

人工智能教育应用的算法风险

谭维智^{1 2}

(1. 曲阜师范大学 教育学院, 山东曲阜 273165; 2. 曲阜师范大学 中国教育大数据研究院, 山东曲阜 273165)

【摘要】 教育已逐渐进入人工智能时代。人工智能技术在改变教育的同时,也在给教育带来不可预知的风险。本研究旨在分析人工智能教育应用中使用算法及计算模型对教育进行量化和计算所造成的潜在风险,并分析其产生根源,提出风险管控建议。研究首先采用预测分析法,对算法本身存在问题进行分析,预测算法的简约化、算法的大规模应用、算法黑箱、算法偏见、算法鸿沟以及过度依赖算法给学生、教师学习成长和个性发展带来的风险;然后采用矛盾分析法对比分析教育的模糊性与算法的确定性之间的矛盾、学生发展的个性化与算法公式化之间的矛盾、教师专业能力具身性与算法去技能化之间的矛盾、教育规则的灵活性与算法硬规则之间的矛盾等,探讨风险产生的原因;最后在综合分析的基础上,按照以人为本的原则提出保持教师在人工智能教育应用中的主导地位、以教育学的方式应用和管理人工智能算法、保持算法和计算模型的透明化和可解释性、建立算法教育应用风险评估和风险管理机制等管控算法风险的对策建议。

【关键词】 人工智能; 算法风险; 教育规律; 以人为本

【中图分类号】 G434

【文献标识码】 A

【文章编号】 1007-2179(2019)06-0020-11

当今,教育已逐渐进入人工智能时代,人们正试图以“人工智能+教育”解决工业革命发展形态下难以得到有效解决的各种教育问题。人工智能技术在改变教育的同时,也给教育带来不可预知的风险和伦理问题。人工智能的核心是大数据算法和模型,其教育应用本质上是数据驱动的算法应用。如果我们要用人工智能解决教育问题,必须把教育问题抽象为数学问题,针对相关教育实际问题 and 教育自身的逻辑搭建模型、设计算法、利用数据进行计算。人工智能教育应用的风险主要是使用算法和计算模型对教育现象进行量化和计算造成的风险,风险的发生取决于计算模型和算法是否符合教育逻辑、教育过程和教育中的人是否可以被量化和计算、对教育过程的量化是否反映了教育本真、算法和计

算模型是否安全可靠、算法和计算模型是否存在滥用误用等。

一、潜在风险

当今教育发展的趋势之一,是科技成了驱动教育的主要力量,调节着教育的发展动态。无论出现什么新技术,人们都会尝试将其应用于教育,希望给教育带来特别的“进步”或“创新”。随着“一切皆可计算”的算法崇拜在教育领域的蔓延,教育不断被算法化和编码化,算法日益成为教育权威的代名词。算法在教育中的应用涉及教育教学和管理的方方面面,被用来计算学生的知识存量、学科倾向、思维类型、情感偏好、能力潜质,评估与动态监测学生个体、群体的学业水平,推送学生个性化课程、学习资源、

【收稿日期】2019-09-21

【修回日期】2019-10-24

【DOI 编码】10.13966/j.cnki.kfjyyj.2019.06.003

【基金项目】2017年度山东省社会科学规划新型智库研究专项“‘互联网+’时代教育评价体系研究”(17CZKJ04)。

【作者简介】谭维智,教授,博士生导师,曲阜师范大学教育学院,中国教育大数据研究院常务副院长(tanweizhi@sina.com)。

学习建议和学习策略,分析评价教师教学表现,自我调节、校正和监控教师教育教学活动,决策学校未来发展及日常运作等。在形成新的问题意识、发现新的意义关联的同时,算法所带来的教育风险和伦理问题也越来越受到人们的关注,正如联合国教科文组织(2019a)最近所警示的那样,人工智能在提供教育机会、向学生提供个性化建议、个人数据的集中、责任归属、对工作的影响、数据隐私和数据馈送算法的所有权方面引发了许多伦理问题。人工智能教育应用的算法风险包括算法的简约化、算法的大规模应用、算法黑箱、算法偏见、算法鸿沟以及过度依赖算法给学生、教师学习成长和个性发展带来的潜在风险。

(一) 算法的简约化有导致教育被形式化的风险

算法和计算模型是高度简约化的。模型的本质就是简化。没有模型能囊括现实世界的所有复杂因素或者人类交流上的所有细微差别(凯西·奥尼尔 2018)。形成算法的数学函数也不能包含其计算的一切内容。数学的成功是有代价的,其代价就是把世界用长度、质量、重量、时间等简单概念看待。这样的表达不足以表示丰富多彩的体验,就如同人的身高并非此人本身一样。数学最多只描述了自然的某些过程,但其符号并未包含所有的一切(莫里斯·克莱因 2019)。运用算法分析学习或教育过程将付出丧失部分重要信息的代价。用简化为本质的计算模型和算法对教育对象和教育过程进行计算,所使用的信息多为替代变量,而不是可靠的直接变量,计算模型对教育对象和教育过程的量化和简化使教育失去了丰富的内涵和诸多有价值的成分,包括学生对知识的好奇心和神秘感、对教师的崇拜感等教育过程中的“神秘”元素被消解了。被简约化的教育过程也因为失去了这些“神秘”元素而逐渐走向形式主义,将本来丰富多彩的教育生活和教育过程中的人简约到像电脑行为一样可以预测,其结果就是机器越来越像人,而人越来越像机器,教育活动丧失最根本的人格和人性,教育变得越来越没有灵魂和情感。基于高度简约化算法设计的虚拟仿真、虚拟训练并不真正等同于真实情境的学习,模型不能模拟出教育情境和教育过程的复杂情形,算法无法像教师那样直接回答学生的疑问,更无法像教

师那样以具身的形式把实践操作经验等默会知识教给学生,学生按照算法设定的程序进行操作,使需要智慧融合、思想碰撞的学习变成一种套路固定的“游戏”,有导致学生常识性知识碎片化、缄默性知识缺失、独立思考、逻辑推理、信息加工等高阶思维得不到发展、学科视野狭窄、协作沟通能力和应对复杂问题的能力不足、教育被形式化、浅层化、表面化等一系列风险。

(二) 算法大规模应用有导致学生同质化发展的风险

算法的进化有赖于大规模的数据,每一种算法都包含着大规模、大样本的基因。在人们对算法盲目崇拜的时代,教育正被醉心于量化运动的工业家和商业家接管,他们不遗余力地大规模推行算法的教育应用,使教育更加程式化,比人们所批评的工业化时代“流水线”式教育对学生个性的扼杀有过之而无不及。技术在追求效率的同时总是在谋求扩大受众规模。与传统的班级授课制相比,算法的大规模使用将受众通过技术分流到不同地点、不同时间,表面上看,人工智能实现了教育方式的分众化、一对一、个别化,但当所有的受众背后都面对同一个计算模型和算法时,实际的结果反而是合众化的。算法将更多的人合并在一个更大的虚拟社区或班级,形成一种算法班级、算法“教育流水线”。学习是个性化的,每个学习者的问题表现不同、原因不同,算法如果规模化使用,以一种模式对待所有学生,便会成为伤害学生个性、固化学生思维的“杀伤性武器”。即使是优质的教育算法,一旦大规模使用,也会带来湮灭学生个性的风险。各种个性化推送算法的教育应用本意在追求培养目标和学习方式的个性化,从学习者的思考方式、兴趣爱好、学习特点等方面为学生提供个性化、定制化的学习内容和方式,从知识关联和群体分层层面面向学生推送学习建议和学习策略,力图做到因材施教,而大规模使用反而有导致学生发展同质化、学校教育趋同化的风险。传统教育模式中教育问题与班级规模或学校规模有很大相关,某种程度上可以说班级规模越大,教育问题越多,算法时代同一算法的大规模应用也会带来类似问题。

(三) 算法黑箱有导致教育被算法控制的风险

算法输入的数据与输出的结果之间,存在着我

们无法洞悉的“黑箱”。算法“黑箱”不仅意味着不能观察,还意味着我们无法理解算法和它生产的内容。这一方面源于算法本身的复杂性,即使其设计者也很难清楚解释整个算法的运行过程;另一方面则源于第三次人工智能发展浪潮背景下机器学习算法本身脱离人类表达能力,不仅其基于大数据集的自我学习、自我训练过程不为人所知,甚至最终形成的规则集也往往不能转换为可为人所理解的自然语言(贾开,2019)。算法时代是“技术对人类的理解越来越深刻,而人类却无须理解技术的时代”(卢克·多梅尔,2016:123)”。当将算法应用于教育时,我们很难理解算法处理教育问题的逻辑,算法也很难向我们解释它处理教育问题、看待教育过程的方式。算法按照它自行演化的规则处理教育问题,以其自身的逻辑对待学生、对待教育,学生和教师却毫不知情,教师无法判断他们依据算法做出的决定是否正确,是否符合学生需求,会不会对学生成长造成伤害。这样,无论我们理解与否都将被动接受算法输出的结果。数据不被我们所知,算法实现逻辑不被理解,在不知道它是如何工作的情况下贸然将其应用于教育过程,其中蕴藏着非常大的风险,使用规模越大风险越大。可以说,算法正在重新塑造教育,并成为隐藏在学校幕后的控制者,我们正面临从根本上失去教育自由和教育民主的风险。

(四) 算法偏见有导致教育弊端被放大的风险

再好的算法也不能完美反映客观实在,其中必然存在诸多偏差与偏见。各种计算模型为了让教育过程、教育对象、教师行为便于量化分析,必然对其进行提炼,去除各种无法处理的复杂因素,再赋予看似合理的某些数值,以便将其纳入算法公式中。可计算的东西都被算法装入“黑盒子”,而不可计算的东西则被摒弃,取舍之间算法的偏见就产生了。打开黑盒子,设计者与用户面对的将是一堆可以得出某种答案的主观偏见与程序。而合上之后,它体现的就是客观性——一种无须满足任何更多的条件即可生成“是”与“否”的二元选项的机器(卢克·多梅尔,2016:220)。多种过滤层不断削弱客观且无偏见地理解实在这一幻想(卢恰诺·弗洛里迪,2018:13)。无论我们的意图多么符合科学精神,只要开始量化,就会造成狭隘性,其根源不是量化工具,而是当时认为有量化价值的各种指标(卢克·多梅

尔,2016:83)。教育过程存在大量无法评估或测度、无法将其植入算法的“知、情、意”等缄默知识、暗知识,以及因不能被计算而被去除在计算模型之外的体现教育价值的公平、正义、认同、爱、创造力等。由于计算模型缺失了这些富有价值的信息,不可避免地造成教育内容窄化和教育方法狭隘化等问题。计算模型能体现的只是教育的知识传授部分,仅仅是知识的显性部分,算法的优化目标易被设为提高考试成绩或学会更多显性知识,进一步将显性知识传授和记忆价值放大,使教育方式演变为偏狭的“刷题”式的逻辑计算和识记写作,放大、强化既有教育体系中错误的教育质量观、题海战术、灌输教育等痼疾,把应试教育“精细化”,为教育变革带来不可预知的风险。

(五) 移植开源代码和现成算法有导致教育被异化的风险

人工智能教育应用的代码和算法多是从其他领域移植过来的开源代码和现成算法,这些代码和算法是设计者为解决某些特定任务拟定或设计的,与特定的教育任务、教育过程中使用者的意图并不一致,不能完全吻合教育本身的实际要求。每种算法背后都有其底层逻辑、应用边界和条件,不同行业领域的抽象与量化过程不同,其他行业领域的算法移植到教育会存在专业性和针对性不足等问题。应用于教育的算法应以教育常识、教育知识和教育规律为判断标准和规则,使用开源代码和现成算法解决教育问题,采纳的是为其他领域设计应用程序的共同规则,是按其他领域的逻辑思考、分析、处理教育问题,计算、分析教育中一切可以被计算和不可被计算的人、目标和方法,是用计算机逻辑和其他领域的逻辑及规则取代教育逻辑和教育规则。算法的计算结果归根结底只是一种统计概率(人工智能的主流方法是基于统计或者说数据驱动的方法),而非基于逻辑推理得出结果。由于算法本身不具备推理能力,算法的适用范围注定是有限的,为某一领域开发的算法可能会胜过最高明的人类专家,但将其运用于其他领域可能还不如常人。相较于工业、商业等特定领域,教育领域显得更宽泛,涉及的对象和问题更复杂。那些针对工业、商业领域的应用场景开发的人工智能算法,被用于分析教育过程时,虽然同样是分析人的行为,但因应用场景的变化,需处理的问

题更宽泛,算法可能会得出与实际风马牛不相及的计算结果,且由于训练框架固定、算法限制,使用者难以对算法进行修改、完善和优化。把以物为指向对象的算法应用于人身上,特别是应用于以影响人的发展为旨趣的教育过程中,会颠覆人性和教育的基本预设,使教育异化为工业、商业,使人异化为物、数据或信息,给人的生存和发展带来巨大的风险和伦理问题。

(六) 教师过度依赖算法有导致专业技能丧失的风险

算法是以教师智能助手的面目出现的,它利用收集到的教学数据,代替教师对教育过程发生的事件进行分析判断、做出决定,比如帮助那些不了解学生、做不到个性化教学的教师更精确地推送、有效地实施个性化教学和因材施教。在算法时代,教师正成为教育内容的消费者,而不是教育内容的创造者,教师依靠算法决定传递何种教育内容、采取何种教育方式,其教育行为受算法的调节和限制,可能会被导向固定的路线,就像演员只能照剧本表演一样(卢克·多梅尔 2016: 126)。消费者的角色将使教师失去课堂洞察力和创造力,失去富有魅力的、对学生发挥示范性引领意义的创造者角色。由于算法的不透明和不可解释性,教师对算法背后的理论假设和做出某项具体决定的逻辑一无所知,只是机械式运用,这就使教师的专业素养很难得到提升。教师越依赖各种算法,自身的专业素养就越可能低下。过度依赖算法,将使教师逐渐失去认识学生、了解学生、因材施教的能力,丧失对教育情境、教育问题的独立思考和自主判断能力,就像很多教师过于依赖PPT,离开电脑、多媒体就无法上课一样。依赖算法做判断和做决定、按照“程序化”的方式处理教育问题,必然导致教师教学技能的退化乃至教师的去智慧化、“痴呆化”,离开智能助手就无法处理任何教育问题,最终只能依赖算法从事教育工作。人工智能的教育应用或将从根本上毁掉教师职业,不是人工智能取代教师,而是教师因过度使用人工智能而丧失教育智慧最终自我毁灭。

(七) 算法鸿沟导致教育风险不能及时被化解

计算模型和算法的开发设计是非常复杂的数学问题。算法的教育应用本质上是个教育问题,在算法的设计者和使用者之间、算法技术和教育方法之

间存在巨大的算法“鸿沟”。模型的建立不仅基于数据,也基于我们关注或忽视哪些数据(凯西·奥尼尔 2018: 255)。开发人员的教育观念、教育常识直接决定了他会搜集什么样的数据、关注什么样的问题、对某事成功或失败或符合标准的定义,有什么样的程序员就会开发出什么样的模型和算法。开发人员通过将价值观念嵌入算法的方式彰显自己的价值取向。由于算法设计人员可能缺乏相关的教育本质、教育需求、教育过程、教育方法等背景知识,不了解学生和教师真正的技术需求,因此算法不可避免会被植入各种误解、臆断或错误观念。基于错误的教育认识只会产生错误的数据收集模型,进而收集漏洞百出的数据,这样的算法应用于教育蕴藏着极大的误判风险。计算模型和算法只有不断进行修正才能反映教育实际,及时化解各种风险。开发人员“需要错误反馈来做取证分析,查明系统哪儿出错了,什么变量被误读了,什么数据被忽略了。而系统则在此过程中得以学习并进化(凯西·奥尼尔, 2018: 150)”。使用这些算法的教师由于不掌握衡量应用于教育的人工智能的质量和正当性的技术,无法发现算法本身的问题,即使能发现问题也缺乏完善和更新算法的技能,无法化解算法的风险。算法鸿沟使算法设计者和使用者之间缺乏有效的沟通反馈机制。开发人员接收不到可用于修正的反馈数据就不可能对算法错误进行修正与改善,算法得不到进一步学习的机会,错误就有可能一直延续,算法应用的风险就得不到及时化解,而教师对于这种风险缺乏应有的认识和足够的警惕。

二、人工智能教育应用算法风险产生的根源

人工智能在教育领域的广泛应用给管理者、教师和学生带来了双重挑战:一方面是人工智能的快速发展加快了知识生产,知识出现了指数级增长,需要借助新的智能信息技术提升我们对世界的把握能力、应对知识过剩问题。反思和开发新的教育形式,无疑是新时代最激动人心的挑战之一。有很多机遇,但也要面对错失机遇的严重风险(卢恰诺·弗洛里迪 2018: 27)。另一方面是面对人工智能教育应用所带来的新的教育问题,我们所掌握的知识又严重不足,需要生产新的知识以应对人工智能教育应用带来的风险,机器为我们做决定的底线在哪里?

谁来告诉机器算法什么重要、什么优先？如果人工智能的误判导致事故，该谁承担责任？如果人工智能有禁区，那什么该划入其中（联合国教科文组织，2019b）？为充分应对这些挑战带来的矛盾，我们需要思考：算法能有效反映教育问题吗？教育是可计算的吗？教育是否是适宜使用数学公式的领域？对这些问题的回答有助于我们明确人工智能教育应用的边界和条件，厘清算法风险产生的根源。

（一）教育的模糊性、不确定性与算法的确定性之间的矛盾

教育处理的是人的问题。教育过程和教育中的人充满不确定性，教育的世界是不确定性的世界。人的成长和发展没有明确的路径和清晰的评价标准，什么是教育、教育的规律是什么、教育应如何培养人、学生应如何学习，这些问题永远没有标准答案。如果人的成长有某种确定性规律，教育就可以被打造成流水线、实现标准化培养，学校教育将被算法同化，多样化的学校也没有存在的必要了。学校生活对学生的意义在于不确定性，任何试图改变教育的不确定性，以某种确定性取代不确定性的努力，包括把教育的输入端（学校和教师）和输出端（学生）的任何一端变成确定性的，都要付出无比高昂的代价。

算法处理的是确定性问题。算法的世界是确定性的世界、物化的世界，物的性质无论什么人研究、在哪里研究，发现的结果都是同样的。算法可以计算、自动化处理的都是有精确定义的概念和现象，这样才能保证数据的多样化和完全性，这是算法能够发挥作用的重要条件。数学处理的是物理世界中最简单的概念与现象，它的研究对象不是人而是无生命的物质，它们的行为是重复性的，因而可以被数学描述。但在经济学、政治理论、心理学以及生物学领域，数学就无能为力了。即使在物理王国，数学也只研究简单化的事物，这些简单化的事物与现实的连接就如同曲线的切线仅切曲线于一点一样（莫里斯·克莱因，2019：425）。教育领域是数学无力处理的领域。教育的过程、目标、方法充满模糊性、复杂性，涉及要素复杂且没有终极的、可量化的标准。教育不像法律那样有明确的定义，好奇心、创新性、爱、同情心、同理心等诸多教育范畴都是模糊定义。教育过程看似微不足道，但难以对其加以描述和构造

以满足算法的需要。如果我们要对教育进行计算，就需要某种机制将不确定性和模糊性转化成确定性、精确性，明确定义教育的诸多范畴。而教育需要不确定性和模糊性，教育的模糊性保护了孩子的不同天赋和个性。如果试图对教育进行精确定义，将会把教育变成真正的流水线，培养出一模一样的人，比我们所诟病的工业时代的学校教育更可怕。教师的教育艺术也依赖于模糊性，教师课堂上作出的判断和决策，如对学生知识掌握情况的判断、听讲专注度的分析以及根据课堂情境调整讲授方法乃至声音大小、提出什么问题等都包含着教师的个人化算法，每个教师的算法都是在具体情境下即时生成的，也是复杂无比的。教师的个人化算法本身涉及的教育理念、教育思想、教育哲学以及他的学科知识和教育智慧都是模糊的、缄默的，教师的个人教育算法是模糊算法、随机算法，无法公式化。对教育精确定义、去模糊性的结果必然导致数据的单一化和残缺化。就教育而言，算法的应用范围是有限的，不是所有的教育过程和教育工作都可通过算法完成。

（二）学生发展的个性化与算法公式化之间的矛盾

算法和计算模型本质上是数学公式，其教育应用是对教育进行公式化处理的过程。算法时代，每个人都被抽象为一组数据、一组标签或一组特征。机器学习算法根据学生的学习习惯、学习轨迹、学习偏好、知识掌握程度、长处和弱点等粒度对其进行统计推理和算法分拣，在细分的基础上对学生分类、画像，为每个学生打上不同的“标签”，赋予其公式化的“算法身份”。为了便于算法分拣，学生必须接受分解程序的处理，被拆解成一个个便于分析的成分。他们不再是不可分割的“个体”，而是可以被细分的“分格”……可以利用算法等工具无限细分、最终变成用数据表示的物理形态的人……公式把人变成了“可分解动物”（卢克·多梅尔，2016：43）。被算法打上不同“标签”的学生会接受不同的推送内容以及不同的推送方式和数量。每个学生对其他同学收到的信息可能一无所知。机器学习算法基于社会整体“大数据集”而形成“规则集”并应用于具体场景的过程，暗含着以整体特征推断个体行为的基本逻辑，这可能造成“算法歧视”问题（贾开，2019）。这种通过挖掘数据建立学生个人档案，预测个人行为

的微目标锁定技术不断发展,将形成学生之间的知识区隔,加大学生之间的信息“鸿沟”,造成学生之间的信息沟通、学术交流、合作学习等方面的困难,最终扭曲教育,不仅无法实现人工智能教育应用所承诺的效率、公平,还会将学生的个性蜕变成碎片化的数据标签、算法分格,造成人与人之间的疏离。算法对于人类行为特征的精准识别有利于为不同个体提供差异化、个性化服务的同时,也将强化个人偏好甚至可能催化极端倾向(贾开,2019)。被算法标签化的新算法身份与个人、个性化具有完全不同的含义,是一种公式化的自我、完全被量化的自我,是与身体完全割裂开来的、缺乏自我认同的算法自我。学生对自我的认知依赖于算法的结果,必然造成个人的认知与自我的疏离、对学习的涉身性的背离。

让每个学生按照自己的需求和喜好实现个性化发展一直是人们的美好教育理想,这也被认为是突破工业时代流水线式学校教育磨灭学生个性弊病的方向。但是,我们有必要认识清楚,运用算法计算出来的学生个人“画像”、公式化身份并非真正的个性化,推送算法对学生个性化学习内容的推送也并非真正的个性化学习。真正有利于个性化发展的学习不是每个学生按照自己的喜好学习不同的内容,而是每个学生按照自己的学习风格采取适合自己的学习方式和学习进度学习共同的知识内容,个性化发展的前提是每个学生必须与其他学生共同学习人类的共同知识中精选出来的那部分。如果把人类的大脑看作一个智能系统,我们学习的知识就是各种数据,人类学习的最终目的不是记忆知识(数据),而是通过学习知识保持人之为人的不同于机器的好奇心、想象力、创造力,训练提升个人算法,形成更复杂的大脑神经连接,塑造大脑的发展,以便于更好地识别和使用各种知识(数据)。个人算法不单纯是传统意义上的听、说、读、写、算等能力,也包括更重要的反思能力、思辨能力、思维能力、创新能力、解决复杂问题的能力、沟通协作能力以及宽阔的学科视野,也就是知识运作模式以及产生知识的方式。学生的个性化发展是在学习共同知识的基础上人的全面发展个性化,而不是根据个人的偏好学习个性化知识内容。公式化的个性化知识推送不等同于学生的个性化发展。公式化的个性化推送和量化自我与其说是有利于学生兴趣爱好和天赋的发展毋宁说是

对其个人发展潜能的抑制。

人的个性化发展拒绝被公式化。教育的公式化只会培养出视野狭窄、目光短视、具有偏见的,而不是德智体美劳全面发展的人。教育的目的是使人成为人,成为人意味着牢记生活是连续的、多样的且被体验的,而不是被计算的(卢恰诺·弗洛里迪,2018:33)。算法的有效性需要完整信息和海量优质数据的支撑,只有当数据集足够庞大、具有完整信息时,才能增强算法的适应性和可靠性。学生的个性化发展是其性格特征、好奇心、想象力、创造力、合作能力、理解能力、思维能力等多种因素综合作用的结果,而这种综合作用是无法精确量化、测量的。教育过程的很多方面不能提供关于教育问题的完整信息和足够多的数据,缺乏海量数据资源支撑的算法就像无本之木,会出现失真和错误。我们不能把不可以计算和量化的教育对象强行用人工智能去解决,甚至为了使用人工智能技术强行去采集数据,比如让学生穿智能校服、戴智能手环、在教室安装人脸识别系统、将学生生活中使用的物品装上传感设备,监控他的一言一行。在不需要和不可以使用人工智能解决的教育问题上强行使用人工智能技术和算法,只会导致数据量不足或是数据采集的谬误,都是颠倒了目的和手段,无谓地增加了教育风险和伦理问题。

(三)教师专业能力具身性与算法去技能化之间的矛盾

教师的专业知识和专业能力是具身性知识,教师的具身性知识是教师专业知识和专业能力的综合体,包含教育理念、教育智慧、学科知识、教学技能、个人才能、自发而持久的思维、感觉以及那些模糊的、无法量化的、非显性的知识、能力和教学经验。教师利用具身性知识发现教育过程中出现的各种问题,归纳和反思问题背后的逻辑,创造性地提出解决问题的办法。教师的具身性知识赋予教师教育工作的灵活性和创造性,其思维和知识的宽度和广度是算法无法比拟的。教师以具身性知识缄默地影响学生,教师的专业成长是学生成功的最好路径,理想的学校应该是既能帮助学生成长,又能帮助教师实现职业发展的地方。与教师的专业成长追求知识的具身性相反,算法追求的是知识与人身分离,其处理过程和结果均需要将知识从人身上抽离出来,机器

学习算法基于大规模数据集形成的“规则集”不依赖于人类的表达能力,这样的技术突破不仅意味着人类行为自动化程度的又一次提升,也反映了算法生产过程及其应用结果与人类行为本身的分离(贾开 2019)。在不断复杂化的人机混合系统的背景下,教师处理教育事件关注的主要内容在教育事件本身转移到智能助手上,原本用来支持旧有智力功能和精神追求的神经回路逐渐弱化,且开始分崩离析。大脑会回收那些闲置不用的神经细胞和神经突触,将其用于其他更迫切的工作。我们会获得新的技能和新的视角,可是旧的技能和视角也会因此而丧失(尼古拉斯·卡尔 2015: 151)。教师和智能算法之间的相互作用会系统性地影响教师专业发展的进程,教师所拥有的那些非技术的、独特的教学技能和个人才能,如同情心、社交能力、创造力、沟通能力、推理能力和判断力将被人工智能技术迅速而彻底地重塑,人工智能助手作为教师教育智慧的延伸,将使教师付出丧失教育智慧及各种具身性教育知识的代价。

算法与教师的工作方式根本无法匹配。教师的教育活动、教育行为不需要结果复杂的运算工作。根本上说,教育不同于那些离开人工智能就无法对海量数据进行处理,目前用海量数据训练机器学习的技术在一些领域的广泛应用,很大程度上是因为这些领域存在人脑无法处理的海量信息。教育领域没有被互联网、计算机改变,部分源于这方面的原因,教育技术的热衷者普遍感到大数据、人工智能在教育领域的应用进展缓慢,没有像其他领域那样运用海量数据喂养人工智能。教师在课堂上以及课后的教育和教学工作是极其复杂的,对教师的专业知识进行编码是个非常棘手又浩大的工程,目前的人工智能算法还不能与教师的理解力和洞察力相匹配。如果我们试图对教师的教学进行量化分析,就会将其破坏殆尽。教师的教学表现对于数学模型来说太过复杂,可用来训练模型的数据都是一些粗糙的替代变量。这种模型还没有完善到能够对我们信任的人做出重要评判。教师考核需要考虑各种细微之处和具体情境,即使在大数据时代,这仍然是个很有挑战性的问题(凯西·奥尼尔 2018: 244)。教育过程中教师凭借掌握的具身性知识,无需算法的协助进行分析判断,无需通过语音识别、人

脸识别、情绪识别等人工智能技术就可以对课堂发生的各种情境进行分析判断和处理,教师不需要人工智能抓取学生瞳孔和表情的变化,也不需要算法提供学生专注度、开心值、惊喜值、兴趣值、理解值等数据,就可以分析判断学生的课堂表现、知识点的掌握程度,并做出即时的互动与相应的教学策略。充分相信教师的教育智慧和把控课堂教学的能力,这是我们保持教师教育过程中的主导地位的一个重要原因。

(四) 教育规则的灵活性与算法硬规则之间的矛盾

算法处理的是静态的、可预测的、按确定性规律演化的问题,数学模型的本质是基于过去的数据推测未来,其基本假设是:模式会重复(凯西·奥尼尔 2018: 31-32)。这种对未来的确定性预测是反教育的。教育是动态的、主动求变的,体现在每个学习者身上最大的变化是教育结果的不断变化,教育中人的思维、判断、推理都是随机应变、因时因事不断变化的,教育中的人是变化中的人。人的可变性、可塑性是教育赖以存在的基础,教育存在的意义就是促进人的变化。没有促动受教育者发生变化的教育是不成功的,教育者追求教育对象当下和未来对过去的超越,而非过去与未来的一致。人是如何变、通过什么生变、在什么时候什么情况下变,又是极其复杂、不可预知的。算法通过分析过去的数据来学习,用来计算教育的数据都是“过时”的数据,是教师和学生过去的教学行为中留下的“信息足迹”。这些数据是基于过去的知识点或教育情境形成的,当我们用旧的“信息足迹”形成的算法分析他在新的知识点或教育情境中可能的行为时,就会产生可变性不同步的问题,无法预测当下或未来的变化。另外,量化数据也无法预测质性变化,特别是学习过程中发生的涌现、顿悟等质变,因为这些变化往往是没有规律可循的,不可从过去的数据中求取的规律来进行预测。因此,教育过程的数据往往没有什么太大价值,最起码在改进教育过程、提升教育质量上是这样的。在绝对意义上,根据一个人的数据开发出的模型不可用于另一个人,不能对每名学生的采取同样的算法,适用于今天的模型明天就不一定那么有效了。算法提供的答案不一定是我们想要的。根据大数据计算出来的学生特征和偏好是固定的,学习以

获得新的变化为旨趣,而不是固循学生已有的东西,学习中的求变不能以尊重其偏好而行动。教育算法的规则截然不同商业领域的投其所好式的规则,教育从根本上很难也不能被模型化、程式化。

教育的求变特质决定了教育规则是灵活的、可变通的,这样才能适应教育过程的复杂性、不可预测性,而算法规则往往是简单的、不可变通的、缺乏灵活性的。很多研究者相信通过深度学习,人工智能最终或能表现出想象力和创造力,他们认为“只要把数字输入全世界最大的计算机,统计算法就可以帮助我们找出科学无法发现的规律(卢克·多梅尔 2016: 208)”,正如威廉·J·米切尔(1999)所言,代码就是法律。各种算法和计算模型都力图将复杂的教育过程简化为可以测量的秩序与规则。算法在教育中的大规模使用会在教育过程中逐渐确立为普适性准则,学生在学校所接受的一切教育都将由算法来决定,对教育过程的控制愈细致,甚至会产生一种类似于法律的权威性影响,教师和学生只能被动接受,而极少有力量反驳它、质疑它。套用米切尔的话,算法正演变为教育规则。随着我们在教育过程中越来越依赖人工智能,教育面临着被代码化的风险,教育规则变得越来越格式化,正在由“算法即教育规则”向“教育规则即算法”演进。实质上,即使算法能自行演化出一套“教育规律”或“教育法则”,它也不是具有教育意义的“规律”,只是算法通过概率统计找到的一些在教育过程中重复出现的模式。算法并不能真正理解教育,其有效性完全依赖于数据,也就是过去的经验,算法完全是一种基于经验的判断,是对既往经验的固化和放大,它不能无中生有,这方面完全异于教育的求变特质。算法不具备价值判断和逻辑推理能力,它不会产生经验(数据)中没有的东西。机器学习算法的成功在于将人类积累的经验加以汇聚,从中发现那些成功率高的模式,并反复将其应用于新的场景,而不会考虑它是否适合应用于新的场景、在新场景的应用是否会造成荒唐的结果。反映教育规则的概念大多是抽象概念,如教育公平、教育质量、学习兴趣、创造力、好奇心、洞察力等,这些概念往往是不易量化、不可计算的模糊概念,程序员也不知道如何为公平、兴趣、创造力、好奇心、洞察力编码,这些抽象的教育概念不能被编入模型,模型所计算的仅仅是它能够计算的

那一部分,这就决定了教育规则和教育规律很难被计算、很难用算法表达出来,教育规则与算法规则是性质完全不同的规则。

三、以人为本管控人工智能教育应用的算法风险

联合国教科文组织总干事阿祖莱(Azoulay)说:人工智能是人性的新前沿。人工智能的指导原则不是完全自主或取代人类智能。我们必须确保人工智能是以人性化的方式发展起来的(UNESCO, 2019b)。人工智能教育应用所引起的问题并不完全是技术问题,更多的是关于人类自身的人性问题和伦理问题,我们需要清晰认识算法和模型带来的教育方法和过程的变革及对人性的侵蚀,按照以人为本的原则管理人工智能的教育应用,在算法面前保持人类主权,使算法服务于人性,通过强化人的控制地位、限制算法的应用范围、加强模型审查和监管等途径削弱它们的神话地位和美丽想象,避免算法风险的发生。

(一) 保持教师在人工智能教育应用中的主导地位

降低算法风险的关键是对教育保持必要的管控。算法时代,应该比任何时候都要重视教师在教育中的重要作用。教育需要创造未来,这个任务不能交给算法去完成,大数据程序只能将过去编入代码,而不会创造未来。创造未来需要道德想象力,而想象力只有人类才有。我们必须明确地将更好的价值观嵌入我们的算法代码中,创造符合我们道德准则的大数据模型(凯西·奥尼尔,2018: 238-239)。算法是硬规则,教育需要软规则,不同的情境下,不同学生出现的不同表现需要教师运用教育智慧进行弹性处理。越是算法时代,越是大规模使用模型,越是需要教师的参与,需要教师的智慧,需要保持教师在教育中的能动性和自主性,不能让算法替代人变成教育的“决策者”。算法有其应用的边界和条件,算法可以发挥价值的地方,仅仅是那些存在最优解或唯一答案的地方,通过算法可以直接寻求最优解或唯一答案,而教育问题往往没有最优解或唯一答案,因此更需要发挥人的自主性,运用智慧处理各种情境性问题。教师要深入了解算法世界,通晓算法的本质和局限性,在教育应用过程中驾驭算法、超越

算法、摆脱算法的控制,保持在教育中的控制地位,更多地运用自身的教育智慧处理教育问题,而不是一味地迷信算法、盲目接受算法给出的答案、最终被湮没在算法世界中。教师要掌握新的数字技能以确保不被算法操纵和控制,在使用算法的过程中能够干预算法输出的结果。教育世界是个“多种主观性并存的世界”,教育过程中,教师对学生的影响是“错综复杂、粒度极高且常常十分隐晦的”(卢克·多梅尔 2016: 46),再先进的算法也不能复制真爱、关心、同情、启发,不能取代教师存在的教育价值。要看到人工智能的教育应用给教师专业成长带来的去技能化的趋势,探索人机共生环境下教师专业技能的再生路径,以保障教师在人工智能教育应用过程中的自主性和主导权,不要陷入“效率陷阱”而罔顾人工智能教育应用对教师专业成长的侵害。

(二) 以教育学的方式应用和管理人工智能算法

人工智能教育应用最大的风险,并不在人工智能本身,而在于人工智能背后设计模型和算法的公司和个人对算法的滥用。算法滥用往往是因为算法设计者过于追求经济利益、使用者过度依赖算法、管理者盲目追求效率扩大算法的应用范围造成的。在算法时代,教育不可能完全弃绝算法和模型,我们需要使用相关算法为学校教育教学工作提供有价值的参考意见,发现教育教学中隐藏的教育模式。但是算法的使用者和开发者必须保持清醒,教育问题不是概率统计问题,不是数学问题,也不是计算机问题,解决教育问题最有效的手段不是算法模型,也不是人工智能。教育领域,特别是面对成长中的学生时,不能过分倚重数学分析,尽可能地避免对算法的滥用和误解。我们首先应该停止我们的科技乌托邦幻想,即无根据地寄希望于用算法和科技解决一切问题。要想让算法更好地服务于人类,我们必须承认算法不是全能的(凯西·奥尼尔 2018: 243)。算法的滥用和误解的部分原因是我们在教育概念上对新技术的挑战准备不足,人类心灵通过概念来把握世界:感知必定以概念为中介,它们像是界面,我们通过它们体验和解释实在。概念提供了对周围实在的一种理解,以及领会它们的一种手段。然而,当前的概念工具箱并不适合解决新的与智能信息技术相关的挑战,并且导致了对未来的消极设想:我们恐惧

和拒绝不能理解和不能赋予意义的东西(卢恰诺·弗洛里迪 2018: 9-10)。目前的教育概念工具箱不适合用来解决教育与算法、人工智能等问题,我们通过清晰的概念把握实在,因此应根据实在的变化重构概念范畴、更新概念框架,丰富概念工具箱,提高教育概念的确定性和可解释性,逐渐将更多的教育核心概念融入算法框架,开发新的系统和形式以增强对教育过程和教育中的人的解释力。解决算法滥用问题的根本还是根据教育的实际问题、按照教育的逻辑、教育规则、教育常识和教育内容开发真正属于教育的原始核心模型、代码和框架,真正通过人工智能技术触及到“人的学习”这一教育的根本问题,整合跨越教育学、人工智能等学科的知识和见解,实现知识驱动和数据驱动的融合。教育算法必须是基于教育实际问题和应用场景的“原创者”,以教育理论为底层理论,从教育的底层算法做起,实现模型、算法设计、模拟训练协同优化、随时修改,而不是做其他领域应用场景的“跟随者”。人工智能的教育应用要坚持教育学导向,而不是人工智能算法的技术导向,由我们想要什么样的教育来决定使用什么样的算法,而不是由我们使用的算法来决定形成什么样的教育。要进一步清晰人工智能教育应用的边界和条件,推进人工智能教育应用需要思考这样的本源性问题:人工智能不能做什么。需要考虑教育问题是否需要以及是否可以用人工智能来解决,不要把原本不需要用人工智能解决的问题强行用人工智能去解决。

(三) 保持算法和计算模型的透明化和可解释性

教育是一种解释性、理解性行为,教育过程是将缄默知识、黑箱知识透明化的过程,它要求教育者必须通晓教育的原理、过程、内容和方法,必须能够全盘把握和调控教育对象和教育过程,具有全面的风险管理和控制能力。算法应用于其他领域可以不需要理解它的运算原理和运算过程,而教育需要保持透明化和可解释性,算法的教育应用对教育利益相关者(包括教师、学生、家长和教育管理者)必须是透明的、可理解的、可解释的。同时算法的使用者也需要认清算法和模型的局限以及可能带给教育的种种风险和伦理问题,保持开放的心态,假定提供这种关键的基础架构的计算决策系统,同任何智能系统

一样,必定是有偏见的和可错的(卢恰诺·弗洛里迪 2018: 34)。提高算法透明性和可解释性的重要途径是让教师参与决策过程,包括算法的形成和修改过程。没有一直有效的算法或模型,保持算法或模型有效性的唯一办法是不断修改它,使其不断迭代完善。没有固定的、客观的、唯一的数学体系。进一步来说,如果历史具有某种指导作用,将会不断有新的内容添入数学之中,因而呼唤新的基础。在这方面,数学就像任何一门自然科学。当与先前的理论相抵触的新现象或新的实验结果出现时,就必须修改这些理论,且必须将这些新现象、新结果纳入其中,没有时效的对数学真理的描述是不可能存在的(莫里斯·克莱因 2019: 388)。模型和算法的设计者要主动打破与使用者之间的算法鸿沟,统计学家可以把了解到的任何信息纳入模型对其加以完善。这就是可靠模型的运作方式。可靠模型的开发者会对自己想要理解或者想要预测的所有事情进行反复的核实查证,模型必须随着具体情况的改变而改变(凯西·奥尼尔 2018: 6)。在使用人工智能技术时,不能夸大算法的精准性,不能夸大所谓的精准教学的价值,算法的开发者和使用者必须明确地理解和接受算法的前提假设、知晓算法和计算模型所忽略的因素,明确假设条件和忽略的因素可能带来的负面影响,并及时提出和接受反馈信息,形成积极的信息反馈回路,适时对算法做出校正和改善,以降低各种应用风险。只有当模型是个拥有积极反馈回路的生态系统,才有可能利用数据改善教育(凯西·奥尼尔 2018: 244)。

(四) 建立算法教育应用风险评估和风险管理机制

要确保算法的使用透明且可审查,决策者不仅需要未来的技术发展及其社会影响保持开放态度,而且需要对不能预料的未来做好充分准备(卢恰诺·弗洛里迪 2018: 29)。目前人们对大数据、人工智能等新技术教育应用的社会监管能力的提升速度远远慢于技术革新的速度,要保证算法和模型的良性运转,必须对其进行必要的审查和监管,及时进行风险评估和管理。这些模型是强大的机器,必须看好它们(凯西·奥尼尔 2018: 255)。算法审查首先要对模型的前提假设进行教育本质层面的剖析,要以尊重教育共识和教育常识为前提;其次要对

数据的质量进行管理,看计算模型是否丢弃了重要数据,或者为了模型的效率和经济利益而使用了替代变量,是否将学生分为不同的群体进行区别对待,模型的输入数据和输出结果是否透明、公开,是否能够满足教师和学生访问的需求;最重要的,要看算法系统是否包含算法歧视和滥用问题,避免人工智能对教师和学生造成侵害。算法的设计和开发要基于教育的逻辑、对未成年人教育、教师工作方式和学校教育现实环境的深刻理解,算法和计算模型设计完成后要进行系列的测试,通过包容性、公平性、可靠性、透明性、可解释性、鲁棒性等的检验和风险评估,才可以投入教育实践应用。要组织专业人员对模型的开发、算法的设计、使用环节进行跟踪监控,并有权力随时纠正模型和数据包含的任何错误。算法和模型的开发者应清晰地界定算法的使用范围,明确告知教师和学生算法能够发挥哪些作用、最初的设计目的是什么、模型是如何运作的,要向所有用户解释模型输出的结果,事先告知和报告算法和模型可能产生的负面影响和潜在隐患。要从制度上防范算法的不透明以及大范围应用可能导致算法开发人员权力过大的问题,通过立法划定算法开发者的权力范围,制定算法开发者权力越界的纠错制度和问责制度。要制定人工智能教育应用的道德伦理准则、技术标准,实施教育算法和模型准入制度,没有经过审查的教育模型和教学系统禁止在学校使用。

【参考文献】

- [1] 贾开(2019). 人工智能与算法治理研究[J]. 中国行政管理, (1): 17-22.
- [2] 凯西·奥尼尔(2018). 算法霸权[M]. 马青玲译. 北京: 中信出版社.
- [3] 联合国教科文组织(2019a). 人工智能给教育带来的机遇和挑战[EB/OL]. [2019-03-07]. <https://zh.unesco.org/news/ren-gong-zhi-neng-gei-jiao-yu-dai-lai-ji-yu-he-tiao-zhan>.
- [4] 联合国教科文组织(2019b). 教科文组织举办首次推动人性化人工智能的全球会议[EB/OL]. [2019-02-25]. <https://zh.unesco.org/news/jiao-ke-wen-zu-zhi-ju-ban-shou-ci-tui-dong-ren-xing-hua-ren-gong-zhi-neng-guan-quan-qiu-hui-yi>.
- [5] 卢克·多梅尔(2016). 算法时代[M]. 胡小锐, 钟毅译. 北京: 中信出版社.
- [6] 卢恰诺·弗洛里迪(2018). 在线生活宣言: 超连接时代的人类[M]. 成素梅, 等译. 上海: 上海译文出版社.
- [7] 莫里斯·克莱因(2019). 数学简史: 确定性的消失[M]. 李宏魁译. 北京: 中信出版社.

[8]尼古拉斯·卡尔(2015). 浅薄: 你是互联网的奴隶还是主宰者[M]. 刘纯毅译. 北京: 中信出版社: 151.

(编辑: 李学书)

[9]威廉·J·米切尔(1999). 比特之城: 空间·场所·信息高速公路[M]. 范海燕, 胡泳译. 北京: 生活·读书·新知三联书

The Algorithmic Risk of Artificial Intelligence in Education

TAN Weizhi^{1 2}

(1. Faculty of Education , Qufu Normal University , Qufu , Shandong 273165 , China

2. Chinese Academy of Education Big Data , Qufu Normal University , Qufu , Shandong 273165 , China)

Abstract: Education has gradually entered the era of Artificial Intelligence. While changing education , Artificial Intelligence technology is also bringing unpredictable risks to it. The purpose of this study is to analyze the potential risks caused by using algorithms and computing models to quantify and calculate the education. The aim of study is also to discover the basic causes of the risks in the application process of Artificial Intelligence in education , and to propose the solutions for risk management and control. First of all , predictive analysis is adopted to analyze the problems existing in the algorithm for students' and teachers' learning growth and personality development. These problems include the simplification of the prediction algorithm , large-scale application of the algorithm , algorithm black box , algorithm bias , algorithm gap , and the risks brought by teachers' over-reliance on the algorithm. Then the causes of risk are discussed by using the method of contradiction analysis. The analysis contrasts between the fuzziness of education and the certainty of the algorithm , between the personalization of students' development and the formulation of the algorithm , between the embodiment of teachers' professional ability and the demilitarization of the algorithm , between the flexibility of education rules and the hard rules of the algorithm. Finally , on the basis of comprehensive analysis , according to the people-oriented principle , put forward countermeasures and suggestions to control algorithm risks. These suggestions include maintaining the dominant position of teachers in the application of Artificial Intelligence education , applying and managing Artificial Intelligence algorithms in a pedagogical way , and keeping algorithms and computational models transparent and interpretable. Risk assessment as well as risk management mechanism of algorithm education application should be also established.

Key words: Artificial Intelligence; Algorithmic Risk; Law of Education; People Foremost