

【网络社会变革与管理】

应用伦理学视阈下人工智能教育的反思与应对

张志华, 季 凯

(南京邮电大学 教育科学与技术学院, 南京 210023)

摘 要:人工智能的深入发展和广泛应用,深刻影响着教育领域的变革与创新。人工智能教育伦理是对技术伦理视角下教育伦理的再塑,主要面向解决人工智能教育的应用困境。应用伦理学的理论基础来源于五种规范伦理学进路:功利论、义务论、契约论、正义论和美德论,以该理论为基础分析人工智能教育当前主要伦理困境,呈现出功利主义技术观与非功利主义教育观的矛盾、技术与教育多重主体行为的矛盾、技术正当性与个人权利的矛盾、技术引发教育平等与社会不稳定性的矛盾、技术动机与教育本质的矛盾,需从制度、个体和组织三个层面建构人工智能教育生态。

关键词:人工智能;应用伦理学;技术伦理;教育伦理

中图分类号:G40-059.1 **文章编号:**1673-5420(2021)05-0001-10

引 言

人工智能的发展促使许多行业发生了深刻变革,其与工业、医疗、金融、交通等呈现深度融合趋势,但在教育领域却进展缓慢。当前的教育模式仍以传统教育为主导,人工智能虽作为一种辅助性工具被广泛运用于教与学的各个环节,但革命性的变革尚未出现。也正因如此,构建规范的人工智能教育伦理框架,预见性地规避人工智能教育伦理风险,结合现实开展理论研究成为一种现实需要。

收稿日期:2021-08-13 本刊网址:<http://nysk.njupt.edu.cn>

作者简介:张志华,博士,南京邮电大学副校长、教授,研究方向:技术创新与知识管理。

基金项目:国家社科基金一般项目“乡村优质教育可及性与实现机制研究”(2020SJDZDA117);江苏高校哲学社会科学重点研究基地“教育人工智能研究中心”资助项目;2021年江苏省研究生科研创新计划项目“数智化驱动教育形态嬗变的作用机制与发展路径研究”(KYCX21_0819)

20世纪20年代美国俄亥俄州心理学教授西德尼·普雷西(Sidney Pressey)设计了一系列机器用于教学,“机器教学”(Machine Teaching)这一概念由此提出。其中隐含的教育理念是对学生进行个性化培养,这与人工智能发展的本质不谋而合。人工智能发展的本质是追求人机协同,实现在不同应用场景下自主配合要素变化,协助人的工作,重新拉回“以人为本”的组织模式^[1],构建以人的需求为中心的生产模式^[2]。人工智能教育呈现智能化、自动化、个性化、多元化、协同化等特征^[3],有学者认为以人工智能为代表的数字技术与教育融合,将带来第四次教育革命^[4]。

技术的进步促进了教育发展,伴随着人工智能技术的发展和应用,教育生态逐渐被解构并重新建构。人工智能促进教育变革的同时,也引发学界在哲学层面对其进行了反思。从角色层面来看,邓国民等认为教育利益相关者将面临更加复杂的伦理困境^[5],教育的主体出现“类人”概念;从技术层面来看,苏明等认为需要通过伦理批判来改造人工智能教育异化的现象^[6];从社会层面来看,Luckin等认为人工智能教育存在数字鸿沟,并将带来社会不公平问题^[7]。同时,一系列问题接踵而至,如人工智能是否会取代教师?人工智能教育是解放还是弱化教育?人工智能教育行为规范是否道德?人工智能教育的合法性如何?伦理是维系社会秩序的规范,也是维系教育不偏离本质的规范,即将到来的人工智能教育变革要警惕存在伦理问题隐患。

一、应用伦理学:人工智能教育伦理反思的理论依据和遵循

人工智能教育已成为当前社会关注的热点问题,运用应用伦理学应对人工智能教育以及衍生的道德困境问题,是对人工智能技术发展的合理约束,也是确保教育向善发展的迫切需求。

应用伦理学的基础理论主要源于五种传统规范伦理学进路:功利论、义务论、契约论、正义论和美德论,其划分标准依据的伦理原则、主张等或有不同,却逻辑自洽^[8]。应用伦理学的发展离不开传统理论的支撑和指导,具有历史性、具体性、发展性等特点。对人工智能教育伦理问题的探究应遵循这些特点。人工智能教育的形态是技术和教育的交融,因此有必要多维度、多视角地探讨具体问题,规范和促进人工智能教育发展,并为其提供理性的道德指引。

技术伦理的现实道德问题集中在技术的本质、技术伦理评估、技术的可控性、技术的主体等。教育伦理的现实道德问题集中在教育的本质、教师的职业道德、教育公平等。人工智能教育伦理问题不应囿于对技术伦理或教育伦理的单向探讨,也不应带有

经典技术哲学的决疑论色彩,一味地强调其危害^[9],而应对人工智能技术与教育的关系进行解构,在伦理—人工智能—教育的整体关系视阈中将传统伦理理论植于人工智能教育,解决实际问题。

二、人工智能教育的现实道德问题及其应用伦理学分析

技术进步是社会发展的重要驱动力,初期往往面临着相关技术领域内的治理失调问题。人工智能教育伦理问题的本质在于技术与教育在应用中的伦理冲突,为契合应用伦理学解决实际道德问题之要义,本文依据功利论、义务论、契约论、正义论和美德论进行分析,以尽可能地展示问题的全貌。

(一)功利论:功利主义技术观与非功利主义教育观的矛盾

功利论,又称“功利主义”,主流学说为“行为功利主义”,主张把行为的结果作为判断行为是否道德的标准,认为正确的道德行为能够带来最大幸福。这里的幸福是指快乐或免于苦难,功利主义最大的追求是结果幸福,主张结果幸福大于行为正当。

功利主义技术观的目的是借助技术改善现状。人工智能技术介入教育是为了克服工业教育模式的弊端,如同质化的学生培养、抑制学生全面发展的需求等。非功利主义教育观与功利主义教育观相对,主张过程中的道德教育,教育伦理的最高价值目标是善^[10],培养个体的善,实现人的自由和价值,达到“教育即生活”和“生活即教育”的目标。

功利主义技术观与非功利主义教育观价值取向的冲突,引发了现实的道德问题。个性化教育被认为是人工智能教育的一个重要实践方向,人工智能教育呈现出“深度学习+大数据”的优势^[11],利用大数据分析和算法推荐,能够基于学生的学习能力、兴趣等提供定制化的教学内容。但是,利用人工智能实现个性化教学,并不一定能促进学生的个性化发展。

人工智能教育的原则需要与学生的主体性地位相适应,个性化的算法推荐提升了教学内容的精准传播,但也导致了学生获得的知识域变窄。长此以往,学生丧失了获取信息的自主权,学生的主体性逐渐被消解^[12]。现实中的一个重要表现就是个性化教育下学生“幼稚化”^[13],学生依赖于人工智能教育认知,失去了对现实生活的体验。如学习看地图、规划路线可以带给人满足感,能够帮助人理解空间关系,懂得如何通过历史、地理知识、社会学理论等来欣赏环境,而不单单以纯功利的心态想着怎么最快、最轻松地到达目的地,这其实是学习的重要乐趣之一。技术作用于教育将造成教育异化,导致

人工智能教育背离“教育即生活”和“生活即教育”的教育本质,因此,要坚决防止被技术异化的假象蒙蔽和迷惑。

(二)义务论:技术与教育多重主体行为的矛盾

义务论,又称道义论,侧重于评价人的行为或者某项活动的道德性。义务论追求行为与道德准则的一致,其正当性取决于所遵循的道德准则是社会公众普遍认可的,而非个体所信奉遵循的道德信条。

人工智能教育的动机、行为的道德评价,取决于不同的“人”,也就是主体。技术催生了“类人”的行为,改变了人与人的行为关系。人工智能教育过程中存在技术、教师、学生多重主体及主体异位现象,主体所遵循的道德准则不尽相同。人工智能教育动机受多重主体道德准则影响,因而判定人工智能教育行为是否具有正当性,需要从技术、教育及二者的辩证关系着手,运用逻辑学解析伦理问题中行为动机的多维道德准则,分析矛盾产生的根源,进而完善人工智能教育伦理准则。

用义务论审视人工智能教育伦理问题可以发现,技术主体的介入冲击了传统教育主体间的伦理关系。人工智能越界行为屡见不鲜,例如,2019年发生的亚马逊 Alexa 人工智能助手劝人自杀事件,就引发了责任界定问题。人工智能教育同样会出现越界问题,基于行为、动机的道德评价,其主体到底是技术、教师还是学生?随着深度学习算法的自我迭代,人工智能的行为变得越来越不透明,使得人工智能行为在出现偏差后,出现纠错困难。人工智能行为选择的原因日益变得不可解释、问题解决变得日益不可控。依据现有法律,未来可能产生权责混淆、责任鸿沟等问题,进而对行为方式的道德性产生质疑。

主体关系的冲突也会引发其他现实道德问题,施教者的角色调整和学习者的角色转换,教师被人工智能取代的恐慌,学生在现实和虚拟之间出现“本我”与“非我”主体性异位^[14]。施教者过分依赖人工智能使得教师的身体缺位,可能会遮蔽教师在教育教学活动中的重要性,削弱教师的主体性地位;学习者脱离外部束缚,放大自身个性,甚至会形成“虚拟人格”,表现出一些“非常”行为,导致自我的现实性和虚拟性之间的张力和平衡被撕裂,自我陷入虚拟性异化陷阱。

(三)契约论:技术正当性与个人权利的矛盾

契约论,此处意指自利契约论,关注的重点是个人权利,判定标准取决于是否遵循道德原则,可以说是义务论的发展。自利契约论可以用一个公式表述:一个行动是道德的,当且仅当它符合自利理性的人们在非强制性社会条件下可能达成和遵守的契约或协议^[15]。契约论中的主动的权利指自由地追求自身利益的权利,被动的权利指言论自

由权、人身自由权、隐私权等。

人工智能教育中的“类人”是否具有权利是存疑的,“类人”的出现源于技术设计者,因此探讨技术是否具有正当性问题,可将技术设计者作为主体分析其权利。人工智能主体与教育主体间达成合作契约,人工智能主体有义务为教育提供服务,有权利要求教育遵循技术规律;反之,教育主体有义务为人工智能主体提供资源,有权利要求人工智能主体遵循教育规律。

以自利契约论探究人工智能教育伦理,按照霍布斯的社会契约论,契约的双方从自利出发都应放弃一部分权利,实现共同存在。应用于人工智能教育的现实情境中,以隐私权为例,利用人工智能技术挖掘数据是服务教育,而学习者放弃一部分隐私权等权利,存在一定的正当性。表面上看这一解释符合契约论的要旨,但从法律层面看,《民法典》第一千零三十五条强调了处理个人信息的正当性原则,个人信息的基础属性是个人身份信息,不完全属于个人隐私权,法理上无法明确界定数据挖掘的边界,这也为保护个人隐私权增加了困难。

人工智能教育中需要大量数据支撑理论模型,教师和学生的行为、情绪、学习效果等都是数据收集和实验的对象,算法对数据进行再加工,解读教师和学生行为的深层逻辑,这就会带来算法和数据的伦理问题。哪些数据该搜集?会不会发生数据意外泄漏?如何分析、解释和共享数据?如何在算法和数据挖掘中保护个体权利?一旦无法判定技术的正当性与否,技术的正当性和个人权利保护的矛盾必然会出现。

目前的人工智能教育是基于一定的技术规则而设计的,保留和放弃的权利都已被限定,需要考虑的是某些特定行为是否违反了个人的主动和被动权利,框定和约束技术的某些特定行为。

(四) 正义论:技术引发教育平等与社会不稳定性的矛盾

正义论,也称非自利契约论,是义务论的另一种发展方向。其与契约论的相同之处在于都主张权利优先于善,关注的重点是个人权利,不同之处在于非自利契约论以正义原则作为首要原则,在该原则的指导下人们自愿结成社会契约,由正义原则确定社会基本结构。

罗尔斯在20世纪70年代提出了两个正义原则。第一个原则主张同等自由的制度,自然状态下正义社会中的公民拥有平等的权利。对于人工智能教育而言,人工智能追求在教育中保障人的平等受教育权利,在技术的最大限度下每个人都有平等的权利选择教育,而不是每个人都必须接受同样的教育。第二个原则也被称为差异原则,认为正义的社会未必是平均主义的社会,自然状态下权利的分配有利于“最小受惠者”,即符合

正义原则。人工智能教育能够改变教育资源分配,使处在社会底层的也能享有受教育的权利。然而,阶层间的教育水平差异实然存在,人工智能加大了教育水平的阶层差异,加剧了社会阶层的不稳定。

由第一个原则推论出人工智能教育促进教育平等,由第二个原则推论出人工智能教育引发社会不稳定性,两者间的矛盾将成为人工智能教育发展面临的问题。教育平等指人们不受政治、经济、社会地位和民族、种族信仰及性别的限制,都享有受教育的权利。教育智能化、自动化等能够促进教育平等实现,教师、学生能够不受客观因素限制享受技术带来的益处。

人工智能算法模型和数据挖掘具有技术偏向性,能够根据主体特征提供定制化服务,但教学质量、学习质量的构成维度仍有待商榷。人工智能的本质是追求人机协同,即使在设计环节全力降低人为因素的影响,但仍会向行为主体的偏好倾斜。因此,人工智能教育产品仍有可能生成带有部分群体价值偏向的决策^[16]。这是人工智能教育发展初期阶段必然要面临的问题。

(五)美德论:技术动机与教育本质的矛盾

美德论,又称“美德伦理学”,通常以人的美德和品质为中心,将培养人的美德作为道德教育和道德实践的首要功能。功利论和义务论的伦理学在讨论道德的善恶时,主要集中在“我们应当做什么”,注重行为或实践应当遵守的原则、规范;美德论的伦理学,则主要讨论的是“我们应当成为怎样性质的人”。实际上,儒家的育人理念一贯倡导培养有德性的人,中国传统文化具有美的向度^[17]。教育以培养人、塑造人、解放人为目的,本质上是一项崇善的事业。

人工智能的道德立场并非完全中立,在其设计和使用过程中也蕴含了道德价值,如不伤害使用者等。其道德立场存在两个场域:一是设计者场域,人工智能受到设计者的道德价值影响,产品往往带有设计者的价值偏好;二是使用对象场域,使用者的价值取向会逐渐改变人工智能产品,使其道德价值与使用者趋于一致。

人工智能教育的研发主要是以企业为主,以盈利为最终目的。资本的逐利性决定了人工智能教育存在利益上的追求,而教育作为一项公共事业难免会被侵犯,无法实现纯粹教人向善的初衷,这与美德论强调的教育本质背道而驰。国外学者在测试 IBM、旷视、微软人脸识别产品中发现,识别的准确率受性别影响,三个产品中对男性人脸识别的准确率均高于女性^[18-19]。由此可见,现有的人工智能技术尚不成熟,在教育领域中难免会受到性别等外部因素影响,进而对决策产生误导。

目前关于人工智能教育对学生成长 and 发展的影响尚无定论,但企业为了抢占市场,

单纯宣传人工智能教育产品对知识管理、学习效率等产生的正向影响。在未经过严格的科学验证的情况下推广人工智能教育产品,将对教育产生不可预估的负面影响。如果学校管理层或老师基于这些数据做决策,有可能会给学生带来心理伤害^[18]。因此,有必要去讨论技术的德性问题。

三、应对人工智能教育伦理行为失范的可行策略

人工智能教育的现实道德问题主要是在技术与教育耦合过程中出现的,问题产生的深层原因是制度不完善、个体意识缺失和组织结构固化,因而,需要采取相应对策,建构技术的非功利性约束机制、动机过程的透明机制、个人权利的法律保障、社会阶层的道德教育和技术德性的准则制度。

(一) 制度层面:完善人工智能教育伦理规范

人工智能教育的发展超前于其伦理发展,这也是造成人工智能教育伦理行为失范的客观原因之一。从逻辑上讲,完善人工智能教育伦理规范,主体方能有章可循、有规可依。因此,首先是从制度层面完善人工智能教育伦理规范,最大限度地呈现人工智能教育主体的共同伦理价值及伦理承诺。其次是抓紧推进人工智能领域法律体系的完善。当前,人工智能相关的法律大多体现在网络信息安全的相关规定中,内容滞后于人工智能技术发展,需要进一步修正、充实,并允许部分地区、行业先试先行。再次是推动人工智能教育伦理评价制度化。建立第三方伦理评价机构,健全伦理评价机制和伦理问责制度。在欧盟发布的《可信赖的人工智能伦理准则》中,问责是重要内容,其包含着明确利益相关责任主体,并对算法、数据等审查、评估和监管等,同时还可以记录和修复负面结果,即出现问题后能够确定责任方是谁,及时进行评估、监管和修复。最后是推动社会舆论对人工智能教育的监督。努力为社会舆论提供发声渠道,建立舆论监督的长效机制。

(二) 个体层面:培养人工智能教育伦理意识

人工智能教育的行为主体是人,人的道德准则和伦理规范受政治、经济等社会因素以及个体心理因素的影响。因此,需要在个体发展的各个环节长期培养伦理意识,将伦理意识内化为行为规范,最终推进社会共识的形成。

首先,高站位开展人工智能教育伦理普及与教育。提升人工智能相关学科教师的职业道德素养,教学内容规划应明确人工智能教育伦理的教学目的、内容及方法,对企业或其他利益组织通过“组织培训+自我学习”的方式进行培养。其次,高起点培养人工

智能教育安全监管人员。人工智能教育监管的困难在于对技术、教育和法律兼顾的人才较少,亟需培养复合型人才。最后,高标准提升社会公众尤其是青少年的道德素养。各类学校要结合人工智能教育的特征,尊重虚拟环境中青少年的特点和思想品德建设规律,将道德培养融入课堂。

(三)组织层面:重构人工智能教育内外组织结构

影响人工智能教育伦理行为的因素,除了个体自身因素外,人工智能教育的内在组织和外部组织也将影响个体伦理行为,如组织结构、伦理氛围等。由于技术动机受技术设计者所在组织的伦理氛围影响,人工智能教育主体的伦理自主性极易被抑制。

首先,建立网络组织结构。对人工智能教育主体所在的组织进行去科层化的组织结构改造,依托人工智能等技术实现信息共享和无障碍沟通,重塑教育组织形态。其次,建设关怀型组织伦理氛围。加强组织的人文关怀建设,尤其是对人工智能企业加强管理,营造向善的伦理氛围;加强组织德性教化,唤醒组织成员的良知,重视道义与利益的平衡,促进技术动机与教育本质的协调发展。最后,建立人工智能教育道德委员会,成员应包含人工智能与教育行业管理人员、技术专家、伦理学家及师生代表等。

人工智能与教育深度融合,给教育发展带来无限可能。采取应用伦理学视角分析人工智能教育伦理问题,在无限可能和无限危机的发展过程中,实现教人向善是人工智能教育的发展趋势,坚守教育本质是人工智能教育的发展共识,促进教育平等是人工智能教育的现实展望,而支撑这一切的则是社会认可的合理、健康、可信赖的人工智能教育伦理准则。

参考文献:

- [1] 李晓华. 人工智能+制造转型升级[N/OL]. [2018-06-13]. http://dz.jjckb.cn/www/pages/webpage2009/html/2018-06/13/content_44297.htm.
- [2] “新一代人工智能引领下的制造业新模式新业态研究”课题组. 新一代人工智能引领下的制造业新模式与新业态研究[J]. 中国工程科学, 2018(4): 66-72.
- [3] 梁迎丽, 刘陈. 人工智能教育应用的现状分析、典型特征与发展趋势[J]. 中国电化教育, 2018(3): 24-30.
- [4] 安东尼·塞尔登, 奥拉迪梅吉·阿比多耶. 第四次教育革命: 人工智能如何改变教育[M]. 吕晓志, 译. 北京: 机械工业出版社, 2019.

- [5] 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(6):39-45.
- [6] 苏明,陈·巴特尔.人工智能教育伦理的多维审视:基于马克思技术批判和人的全面发展理论[J].西南民族大学学报(人文社科版),2019(11):223-228.
- [7] LUCKIN R, HOLMES W, GRIFFITHS M, et al. Intelligence unleashed: an argument for AI in education [M]. London: Pearson, 2016.
- [8] 王正平.应用伦理学[M].上海:上海人民出版社,2013.
- [9] 芦文龙.技术主体的伦理行为:规范、失范及其应对[M].北京:人民出版社,2019.
- [10] 糜海波.教育伦理建设引论[M].北京:中国社会科学出版社,2015.
- [11] 李晓华,曾昭睿.前沿技术创新与新兴产业演进规律探析:以人工智能为例[J].财经问题研究,2019(12):30-40.
- [12] 冯锐,孙佳晶,孙发勤.人工智能在教育应用中的伦理风险与理性抉择[J].远程教育杂志,2020(3):47-54.
- [13] 袁春艳,刘珍珍.人工智能时代大学生思想政治教育的变革与因应研究[J].重庆邮电大学学报(社会科学版),2020(4):93-100.
- [14] 孙田琳子.虚拟现实教育应用的伦理反思:基于伯格曼技术哲学视角[J].电化教育研究,2020(9):48-54.
- [15] 邢雁欣,龚群.规范伦理学的四种进路[N].光明日报,2017-09-25(15).
- [16] 刘金松.大数据应用于教育决策的可行性与潜在问题研究[J].电化教育研究,2017(11):38-42+74.
- [17] 李盛婷.新媒体时代中华优秀传统文化美的向度及功能[J].南京邮电大学学报(社会科学版),2020(2):95-102.
- [18] BUOLAMWINI J, GEBRU T. Gender shades: intersectional accuracy disparities in commercial gender classification [C] // Conference on fairness, accountability and transparency. PMLR, 2018: 77-91.
- [19] 沈苑,汪琼.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(4):18-34+184.

(责任编辑:张秀宁)

Reflection and response to the artificial intelligence education from the perspective of applied ethics

ZHANG Zhihua, JI Kai

(College of Educational Science and Technology, Nanjing University of Posts and Telecommunications,)
(Nanjing 210023, China)

Abstract: The in-depth development and wide application of artificial intelligence have a profound impact on the reform and innovation in the field of education. Artificial intelligence education ethics is the re-shaping of education ethics from the perspective of technology ethics, which is mainly aimed at solving the application dilemma of artificial intelligence education. The theoretical basis of applied ethics comes from five normative ethical approaches: utilitarianism, deontology, contract theory, justice theory and virtue theory. Based on the above basis, this paper analyzes the current main ethical dilemma of artificial intelligence education, showing the contradiction between utilitarian technology view and non-utilitarian education view, the contradiction between technology and multi-subject behavior of education, the contradiction between technological legitimacy and individual rights, the contradiction between education equality and social instability caused by technology, and the contradiction between technological motivation and the essence of education. It is necessary to construct artificial intelligence education ecology from three aspects: system, individual and organization.

Key words: artificial intelligence; applied ethics; technical ethics; educational ethics