

# 人工智能教育应用伦理研究论纲

李晓岩<sup>1</sup> 张家年<sup>1</sup> 王丹<sup>2</sup>

(1. 淮北师范大学 教育学院, 安徽淮北 235000; 2. 淮北师范大学 体育学院, 安徽淮北 235000)

**[摘要]** 人工智能已成为影响未来社会发展的重要变量,其教育应用价值在获得认可的同时,也引起了伦理隐忧。教育伦理范畴视角下人工智能教育应用伦理的内涵,即教育教学过程中“人”与“人工智能”间的道德关系。人工智能教育应用的伦理旨趣在于“助”,而助什么和如何助是研究的两个关键问题。人工智能教育应用的伦理原则应从教育主体维度展开,其中教育管理者需遵循权力规约原则,具体表现为数据收集的尊重原则、数据处理的合理原则和数据利用的边界原则;教师需遵循和谐共生原则,具体表现为以人为本原则、适度原则和关怀原则;学习者需遵循形塑自我原则,具体表现为良善原则、诚信原则和节制原则。人工智能教育应用的伦理实践过程应遵循从个体的伦理认知、伦理行为到社会的伦理制度,再到个体认知和行为伦理的轨迹,即由个体走向社会,又由社会反作用于个体的螺旋循环过程。但人工智能伦理研究的两种研究范畴——技术伦理和教育伦理杂糅交错,不利于伦理实践的开展,希冀跨学科学者能够专注于教育伦理范畴的研究。

**[关键词]** 人工智能教育应用; 伦理旨趣; 伦理原则; 伦理实践

**[中图分类号]** G434

**[文献标识码]** A

**[文章编号]** 1007-2179(2021)03-0029-08

进入21世纪后,伴随着人工智能技术的巨大发展,“互联网+”教育逐步升级为“人工智能+”教育,尤其是近十年来,人工智能教育应用成为国际教育领域关注的热点。2019年3月联合国教科文组织(UNESCO 2019)发布的《教育中的人工智能:可持续发展的挑战与机遇》,既是对人工智能教育应用前期研究和实践的总结,也是未来各国人工智能教育应用发展方向的指南。习总书记指出,要高度重视人工智能对教育的深刻影响,积极推动人工智能和教育深度融合,促进教育变革创新(新华网,2019)。虽然人工智能目前并未对教育系统产生整体性、颠覆性的影响,但小范围的应用个案,如2019年BrainCo公司生产的“赋思头环”在某小学的应用,

已引起人们对人工智能教育应用的伦理担忧。

人工智能教育应用伦理属于教育伦理学的研究范畴。其研究过程应从“人”的角度出发,规约人在应用人工智能过程中的行为,而非关注人工智能的利害,是本研究的出发点和立场。需注意的是,与普遍意义上的教育伦理研究主体间的道德关系不同,人工智能教育应用伦理研究的是教育主体与具有一定主体属性的人工智能间的道德关系。人工智能同时具有人和物的两种属性:一方面人工智能没有自我意识,还是人造物;另一方面,它们又具有部分人的属性和能力,尤其是算法上,甚至比人类更强(何怀宏 2018)。因而,将人工智能与教育主体并列,而不是主客体二元对立,充分考虑人工智能的复杂

**[收稿日期]** 2021-02-26

**[修回日期]** 2021-04-23

**[DOI 编码]** 10.13966/j.cnki.kfjyyj.2021.03.003

**[基金项目]** 全国教育科学“十三五”规划课题“多维视角下人工智能教育应用的理论框架和实践路径研究”(BCA190082)。

**[作者简介]** 李晓岩,副教授,博士,淮北师范大学教育学院,研究方向:人工智能教育应用、教学系统设计(yt\_xiaoyan@163.com);张家年,教授,博士,淮北师范大学教育学院,研究方向:教育信息化、智库研究、人工智能教育应用(chzjn@126.com);王丹,讲师,淮北师范大学体育学院,研究方向:教学系统设计、信息技术教育应用(357627534@qq.com)。

**[引用信息]** 李晓岩,张家年,王丹(2021). 人工智能教育应用伦理研究论纲[J]. 开放教育研究, 27(3): 29-36.

性,有助于把握人工智能教育应用的本质。

本文主要围绕三方面展开:一是人工智能教育应用的伦理旨趣,即研究“人”与“人工智能”间的道德关系,目的是解决教育主体应用人工智能的伦理意图问题;二是人工智能教育应用的伦理原则,即研究“人”与“人工智能”间道德关系的依据,是教育主体应用人工智能达到教育目的过程中应遵循的道德规范;三是人工智能教育应用的伦理实践,即研究“人”与“人工智能”间道德关系的结果,主要探索教育主体应用人工智能为达到教育目的而开展的道德活动或表现出的道德行为。

## 一、“助”:人工智能教育应用的伦理旨趣

人工智能教育应用的伦理旨趣是人工智能教育应用伦理理论体系的核心,是评价人工智能教育应用善恶的准绳,是确定人工智能教育应用伦理原则和伦理认知与行为规范的依据。研究伦理旨趣就是研究人工智能教育应用的道德价值,是确立其价值体系、明确其价值目标的过程。

### (一) 伦理旨趣在于“助”

课题组在前期研究中已界定了人工智能教育应用的内涵,即在教育理念的引领下,将人工智能技术融入教育系统或教育过程中,支持或帮助实现教育目的,促进问题解决,优化教育过程和推动教育的创新发展(李福华等,2020)。可见,人工智能教育应用伦理旨趣在于“助”。

“助”凸显了人的主体地位,彰显了人工智能的价值追求,也体现了教育的本质。虽然人工智能与人可能形成“教师—机器人—学生”的新型交互结构(苏明等,2019),但二者不是平等关系。德国著名教育家卡尔·雅思贝尔斯(1991)曾言“教育是人的灵魂的教育,而非理智知识和认识的堆积。”人工智能虽然具有人的部分属性,但并不具有人的意识。教育始于“人”,终于“人”,这决定了人工智能只能是在效率、效益、效果上助力这一过程。

研究人工智能对教育之“助”,首先要回答两个问题,即助什么和如何助。第一个问题解决人工智能所“助”的内容。人工智能教育应用依据作用的领域可分为助学、助教、助管;第二个问题解决人工智能所“助”的方式手段。人工智能应用依据作用的层次可分为促进、优化、创新。由此,人工智能对

不同教育对象提供不同层次的助力,形成其完整的伦理旨趣(见图1)。

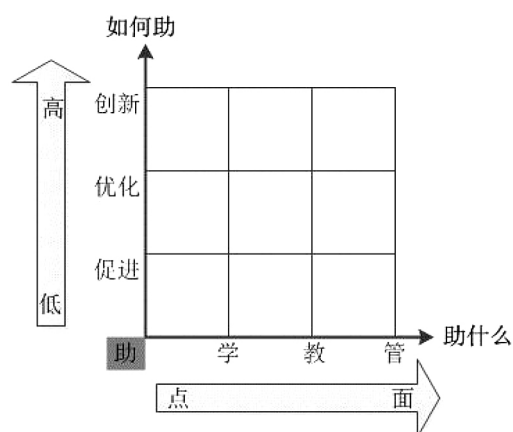


图1 人工智能教育应用的伦理旨趣

### (二) 助力内容: 由点到面

在横向上,人工智能对教育之“助”是由点到面的(见图1)。

首先,“助”个性化学习,即指人工智能助力受教育者学习。学习是个体行为相对持久的变化,是学习者改变原有观念的过程,因此学习者个体具象为一个点,在作用范围上可视为人工智能在教育领域“点”的应用。目前已有较多应用的自适应学习系统、自适应学习平台以及智适应学习平台等,主要通过学习分析、模式识别等技术,动态记录学习者的学习过程和学习状态,诊断并分析学习中的“漏洞”,给出解决方案,提供自动化反馈,提高学习效率;智能学伴主要通过与学生交流、互动和沟通的虚拟伙伴,帮助和引导学习者学习,记录学习进度、成长经历等(张攀峰等,2014),为学习者提供个性化学习服务。它通过智能传感技术和动态模型收集记录学习者的学习、情感及生理数据,感知学习者的学习状态和情绪变化,调整学习方案、给予学习者人文关怀。

其次,“助”精准化教学,即指人工智能助力教师教学过程。在现存教育制度下,班级授课和小组教学仍是主要教学组织形式,由此人工智能助力教师教学的作用范围已由学生个体的“点”扩大到教师引领的班级和小组学生群体的“线”。云班课、雨课堂、慕课堂等智慧教学工具或平台,贯穿于教学前的预习、教学中的互动讨论以及教学后的测试,在学习分析技术、语义识别技术以及模式识别技术的支

持下,充分调动学生的学习积极性,实施精准教学,提高教学效率。自动化短文评价系统、自动化口语评测系统、ICT技能与程序作业的自动化测评系统以及科大讯飞开发的智能测试系统等,依赖自然语言处理技术与机器学习等技术(张坤颖等,2019),提供及时反馈,为教学决策提供更真实可靠的依据。

最后,“助”智慧化管理,即指人工智能助力教育教学管理过程。管理不仅是对教师和学生的管理,还包括对学校层面所有教育教学活动的管理,因而其作用范围再次扩大到学校的“面”。智慧校园是人工智能管理应用的典型代表,它通过模式识别技术、自然语言处理、机器学习以及大数据技术实现对教师与学生活动的动态管理、精准管理,也为更好地服务教师和学生提供决策依据。高校学生普遍使用的一卡通管理已涉及到图书管理系统、打印复印系统、缴费管理系统、身份识别系统、课堂管理系统以及基于这些数据的师生行为轨迹分析和消费特征分析(田丽等,2015)。

### (三) 助力方式: 由低到高

首先是促进,即人工智能的应用旨在解决目前由教育制度不完备、教育资源不均衡、优质师资不足等因素引起的教育问题。可以说,在“促进”层面,人工智能应用围绕教育教学问题解决展开,是人工智能“使能”教育的表现。“使能”教学理念在班级授课制下很难得到实现,而人工智能通过学习分析和大数据技术构建高维学习者分析模型,可以诊断学习者的知识结构、认知特征、学习风格等,在一定程度上实现因材施教。

其次是优化,即在原有基础上应用人工智能优化教育教学,提高效率。“优化”主要通过日常教育教学达到预期,但应用人工智能可以更快、更好地达到目标,是人工智能“增能”教育的表现。某公司开发的AI助教(北极星,2020)利用人工智能深度学习算法,精准定位学习者的知识掌握程度,形成“学情画像”,教师可以准确把握学习者特征,从而设计针对性的教学;在教学过程中,AI助教提供海量优质教育资源,可以实现“智能组卷阅卷、课件制作、协同教研、资源共享、学情监测”等,自动生成多维度考试数据统计报告和考情分析报告,帮助教师提升教学绩效;在管理中,AI助教从多维度对教学行为数据进行挖掘与分析,为学校教育教学管理提供

决策依据,帮助家长了解学情,配合学校教学。

最后是创新,即伴随人工智能应用而产生和形成的新的教育教学和管理模式,是人工智能“赋能”教育的表现。人工智能作为“类人体”的存在,进入学校和课堂后,使教育从“批量灌溉”到“精准投放”、从“固定教材”到“生成内容”、从“流程繁复”到“过程消减”、从“总结评价”到“形成评价”(于英姿等,2020)。这使教育教学和管理必须做出改革,构建新的模式。联合国教科文组织研究报告提出的“双师课堂”便可看作是其创新成果之一(Pedro et al., 2019)。所谓“双师课堂”是一种由人工智能教育机器人承担教师的部分教学任务,并提供个性化学习服务的新型课堂模式(汪时冲等,2019)。人工智能承担部分教学任务后,教师的角色和职能、师生关系等必然进行调整,为新模式的构建提供了空间。

## 二、知善: 人工智能教育应用的伦理原则

伦理学是从善恶角度规约行为的。研究伦理原则就是明确何种行为是善的,何种行为是恶的。人工智能教育应用的伦理原则探讨教育主体应用人工智能达到教育目的过程中应遵循的道德规范,规约不道德行为。不同的教育主体因应用人工智能的目的不同需要遵循不同的伦理原则。教育主体包括教育管理者、教师和学习者。

### (一) 权力规约原则

权力规约原则主要针对作为权力主体的教育管理者,要求他们在应用人工智能的过程中要明确其权力范围,保护被管理者的权益和尊严,主要体现在教育管理者在教育数据的收集、处理、利用过程中,除大多数学者强调的数据隐私保护原则外,还包括数据收集的尊重原则、数据处理的理解原则和数据利用的边界原则。

1. 尊重原则。人工智能数据能力收集有限,收集哪些数据,应用主体具有一定的决定权。教育管理者应用人工智能收集相关数据时,应在尊重原则的基础上实施其决定权;教师和学生等数据被收集者有权利知道个人哪些数据被收集了、何时收集的以及收集的数据将做何用;非必要数据应给予教师和学生等主体一定的选择权,由他们决定这些数据是否被允许收集,以及在何种范围内使用。尊重原则是教育管理者应用人工智能时要遵循的首要原

则,是对其权力进行规约的道德力量,也是以人为本管理理念的体现。

2. 理解原则。在2020年第三届世界顶尖科学家论坛上,斯发基斯(Sifakis, 2020)谈到人与人工智能之间关系时强调,人类应充分理解机器传达的信息。这种理解即指要掌握这些数据的产生机制及其背后的理论机理。由此教育管理者才能进行正确的分析,并在DIKW模型(Data-to-Information-to-Knowledge-to-Wisdom Model)的引导下,得出指导性和决策性结论,减少片面臆断。当人工智能显示学生注意力不集中时,教育管理者要清楚人工智能判断学生注意力的机制(通过面部、脉搏抑或脑电等)。若通过面部识别进行判断,要了解人工智能对注意力不集中时的面部判别标准(眼神或者表情等),真正理解人工智能所谓的注意力不集中指的是什么,在此基础上,结合课堂情境和教师教学才能得出准确的结论。

3. 边界原则。获得数据后如何利用数据与教育管理者应用人工智能的目的直接相关。作为权力主体,教育管理者有权决定何时、何地、如何利用数据,或者说教育管理者对数据有绝对的使用权。因而,有必要对其数据利用权力设置边界,使其在既定范围内使用,防止滥用。一方面,数据利用的边界应在“理”而不在“管”,不能单纯用数据“管”(监管)教师、学生乃至学校,而是以此了解需求、“理”清问题,以更好地服务师生、促进学校发展为目的;另一方面,数据利用的边界应在“发展”而不在“评价”,不能以数据分析的结果作为评价教师、学生乃至学校的证据,而是以此为基础提出改进策略,促进教师、学生、学校的发展。教育管理者通过面部识别、情绪识别等技术收集课堂教学数据,对其分析,发现课堂纪律较差,不能以此判定教师教学能力低,教学考核不合格,而应以此为基础,与教师讨论改善课堂教学的方法,促进师生的共同成长。由此,课堂数据的收集才不会受到教师的心理抵制,更真实自然地表现自我,使人工智能获得更真实、有效的数据。

## (二) 和谐共生原则

和谐共生原则主要针对教师而言。人工智能这一类人体的教育介入必然影响师生的课堂关系。这一伦理原则主要体现在教师如何平衡和协调教师—学生—人工智能三者的课堂关系,确保和谐相处,共

存共生,各行其是而不逾矩,更高效地达到预期教学目标。

1. 人本原则,即学生中心原则,这是教师应用人工智能开展课堂教学要遵循的基本原则,也是处理教师—学生—人工智能三者关系的首要原则。澳大利亚学者塞尔温(Selwyn, 2016)教授曾指出“危险在于把数据和编码看成提供指导和支持的权威而非相对依据。教育如此复杂,远非可以仅被简化为数据分析和算法。”教师利用人工智能获取的学生学习相关数据,只能是教学设计、实施与评价的参考。一方面,人工智能本身获取的数据有限,不可能全方位获取学生的所有信息;另一方面,学生作为完整的个体,其自身所呈现出来的特征也远非数据所能涵盖。因此,教师应用人工智能进行教学时,不能只见“数据”不见“人”,而要从数据中看到一个(群)正在发展中的人因何、为何呈现这样的数据,要看到数据背后活跃的思维和灵动的气息。学生是学习的主体,教师要在这一过程中将焦点放在学生身上,通过数据了解和发掘其潜能,更好地促进其素养的提升,而非冰冷的数据本身。

2. 适度原则,是指教师要适度应用人工智能进行教学,包括适时和适量。所谓适时,指教师要选择恰当的时机应用人工智能,不能随意使用,否则将成为学生发展的障碍。扎瓦克奇-里克特(Zawacki-Richter et al., 2019)等学者对2007-2018年间发表的人工智能在高等教育领域的应用研究发现,目前研究与教学理论联系不紧密,提出要超越工具,重新关注学习和教学法。孔子曰“不愤不启,不悱不发。”人工智能的应用要遵循教育教学规律和原则,根据学生的发展需求,把握时机恰当使用。所谓适量,即指教师控制使用人工智能的频率,不能为了使用而使用,忽视教学目标的达成。人工智能通过“学习者画像”可以帮助教师更好地了解学生的知识基础和学习能力,为教学目标的确定提供依据。教学目标的确定除受课程标准等因素影响外,学习者的非智力因素同样重要,而这些是人工智能难以精准画像的。更重要的是,在学习者大脑发育的关键期过度使用认知技术,有可能带来不可逆转的严重后果,对发展更高级的认知技能 and 创新能力非常不利(Tuomi, 2018)。

3. 关怀原则,是指教师应用人工智能进行教学

的过程中要给予学生足够的关注和关怀,它是处理教师—学生—人工智能关系的润滑剂和催化剂。美国学者诺丁斯(2011)提出:关怀伦理的最基本思想是对每一个个体的需要予以恰当的反应,目的是建立并维护关心关系。一方面,教师要真正“看到”学生,透过人工智能和现实课堂,看到学生的需求,并给予恰当的反应。教师通过人工智能的数据发现某学生未能掌握相关内容,处于迷茫状态,则需要走到学生身边,详细询问情况,指导其学习,并通过鼓励的眼神、手势等给予学生情感投入和行为支持。另一方面,作为完整的个体,学生的发展不仅在于认知和思维的发展,还有社会性交往、情感、态度和价值观等,而目前这些方面人工智能还取代不了教师(唐汉卫,2018)。人类具有多元智能,人工智能可以超越其中理性化的智能,却难以超越人化的情感,人工智能更多是“智”而不是情(刘悦笛,2018)。教师需要在教学过程中消弭学生与机器交互的冰冷感,以人格魅力影响学生,以实际行为关怀学生,促进学生健康发展。

### (三) 形塑自我原则

形塑自我原则是针对学习者而言的。作为被教育者,学习者是一类特殊的主体。顾建军(2020)曾指出学生的主体性主要体现在对教师的主体性作用、属性、功能进行主体性选择,对教师主体性作用信息的破译与转化,以及对教师主体作用的借鉴、创造与超越。可见,学习者主体性的体现依赖于教师主体作用的发挥。但人工智能加入课堂后,作为类主体,其主体作用发挥有限。利用人工智能进行学习时,学习者占据主动地位,可以充分发挥主观能动性。但学习者是正在发展中的群体,其认知、情感、行为处于形成过程中,因而其应用人工智能旨在形塑自我,促进自身发展,须秉持以下伦理原则:

1. 良善原则,即指学习者在利用人工智能进行学习时需要心存良善,将人工智能作为“人”以待之。2016年微软聊天机器人Tay上线不久便学会了辱骂用户、发布带有种族歧视和性别歧视的言论,微软为此紧急下线调整(新浪,2016)。学习者在与人工智能交互过程中,自然语言处理技术和深度学习技术使人工智能习得了学习者语言表达模式。学习者尤其是未成年人,其道德水平尚在发展中,容易在好奇心驱使下使人工智能习得不良语言表达。因

而,学习者在与人工智能交互中,要将人工智能作为共同学习的“伙伴”,积极、平等地与其交互,使其真正成为学习助手。

2. 诚信原则,即指学习者恰当应用人工智能进行学习,不剽窃、复制人工智能的“成果”,坚持诚信为本。目前人工智能在许多专门领域已经持平甚至超越人类,即便是在一些曾被认为是人类所独有的领域,如作诗、绘画、唱歌等。当学习者将家庭作业交由人工智能完成时,教师是否能够清晰准确地判断该作业的真实性?互联网时代的剽窃行为尚可通过查重等手段加以监察,若由人工智能生成的呢?教师如何得知该作业是由学生而非人工智能完成的?解决这一问题,需要教师正确引导,使学习者本着对自己负责的态度,诚信为学,认真完成作业。这既是学生的本职要求,也是学生了解自己和提升自身学习能力的主要途径。

3. 节制原则,即指学习者要适度节制应用人工智能,以免过度依赖乃至上瘾。一方面,人工智能的类人性,易使学习者产生交往客体主体化,不利于其认知、情感的发展。另一方面,学习者应节制应用人工智能是由于其自制力差、心智发展尚不健全,容易在应用人工智能过程中迷失自己,甚至沉迷于与人工智能的交互。互联网时代的网络成瘾便是最好的例证。对于更智能、更个性化的人工智能,学习者将更缺乏抵抗力。

## 三、求真:人工智能教育应用的伦理实践

伦理学是一门实践哲学,不仅涉及纯粹理性的认知和实践理性的认知,还涉及借助意志使认知走向生活的实践活动(李志祥,2020)。人工智能教育应用伦理只有转化为个体的德性、行为,转化为社会的制度,才能发挥作用,实现其意义。其伦理实践具体过程为:首先是教育主体需要“由外而内”形成伦理认知,然后在认知基础上“由内而外”表现出一定的伦理行为,最后是对伦理行为进行规约而形成伦理制度,伦理制度又成为教育主体伦理认知的主要内容。可以说,伦理实践由个体走向社会,又由社会反作用于个体,二者互利共生,相互促进。

### (一) 发轫于教育主体的伦理认知

人工智能教育应用伦理要在实践中落地,融入日常教育教学活动,首先需要教育主体形成伦理认

知,即教育主体要对人工智能教育应用过程中出现的伦理问题、要达到的伦理目标、应该遵循的伦理规范以及涉及的伦理关系有清晰的认识和把握。伦理认知是一个“由外而内”的过程,其中“外”主要指位于教育主体之外的、客观存在的行为规范,这些规范可以是成文的、显性的,如伦理规章等,也可以是不成文的、隐性的,如某些约定俗成的规约。目前人工智能教育应用尚处于初级阶段,相关伦理规范并不健全,教育主体在这一过程中只能以其他教育伦理为参照,这从另一侧面折射了人工智能教育应用伦理研究的必要性和紧迫性。“内”是教育主体将外部行为规范内化的过程,主要包括两个层面:一是教育主体要意识到并了解应该遵循怎样的伦理规范,明晰人工智能教育应用伦理规范的重要性和必要性,尤其是人工智能作为教育系统的新生物,对其伦理规范有清醒的认识非常必要;二是教育主体要从内心接受和认同这些伦理规范,并融入到已有的认知体系中,主动遵循这些伦理规范。

#### (二) 形成于教育主体的伦理行为

人工智能教育应用伦理形成于教育主体应用人工智能进行教育教学活动的过程中,即在这一过程中表现出的伦理行为,是教育主体伦理认知的体现,也是遵循既定行为规范的结果。伦理行为是一个“由内而外”的过程,是内在的伦理认知在理性和意志的协助下,在特定的应用情境中激发出具体行为的过程。人工智能教育应用的伦理行为主要体现在:一是主动性,即教育主体应用人工智能进行教育教学时,能自觉、主动地遵循伦理规范,按伦理制度办事;二是情境性,即在面对特定的伦理困境时,教育主体能够依据伦理规范提出公平、正义的解决方案;三是一致性,即教育主体能够坚持这些伦理规范的合道德性,在面临任何教育情境时都能够始终如一地遵循伦理规范。

#### (三) 完善于社会的教育伦理制度

教育主体在应用人工智能过程中呈现的伦理行为需要伦理制度的规约,尤其是正式的、成文的伦理制度。这一方面便于教育主体省察自己的行为是否合道德;另一方面便于他人对教育主体表现出的伦理行为进行审查,判断其行为是否遵循伦理规范,对不符合的行为给予一定的惩罚。人工智能教育应用伦理制度在现实生活中可以有两种表现方式:一种

是制度伦理化,即指在现有的教育制度中加入人工智能教育应用伦理的相关行为规范,与其他教育制度融为一体。这可以看作是人工智能教育应用初级阶段的表现方式,因为人工智能教育应用伦理相关研究和实践尚不成熟,在原有教育制度的基础上增补行为规范更易被教育主体接受。另一种是伦理制度化,即以制度的形式对人工智能教育应用伦理实践过程进行规约,更具针对性和约束力。这是人工智能教育应用发展到较高阶段的表现形式。但目前,我国人工智能教育应用还未形成系统的、体系化的伦理规范,人工智能教育应用伦理制度化还有很长的路要走。人工智能教育应用伦理制度化,才是人工智能教育应用伦理真正进入实践,成为教育主体日常教育教学活动的准则和规约。

### 四、讨论与思考

人工智能教育应用伦理的诸多研究存在界限不清、主体不明的问题。原因之一在于没有明确人工智能教育应用伦理的研究范畴。

#### (一) “技”与“教”: 两种研究范畴

人工智能教育应用伦理研究的目的是让人与人工智能和谐共处。对于这一研究也因对“人”和“人工智能”当事双方不同的侧重而形成两种不同的研究范畴。一种研究范畴是对人工智能进行伦理规约,属于技术伦理(或工程伦理)范畴,即在人工智能的设计、开发和应用中,以人工智能自身为出发点和落脚点提出伦理要求。另一种研究范畴是对人进行伦理规约,即对人工智能应用领域的主体提出伦理要求。真正让人与人工智能和谐相处、共存共生,仅某一研究范畴是不够的,需要“两条腿”走路,人和人工智能都需要进行伦理规约。但在具体实践和研究中,二者各有侧重。归属为技术伦理的第一种研究范畴应由人工智能的设计者、开发者完成,而第二种研究范畴——人工智能在教育领域的应用应由教育主体完成,因为教师缺乏对人工智能本身进行伦理规约的专业知识和专业能力,只能在运用人工智能过程中对其自身行为等进行伦理规约。

#### (二) 杂糅交错: 人工智能教育应用伦理研究的困境

自欧盟发布《可信赖的人工智能伦理原则》后,人工智能教育应用伦理研究进入新的发展阶段。

2019年5月国际人工智能与教育会议发布的成果文件《北京共识——人工智能与教育》,明确提出要促进教育人工智能应用的公平与包容、促进性别平等、确保教育数据和算法使用合乎伦理、透明且可审核等行动举措(联合国教科文组织 2019)。此后诸多学者就此展开研究。沈苑等(2019)学者基于欧盟提出的伦理框架分析了人工智能教育应用面临的伦理困境及采取的措施,如伦理嵌入设计、包容性测试等技术手段和多层责任分配制度、完善认证系统的非技术手段等;杜静等(2019)结合人工智能在教育领域的应用现状和伦理关键要素,提出人工智能教育应用需遵循问责原则、隐私原则、平等原则、透明原则、不伤害原则、非独立原则、预警原则与稳定原则等;邓国民等(2020)萃取一般的人工智能伦理原则和教育伦理原则的核心交集,提出福祉、是非善恶、公平正义、人权和尊严、自治以及责任和问责等教育人工智能伦理原则。这些研究表明,目前人工智能教育应用伦理研究中技术伦理和教育伦理两个范畴并未各行其是,而是杂糅交错,导致伦理主体不明,在伦理实践层面很难落实,有必要将二者区分开来。教育界学者专注于教育伦理范畴的研究,不仅有利于促进人工智能在教育领域发挥应有的价值,而且也有利于教育领域在人工智能加持下更高效地达到教育目的。

(三)人的主体性:人工智能教育应用伦理研究的着力点

人工智能究其本质仍是一种工具、一类技术,在教育领域的应用不可能改变教育的本质和目的。虽然人工智能作为类人体,具有一定的主体性,但其主体性的发挥以及如何发挥仍取决于教育主体,在伦理维度亦是如此。教育主体在人工智能教育应用过程中有决定权,占主动地位,如何让“人”与“人工智能”和谐共处同样取决于教育主体的合道德行为。

总之,人工智能教育应用的伦理研究明显滞后于人工智能的发展和应用。本文主要从教育主体出发,在教育伦理范畴内对人工智能教育应用伦理在“知善”和“求真”两个层面进行了探析,未来研究还将进一步探索人工智能教育应用伦理目标层面上的“臻美”——培养全面发展的人。期待更多学科的跨界学者能够在这一范畴内,结合教育教学理论,深入系统地探讨教育主体在应用人工智能过程中需要

遵循的伦理规范,为人工智能教育应用伦理制度的建立建言献策。

#### [注释]

①人工智能教育应用是一个多维学科的复杂融合的领域,其主体还包括人工智能教育应用系统的开发者(包括设计)、运营者等,属于第一个研究范畴——技术伦理或工程伦理的研究,本文在此没有涉及。

#### [参考文献]

- [1]北极星. 北极星 AI 助教[EB/OL]. [2020-12-17] <https://bdpc.51polestar.com/index.html>.
- [2]邓国民,李梅(2020). 教育人工智能伦理问题与伦理原则探讨[J]. 电化教育研究, (6): 39-45.
- [3]杜静,黄荣怀等(2019). 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, (7): 21-29.
- [4]顾建军(2020). 浅论教育的双主体性特征[J]. 教育科学, (1): 1-5.
- [5]何怀宏(2018). 人物、人际与人机关系——从伦理角度看人工智能[J]. 探索与争鸣, (7): 27-34.
- [6]卡尔·雅思贝尔斯(1991). 什么是教育[M]. 北京:生活·读书·新知三联书店:4.
- [7]李福华,年浩,张家年(2020). 人工智能教育应用论纲[J]. 现代大学教育, (1): 1-8.
- [8]李志祥(2020). 伦理学研究需要面向现实生活[N]. 光明日报, 12月21日报.
- [9]联合国教科文组织(2019). 北京共识——人工智能与教育[EB/OL]. (2019-08-28) [2020-11-20] <http://edu.cnrcn.cn/list/20190828/W020190828373742676391.pdf>.
- [10]刘悦笛(2019). 人工智能、情感机器与“情智悖论”[J]. 探索与争鸣, (6): 76-88.
- [11]诺丁斯(2011). 学会关心:教育的另一种模式(第2版)[M]. 于天龙,译. 北京:教育科学出版社 2011:6.
- [12]Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). Artificial Intelligence in Education: Challenges and opportunities for sustainable development[R]. [http://www.nied.edu.na/assets/documents/05Policies/NationalCurriculumGuide/Artificial\\_Intelligence\\_\(AI\)\\_challenges\\_and\\_opportunities\\_for\\_sustainable\\_development.pdf](http://www.nied.edu.na/assets/documents/05Policies/NationalCurriculumGuide/Artificial_Intelligence_(AI)_challenges_and_opportunities_for_sustainable_development.pdf).
- [13]Selwyn, N. (2016). Is technology good for education? [M]. Cambridge, UK: Polity Press: 106.
- [14]Sifakis, J. (2020). 世界顶尖科学家论坛之人工智能会议[EB/OL](2020-10-31) [2020-12-24] [http://m.thepaper.cn/rss\\_newsDetail\\_9799618](http://m.thepaper.cn/rss_newsDetail_9799618).
- [15]Tuomi, I. (2018). The impact of artificial intelligence on learning, teaching, and education: policies for the future[R]. European Union: <https://ec.europa.eu/jrc/en/publication/impact-artificial-intelligence-learning-teaching-and-education>: 30
- [16]沈苑,汪琼(2019). 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原则》[J]. 北京大学教育评论 (4): 18-34.

[17] 苏明 陈·巴特尔(2019). 人工智能伦理的多维审视——基于马克思主义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) (11): 223-228.

[18] 唐汉卫(2018). 人工智能时代教育将如何存在[J]. 教育研究 (11): 18-24.

[19] 田丽(2015). 智慧校园环境下校园一卡通建设[J]. 华东师范大学学报(自然科学版) (3): 530-535.

[20] UNESCO(2019). Artificial intelligence in education changes and opportunities for sustainable development [EB/OL]. [2020-06-23] [http://www.nied.edu.na/assets/documents/05Policies/NationalCurriculumGuide/Artificial\\_Intelligence\\_\(AI\)-challenges\\_and\\_opportunities\\_for\\_sustainable\\_development.pdf](http://www.nied.edu.na/assets/documents/05Policies/NationalCurriculumGuide/Artificial_Intelligence_(AI)-challenges_and_opportunities_for_sustainable_development.pdf).

[21] 汪时冲 方海光 张鸽 马涛(2019). 人工智能教育机器人支持下的新型“双师课堂”研究——兼论“人机协同”教学设计与未来展望[J]. 远程教育杂志 (2): 25-32.

[22] 新华网. 习近平向国际人工智能与教育大会致贺信 [EB/

OL]. (2019-05-16) [2020-06-23] [http://www.xinhuanet.com/2019-05/16/c\\_1124502111.htm](http://www.xinhuanet.com/2019-05/16/c_1124502111.htm).

[23] 新浪. 人工智能为什么学会了骂人 [EB/OL]. (2018-05-26) [2020-12-28] <https://tech.sina.com.cn/roll/2018-05-26/doc-ihcaqueu2157317.shtml>.

[24] 于英姿 胡凡刚(2020). 隐忧与消减: 智能技术之于教育的伦理省思[J]. 远程教育杂志 (3): 55-64.

[25] Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators[J]. International Journal of Educational Technology in Higher Education, <https://doi.org/10.1186/s41239-019-0171-0>.

[26] 张坤颖 李晓岩(2019). 大数据环境下的人工智能教育应用[M]. 北京: 学苑出版社: 249-251.

[27] 张攀峰 寻素华 吉丽晓(2014). “智能学伴”在小学游戏化学习社区中的情感交互设计[J]. 中国电化教育 (10): 123-128.

(编辑: 李学书)

## Ethical Research Outline on Artificial Intelligence in Education

LI Xiaoyan<sup>1</sup>, ZHANG Jianian<sup>1</sup> & WANG Dan<sup>2</sup>

(1. Department of Education, Huaibei Normal University, Huaibei 235000, China;

2. Department of Physical Education, Huaibei Normal University, Huaibei 235000, China)

**Abstract:** Artificial intelligence (AI) has become an important variable affecting future social development. Academia and industry favor AI in education, but it also causes people's ethical dimension concerns. This paper mainly clarifies that the ethics of AI in education is the moral relationship between "human" and "artificial intelligence." First of all, the ethical purport of AI in education lies in "support," and what to support and how to support are two key issues in the research. The ethical principles of AI in education are mainly from the dimension of education subject. Education administrators need to follow the principle of power stipulation, which is embodied in the respect principle of data collection, the understanding principle of data processing, and the boundary principle of data utilization. Teachers need to follow the principle of harmonious coexistence, which is embodied in the principles of humanism, moderation, and care. Learners need to follow the principle of shaping themselves, which is embodied in the principles of goodness, honesty, and moderation. Finally, the ethical practice process of AI in education is a process from individual ethical cognition and ethical behavior to the social, ethical system, and then back to individual cognition and ethical behavior. The trajectory is a spiral cycle from the individual to the society and from the society to the individual. But at present, there are two kinds of research categories related to the ethics of AI in education: technology ethics and education ethics, which are mixed and interlaced, which is not conducive to the development of ethical practice. It is hoped that more interdisciplinary scholars will focus on the research of education ethics.

**Key words:** artificial intelligence in education; ethics purport; ethics principle; ethics practice