出解决方案;(2)提高开发人员、管理人员、教师等的教育人工智能伦理素养;(3)以政府为抓手,促进企业合作,形成业界认同的治理规则;(4)降低教育人工智能使用者发生事故概率。

4. 治理伦理问题的实施

治理伦理问题的实施指设计治理系统后实施的过程。系统的管理者需实时录入并监测数据,保证系统能正常运行,在事故发生后按照准备的解决方案及时处理,形成快速反应、系统的循环流程。在实施过程中,仍需注意向可能会受到影响的相关利益者适当披露有关信息。包括:(1)教育人工智能伦理问题治理系统的流程遵守可解释性;(2)主动向外界透露治理过程中所涉及的非财务信息。在此过程中,会受到外界正向或负向的影响,也可能对外界产生正向或负向的影响。

5. 评价治理情况与再分析

评价治理情况与再分析分为两部分:第一是评价教育人工智能伦理问题治理系统的实施情况,评估与最初设定目标是否吻合、是否具有合理性,征集多方意见,进一步改善治理系统;第二是对伦理风险再分析,在征集评价过程中,关注每种伦理风险是否发生变化,若有变化则需再次进入循环,重新设定目标并录入治理系统中,两部分同时进行,以提高治理效率。

五、教育人工智能伦理敏捷治理的实践路径

从教育人工智能伦理治理的五个问题出发,构建教育人工智能伦理敏捷治理的五步模型,以敏捷治理的原则与能力为实现模型的辅助基础,打造教育人工智能伦理敏捷治理的五层实现路径:分析层、目标层、设计层、实施层、评价层,并以本文上述归纳的“PBAS”教育人工智能伦理问题为例,详细阐述通过敏捷治理解决的实现路径,如图4所示。

在五层实现体系中,分析层是第一层,是对伦理问题的预见性分析与收集,多方主体灵活参与,具备快速感知与动员的能力,并将其成因、了解与认知程

度的数据分析提供给下一层。目标层通过分析情况设定符合主旨的治理目标,具有决策能力和快速性。设计层根据设定的目标设计教育人工智能伦理的治理系统,在系统中以数字化工具作为辅助分析的基础,建构每类教育人工智能技术或产品出现伦理问题的解决准则与方案,注重加强主体间合作和提升整体伦理素养水平,具备多元性、主动性和动员能力。实施层按照设计完成的治理系统对出现的伦理问题进行处理,处理过程中遵守可解释性、透明性等原则,具备协调能力、执行能力和灵活性。评价层是采纳各层主体意见并与实施状况做出对比与调整,具有公众性。一旦发现对比情况不理想或与预想目标之间差距较大,则再回到分析层继续循环治理。

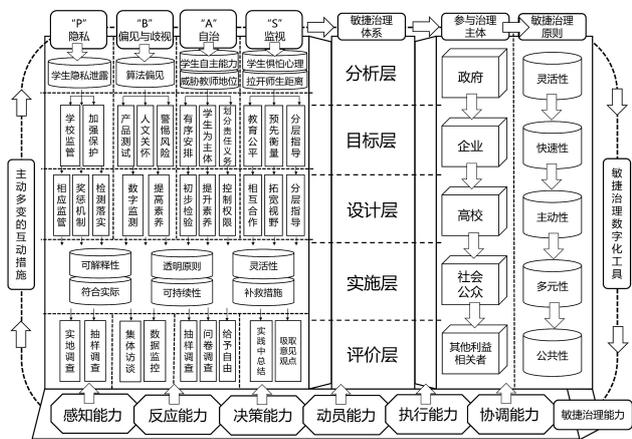


图4 五层路径体系

(一) 隐私问题的敏捷治理

在第一类——隐私问题中,多数以学生隐私泄露情况为主。教师可监测学生使用过程中出现的问题,进行样本采集与存档,上报学校进行汇总处理,再交由企业与政府来分析不同地区、不同人工智能的产品共性,总结优缺点。例如,采集学生数据带来的好处是什么,带来的影响是什么,并积极调查学生、教师、学校管理者对其认知程度,详细分析产生此类情况的原因。在分析完成后,得知发生隐私泄露问题可能是由于学校监管不力,教师对学生隐私保护意识不强,因此造成伤害。治理目标则需加大对学校的监管力度,加强学校对学生的保护。根据治理目标,在治理系统中将解决方案定为:政府发布政策落实监管机制,学校建立合适的监管与奖惩机制督促教师注意保护和关心学生,在教学过程中注意严密性,并利用信息化采集工具检测落实程度,对实施情况进行统计,可通过实地调查法、抽样调查法等对学生是否受到了隐私泄露的威胁进行征集,再调整治理方案。

(二) 偏见与歧视问题的敏捷治理

在第二类——偏见与歧视问题中,人工智能背后的算法偏见与歧视是最为突出的情况,这与开发人员具有较大关系,以企业作为分析主体,教师、学生与政府等作为辅助人员,明晰偏见与歧视的背后逻辑,并调查大众对其的意识。根据掌握情况设定治理目标,主要与开发人员有关,加大对教育人工智能的产品测试,提醒教师仍应多关心学生身心健康的发展,尤其注意对已受到偏见与歧视的学生提供帮助,如调整学习方式与心理爱护。根据治理目标,在治理系统中将解决方案定为:企业要求开发人员在产品中嵌入算法监测系统,通过数字化监测工具掌握使用情况,根据收集数据调整关于男女性、容貌设定、巡逻行为等设定,提高开发人员的教育人工智能伦理素养。最后,剖析日常教育环境中真实发生的现象,可采用集体访谈法、问卷调查法等了解学生是否遭受到或意识到未知的偏见与歧视,加强偏见检测和道德评估,为后续的发展做出改正^[31]。

(三) 自治问题的敏捷治理

在第三类——自治问题中,教育人工智能产品可能会剥夺学生的自主学习能力,威胁教师地位,与教师的教学能力和学校管理者对其规制的要求有关,有许多学生并没有意识到自主决策能力的降低,需采纳各层人士对其辅助教学的看法,采纳是否给教师带来职业焦虑和压力的意见。根据了解的情况设定治理目标,即在人工智能课程中有序安排教学活动,坚持以学生为主体,同时对人工智能的责任与义务进行明确划分。根据治理目标,在治理系统中将解决方案定为:学校需对人工智能课程的教学安排进行初步检验与评估,从小班教学拓展至大规模,借助信息化工具收集教师与学生使用过程的真实感受,提升学校管理人员的人工智能伦理素养,通过企业合作控制人工智能的地位权限。最终采取抽样调查法、问卷调查法等探讨教育人工智能自治问题是否被有效解决,还学生真正地自由,降低教师的职业焦虑。

(四) 监视问题的敏捷治理

在第四类——监视问题中,常见情况是通过数据采集型的教育人工智能产品对学生的面部表情、语言行为、学习方式等进行监视,使学生产生惧怕心理,从而拉开师生间距离。这与教师各类行为控制具有密切联系,其一言一行始终对学生产生较大影响,需通过多方深度调查才能了解学生内心真实想法。因此,治理目标应是预估教师水平、素养、性格等多方面因素,开展考核机制对教师进行预先衡量,并通过考核机

制,大幅提升教师的职业专业度。根据治理目标,在治理系统中将解决方案定为:企业与学校、学校与学校间互相合作传播教育人工智能的理念与优势,培训教师使用的注意事项,拓宽解决教育人工智能伦理问题的视野,利用数字化工具实施教师的分层指导,提供个性化伦理素养辅导。最终通过个别谈话、讲座、座谈会等形式收集教师反馈意见,在实践中吸取每次推进的意见,调整下一次解决方案。

智能化时代,教育人工智能的发展没有终点。因此,教育人工智能的伦理问题绝不仅只有本文中所

提到的四类,以后还将有更多现在已经出现人类并未意识到或还未曾出现的伦理问题^[32]。随着新一代人工智能、大数据等信息技术的快速发展,未来需借鉴敏捷治理的方式才能更加系统化、灵活性地治理教育人工智能伦理问题,按照教育人工智能伦理敏捷治理的五步模型和五层实践机制,推动教育人工智能健康化发展,为未来教育人工智能伦理问题的治理提供理论的支撑和方法的引领,提供教育人工智能可持续发展的机遇,建立人机和谐共生的教育生态体系。

[参考文献]

- [1] 中华人民共和国国务院. 国务院关于印发新一代人工智能发展规划的通知[EB/OL]. (2017-07-20) [2022-12-20]. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.
- [2] 孙田琳子. 论技术向善何以可能——人工智能教育伦理的逻辑起点[J]. 高教探索, 2021(5): 34-38, 102.
- [3] 托雷·霍尔, 曹梦莹, 明芷安, 等. 可解释人工智能的教育视角: 基于伦理和素养的思考[J]. 中国教育信息化, 2022, 28(4): 5-13.
- [4] 胡小勇, 徐欢云. 面向 K-12 教师的智能教育素养框架构建[J]. 开放教育研究, 2021, 27(4): 59-70.
- [5] 胡小勇, 黄婕, 林梓柔, 等. 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, 36.
- [6] 苗逢春. 教育人工智能伦理的解析与治理——《人工智能伦理问题建议书》的教育解读[J]. 中国电化教育, 2022(6): 22-36.
- [7] 邓国民, 李梅. 教育人工智能伦理问题与伦理原则探讨[J]. 电化教育研究, 2020, 41(6): 39-45.
- [8] AKGUN S, GREENHOW C. Artificial intelligence in education: addressing ethical challenges in K-12 settings [J]. AI and ethics, 2022: 431-440.
- [9] 田贤鹏. 隐私保护与开放共享: 人工智能时代的教育数据治理变革[J]. 电化教育研究, 2020, 41(5): 33-38.
- [10] 李世瑾, 胡艺龄, 顾小清. 如何走出人工智能教育风险的困局: 现象、成因及应对[J]. 电化教育研究, 2021, 42(7): 19-25.
- [11] SMITH H. Algorithmic bias: should students pay the price? [J]. AI & society, 2020, 35(4): 1077-1078.
- [12] 赵磊磊, 吴小凡, 赵可云. 责任伦理: 教育人工智能风险治理的时代诉求[J]. 电化教育研究, 2022, 43(6): 32-38.
- [13] 教育部. 联合国教科文组织正式发布国际人工智能与教育大会成果文本《北京共识——人工智能与教育》[EB/OL]. (2019-08-28) [2022-12-20]. http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/201908/t20190828_396185.html?from=groupmessage&isappinstalled=0.
- [14] 王佑镁, 宛平, 赵文竹, 柳晨晨. 科技向善: 国际“人工智能+教育”发展新路向——解读《教育中的人工智能: 可持续发展的机遇和挑战》[J]. 开放教育研究, 2019, 25(5): 23-32.
- [15] 杜静, 黄荣怀, 李政璇, 等. 智能教育时代下人工智能伦理的内涵与建构原则[J]. 电化教育研究, 2019, 40(7): 21-29.
- [16] The University of Buckingham. Interim report: towards a shared vision of ethical AI in education[EB/OL]. (2020-03-04) [2022-12-18]. <https://www.buckingham.ac.uk/wp-content/uploads/2021/03/Interim-Report-The-Institute-for-Ethical-AI-in-Educations-Interim-Report-Towards-a-Shared-Vision-of-Ethical-AI-in-Education-1.pdf>.
- [17] 新华社. 发展负责任的人工智能: 我国新一代人工智能治理原则发布[EB/OL]. (2021-07-03) [2022-12-18]. http://www.gov.cn/xinwen/2019-06/17/content_5401006.htm.
- [18] 张立国, 刘晓琳, 常家硕. 人工智能教育伦理问题及其规约[J]. 电化教育研究, 2021, 42(8): 5-11.
- [19] 高山冰, 杨丹. 人工智能教育应用的伦理风险及其应对研究[J]. 高教探索, 2022(1): 45-50.
- [20] 侯浩翔, 钟婉娟. 人工智能视阈下教育治理的技术功用与困境突破[J]. 电化教育研究, 2019, 40(4): 37-43, 58.
- [21] 钱小龙, 张奕潇, 宋子响, 等. 教育人工智能系统的伦理原则与困境突破[J]. 江南大学学报(人文社会科学版), 2021, 20(6): 96-104.
- [22] 于聪, 刘飞. 人工智能教育应用的伦理风险及其对策研究[J]. 机器人产业, 2022(2): 32-37.
- [23] World Economic Forum. Agile governance: reimagining policy-making in the fourth industrial revolution [EB/OL]. (2018-04-24) [2022-12-21]. <https://www.weforum.org/whitepapers/agile-governance-reimagining-policy-making-in-the-fourth-industrial-revolution>.

- [24] 薛澜. 第四次工业革命将给我们带来什么?[J].金融经济,2018(13):21-22.
- [25] 薛澜,赵静. 走向敏捷治理:新兴产业发展与监管模式探究[J].中国行政管理,2019(8):28-34.
- [26] 发展负责任的人工智能:新一代人工智能治理原则发布[J].科技与金融,2019(7):2-3.
- [27] 吴磊,冷玉,唐书清. 数字化时代敏捷治理的学术图景:研究范式与实现路径[J].电子政务,2022(8):77-88.
- [28] 颜海娜,唐薇,王露寒. 基层重大公共卫生风险防控中的敏捷治理机制——基于 H 街道的新冠肺炎疫情防控个案分析[J].华南理工大学学报(社会科学版),2021,23(1):113-124.
- [29] 詹绍文,赵雪. 基于敏捷治理的智慧社区建设:机理分析与案例考察[J].改革与开放,2022(13):30-39.
- [30] AI Governance Guidelines WG. Governance guidelines for implementation of AI principles [EB/OL]. (2022-01-28) [2022-12-28].
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_2.pdf.
- [31] 沈苑,汪琼.人工智能教育应用的偏见风险分析与治理[J].电化教育研究,2021,42(8):12-18.
- [32] 王佑镁,王旦,柳晨晨. 从科技向善到人的向善:教育人工智能伦理规范核心原则[J].开放教育研究,2022,28(5):68-78.

Agile Governance: A New Model of Ethical Governance for Educational Artificial Intelligence

WANG Youmei, WANG Dan, LIANG Weiyi, LIU Chenchen

(Research Center for Big Data and Smart Education, Wenzhou University, Wenzhou Zhejiang 325035)

[Abstract] Since entering the 21st century, human beings have entered the era of artificial intelligence with exponentially increasing data scale, and artificial intelligence in education has become an inevitable trend, but there are also unavoidable risks and challenges, leading to a series of ethical issues. This paper summarizes the ethical risks of educational AI into four categories of "PBAS", analyzes five existing shortcomings of ethical risk governance, and proposes a new governance model, agile governance, as an innovative response to the ethical governance of educational AI. This study takes the trend of agile governance and its connotation characteristics as the entry point, combines the principles and capabilities of agile governance, constructs a five-step cyclic model of ethical agile governance of educational AI, designs a five-level practice path, and elaborates agile governance solutions using the four categories of "PBAS" as examples. The results provide a new governance paradigm for the ethical issues of educational AI, and provide a reference value for the sustainable governance and healthy development of ethical risks.

[Keywords] Educational Artificial Intelligence; Ethical Issues; Agile Governance; Governance Model