

人工智能赋能教育评价：价值、挑战与路径

郑永和 王一岩 杨淑豪

(北京师范大学 科学教育研究院, 北京 100875)

[摘要] 人工智能为教育的创新发展带来了新的驱动力,也催生教育评价的转型升级。人工智能在教育中的深入应用,将推动教育评价由“经验性评价”向“数字化评价”转变、由“单一性评价”向“综合性评价”转变、由“结果性评价”向“过程性评价”转变、由“诊断性评价”向“反馈性评价”转变。同时,它也会带来一系列问题,如技术至上遮蔽教育评价本质、标准缺失误导教育评价方向、算法偏见影响教育评价效度、反馈缺失削弱教育评价价值、人机互信危机阻碍教育评价实施、技术僭越加剧教育评价风险。智能教育评价设计者需要厘清本质内涵、完善指标体系、优化算法设计、健全反馈机制、强化人机互信、完善伦理规约,以提高教育评价的系统性、科学性和客观性,为教育创新发展提供有效依托。

[关键词] 人工智能;教育评价;核心价值;现实挑战;实践路径

[中图分类号] G434

[文献标识码] A

[文章编号] 1007-2179(2024)04-0004-07

一、引言

2020年,中共中央、国务院印发《深化新时代教育评价改革总体方案》,明确提出“改进结果评价、强化过程评价、探索增值评价、健全综合评价,充分利用信息技术提高教育评价科学性、专业性、客观性”(教育部,2020)。教育评价是教育改革发展的“指挥棒”,影响着人才培养目标的实现和教育发展方向,也决定了教育的标准、底线和原则。近年来,随着人工智能、大数据、生成式人工智能技术的逐渐渗透,教育评价逐渐朝着智能化、科学化、精准化的方向发展,以便有效辅助教育诊断、挖掘教育问题、支撑教育决策,为教育创新发展提供有效依托。从本质上讲,人工智能赋能教育评价

应以新时代教育的高质量发展为目标,以教育规律为参照,以智能技术为依托,以数据为核心驱动力,构建面向多元主体的系统化、科学化、综合化评价体系,驱动教育评价效率和效度大幅提升,为人才选拔方式的改进、教育目标的设计、教学策略的优化、教育政策的制定提供参考(郑永和,等,2023)。其核心逻辑是从长期教育实践经验出发构建系统化的评价指标体系,利用人工智能技术实现多源异构教育数据的采集分析,对评价对象的外显状态和内隐特征进行全方位、时序化的建模分析,并依据预设的评价体系和算法逻辑生成评价结果,在此基础上挖掘潜在的教育问题,并生成定制化的教育改进方案,促进教育实践不断优化。然而,人工智能技术在为教育评价带来诸多“利好”的同时,也不

[收稿日期] 2024-07-10

[修回日期] 2024-07-11

[DOI编码] 10.13966/j.cnki.kfjyyj.2024.04.001

[基金项目] 2022年度新一代人工智能国家科技重大专项“面向智慧教育的学习者认知与情感计算研究”子课题一“面向学习成效评价的认知与情感计算模型”(2022ZD0117101)。

[作者简介] 郑永和,教授,博士生导师,北京师范大学科学教育研究院院长,研究方向:教育信息科学与技术、科技与教育政策、科学教育(zhengyonghe@bnu.edu.cn);王一岩(通讯作者),博士,北京师范大学科学教育研究院,研究方向:智能教育、人机协同教育、教育信息科学与技术(wangyiyang3318@163.com);杨淑豪,硕士研究生,北京师范大学科学教育研究院,研究方向:人机协同教学。

[引用信息] 郑永和,王一岩,杨淑豪(2024). 人工智能赋能教育评价:价值、挑战与路径[J]. 开放教育研究,30(4):4-10.

可避免地带来系列风险和挑战,限制智能教育评价的有序推进。本文旨在厘清人工智能赋能教育评价的核心价值、现实挑战和实践路径,以期为智能教育评价的设计和 implement 提供借鉴。

二、核心价值

人工智能技术的赋能为教育评价的高质量发展提供了有效依托,推动教育评价全面转型,为智能时代教育的创新发展提供了方向指引。

(一)由“经验性评价”向“数字化评价”转变

以往的教育评价大多依赖教师的经验,由教师利用教学经验评判学生的学习能力、学习态度和性格品质,或利用“经验主义”+“专家征询”的方式构建系统化评价体系,再通过问卷、量表、测试题、课堂观察等方式对学生能力进行评价。这带来的问题是,教育评价的设计完全由经验驱动,评价实施也具有较强的主观色彩,评价结果往往取决于教师的认知水平、过往经验和个人喜好,甚至会受教师当下情绪状态的影响。如果教师具有丰富的教学经验和深厚的专业素养,往往能够作出公正、精准的评价,反之,则很难为学生提供客观评判和有益指导,且教师精力有限,难以全面深入地了解每位学生。人工智能技术的应用能有效解决经验性评价带来的主观和低效问题,机器能够对学生多样化学习场景的学业表现数据进行时序性、细粒度的采集分析,再利用机器学习和深度学习算法对学生的表现进行智能评判,助力经验性评价向数字化评价转变,推动教育评价朝智能、精准、客观的方向发展(吴砥,等,2023),为教育评价效率的提升和效能的优化提供保障。

(二)由“单一性评价”向“综合性评价”转变

从单一和综合的视角看,以往的教育评价受教师精力和数据采集方式的限制,往往聚焦于某一细分领域,利用单一证据表征复杂的评价对象,如利用考试数据评判学生的学习能力和教师教学水平,利用课堂测验数据评价课堂教学质量,利用升学数据评判区域教育发展水平等。教育系统的复杂性使得这种依赖单一证据的评价方式难以准确表征评价对象的本质特征,也无法为教育实践的改进提供有效建议。在某种程度上,人工智能技术能有效拓展数据的获取渠道,覆盖更加多元的学习场景、学习主体和数据模态,推动教育评价从单一性评价

向综合性评价转变。例如,对学生来讲,人工智能技术能够实现知识性评价向素养性评价转变(徐瑾劼等,2023),不仅能精准评判学生学业表现,还能对学生的知识基础、解题技巧、学科能力、情感态度、思维品质等进行综合分析,深入解析学业表现的成因,全方位刻画学生的问题解决能力、批判性思维能力、跨学科能力、创造力等高阶思维能力,挖掘学生个性化特征。对于课堂教学评价来讲,人工智能技术能依据学生的课堂表现、课后测试、参与度以及教师的行为、话语、表情等数据有效还原课堂教学过程,深入揭示教师行为和学生学习行为之间的复杂关联关系,实现对课堂教学的精准分析。

(三)由“结果性评价”向“过程性评价”转变

人工智能技术的应用使教育评价从只关注结果转变为关注过程。以往的教育评价受数据采集方式和分析方法的限制,大多采取课后测试或者让学生事后填写问卷和反思日志的方式,其优点是简单高效,能对评价对象的发展状况进行整体评判。缺点表现在:一是证据来源单一,仅仅依据最终的结果,无法对评价对象的发展过程进行动态的建模分析,也难以科学解释最终的评价结果;二是无法及时发现学生学习和成长中的关键问题,并进行干预,造成学业问题累加。人工智能技术的应用能帮助教师强化对教育过程的关注,实现对教师的教、学生的学的时序化、动态化建模分析,丰富教育评价的证据来源,及时发现其中的关键问题,如通过学生话语、表情、动作、课堂测验等多模态数据,测评分析学生的专注度、知识掌握程度、情绪态度等,这一方面能帮助教师科学评判学生的学业表现,另一方面有助于教师深入探究影响课堂教学质量的关键问题,并强化过程性干预,动态调整教学方案。从这一意义上说,人工智能技术的应用能实现面向真实教育过程的实时、动态、精准评价,推动结果性评价向过程性评价转变,为教育评价的有效实施提供可靠保障。

(四)由“诊断性评价”向“反馈性评价”转变

传统的教育评价多以高利害考试为手段,评价目标多定位于选拔学生或者评判学生能力发展水平(如中考、高考),但对评价结果的合理有效利用常处于缺位状态。除选拔和甄选的功能外,现阶段教育评价的目标还要发挥导向、鉴定、诊断、调控和改进

作用(教育部,2020),使其结果能够为教育实践的改进和优化提供有效反馈,避免为评价而评价。但传统的教育评价受评价方法的限制,评价结果仅能呈现特定群体的发展水平。人工智能技术的应用极大地丰富了教育评价的技术手段,使得评价者不仅能够对评价对象进行全方位、过程性的综合评判,还能够对评价结果的成因进行合理分析,提出针对性改进方案。例如,要评价学生学业,教师可以利用知识图谱、认知诊断、情感计算等方式对学生的知识掌握、认知能力、情感态度等特征进行综合分析(王一岩等,2023a),并基于此为学生提供定制化的学习干预措施,帮助学生克服学业困难,提高学业水平。由此可见,人工智能技术的赋能能够推动教育评价从诊断性评价向反馈性评价转变,构建基于“诊断—反馈”的闭环模型,真正发挥教育评价对教育实践的改进作用。

三、现实挑战

(一)技术至上遮蔽教育评价本质

随着教育数字化转型的逐步推进,智能技术之于教育改革的赋能作用越发凸显,大众对智能技术的期待也到了前所未有的高度。在很多论述中,人工智能技术俨然成了数据化、科学化、个性化、智能化的代名词,似乎教育的高质量发展必须依托智能技术来实现(李芒等,2024)。这种观点很大程度上将智能教育评价推向了“技术决定论”的实践误区,似乎只有运用智能技术才能开展有效的教育评价。这种观点过于强调技术的应用,却忽视了教育评价的本质,使得融合人工智能技术的智能教育评价系统沦为“装点门面”的工具,忽视了其对教育改革的赋能作用。具体表现在:一是盲目迷信智能评价系统的分析结果,忽视教师实践经验,认为机器通过数据化建模分析得到的评价结果是科学和客观的,忽略机器在很大程度上是在模拟和延伸人的智能。智能教育评价的核心目标是让机器能够在人类经验的驱使下完成教育评价的中间环节,而不是让机器替代教师;二是极力倡导在教育评价中运用智能技术,却未考虑智能技术的适用性,往往只倾向于将教育过程进行数据化表征,但对数据所揭示的教育现象和教育问题缺乏合理设计,使智能技术的应用脱离真实的教育场景,无法为教育评

价提供有效支撑。从本质上讲,智能教育评价应该以教师的实践经验为指导,合理设计智能技术的应用场景,开发相应的智能评价系统,目标是让机器替代教师完成繁琐、单调的评价任务,以提高教育评价效率,而不能过于强调技术应用,忽略教育评价的本质。

(二)标准缺失误导教育评价方向

评价体系是连接智能技术和教育评价的桥梁。科学合理的评价体系一方面能承载教育评价的本质要求,将复杂的评价内容拆解为具体的、可落地的评价指标,为教育评价的实施提供标准参照;另一方面能支撑数据化建模分析的实际需要,建立教育数据和评价指标之间的映射关系,利用微观数据表征复杂的评价对象。目前智能教育评价往往缺乏科学合理的指标体系作为支撑,限制了智能教育评价的实施。比如,学生评价应关注学生的哪些认知和情感状态、哪些学科能力、哪些综合素质?课堂教学质量评价应关注学生和教师的哪些行为?如何构建数据指标和复杂特征之间的关联关系,逐层揭示评价对象的深层次特征?这些问题的解决均需要系统的指标体系和评价规则作为支撑。但目前智能技术研发人员重视对数据本身的刻画,却难以表征评价对象的复杂特征,教育研究人员过多强调从宏观的培养目标和教育需求设计评价体系,导致指标体系与教育数据脱节,二者难以建立有效的映射关系。由此可见,如何构建既能因应教育发展和人才培养本质特征,又能够支撑细粒度数据建模分析实践需求的系统化指标体系,是未来人工智能赋能教育评价需要解决的核心问题。

(三)算法偏见限制教育评价效度

人工智能赋能教育评价的基本假设是,智能技术能科学、合理、有效地表征教育过程、评价教育对象、挖掘教育问题,并为学生、教师和管理者提供客观精准的评价结果。但在实践中,由于数据类别和算法模型的限制以及开发人员的“偏见”,机器决策往往会失去客观中立立场,产生对特定群体的不公平对待,进而加剧教育评价的不可控风险。其一,算法偏见影响教育评价的公正性。如机器对特定地域、性别、民族、文化背景或某种特定学习习惯的学生存在“偏见”,那么就有可能导致这些学生的评分失准,受到不公平对待。其二,算法偏

见影响学生发展的多样性。如果机器通过数据化的建模分析判定学生喜欢某特定类型的教学资源,那么在评价该学生学习风格和学习偏好时,自然会打上相应的“标签”。这种标签在机器智能分析与决策的过程中会不断地被强化,使机器倾向于为该学生推荐特定类型的学习资源。长此以往,这些学生的学习偏好会被不断固化,影响学生的多样化发展。其三,算法偏见造成教育机会的不均等。算法偏见可能导致机器无法准确评估学生的潜力,过早地将学生分流,导致学生失去参加高阶课程或特定项目实践的机会,引发教育机会不均等。教育评价作为教育改革的指挥棒,决定着教育改革的方向,而算法偏见使得智能教育评价的科学性和客观性大打折扣,阻碍教育改革的深入推进。

(四)反馈缺失削弱教育评价价值

教育评价的根本目标是以科学的标准对教育活动进行客观评定,以促进教育实践的改进和教育决策的优化。在以往的教育评价中,人们容易将教育评价等同于选拔性考试,突出教育评价鉴别和选拔人才的使命。人工智能技术为教育评价的设计与实施提供了“新动能”,能有效推动教育评价从诊断性评价向反馈性评价转变,更加关注评价结果对教育实践的改进作用。然而,由于评价标准的缺失、算法模型的限制和实践逻辑的不完善,教育评价还仅仅聚焦于评价本身,对评价结果的有效应用缺乏足够关注。比如,常见的智能评价系统仅聚焦于评价课堂教学质量,无法根据课堂教学实际对教师教学能力进行诊断和反馈,也无法为课堂教学的优化提供合理建议。其结果是研究机构和企业花费大量人力物力开发智能评价系统,仅能够满足评价的需求,却无法为学习、教学和管理的改进和优化提供有效支撑。因此,未来人工智能赋能的教育评价,需要重视教育评价对教育实践的反馈作用,通过构建“评价—反馈”的闭环模型,挖掘教育评价背后潜在的教育问题和教育规律,为相关教育主体提供智能精准的改进建议,推动智能教育评价效果的提升。

(五)人机互信危机阻碍教育评价实施

从目前来看,机器无法完全替代人类进行教育评价。因此,智能教育评价更多的是由教师或者研究人员制定评价标准和规则,由机器负责数据分析

并得出评价结果,最终由教师、学生和管理者依据评价结果进行教育决策。但机器的“测不准”特性以及机器智能评价的不可解释性,导致教育主体对机器评价结果的“不信任”,从而会引发严重的人机互信危机,阻碍智能教育评价的有效实施。其中的问题包括:一是机器能采集的数据维度有限,难以满足全景化数据建模分析的实际需求,机器提供的智能评价仅仅是基于有限证据的分析结果,与现实可能存在出入;二是由于算法偏见,机器的分析结果可能会误导教育决策,加剧教育不公、影响教育质量、阻碍学生发展;三是依托人工智能算法的智能评价与决策往往局限于呈现评价的结果,忽视得出结果的原因,使机器的智能分析与决策具有较强的不可解释性,为教师和学生采纳评价结果带来阻碍。如果人类难以信任并采纳机器的分析结果,将会给智能教育评价的开展带来严重危机。在此背景下,如何提升机器智能评价的科学性、提高师生的智能素养、建立人机互信机制,是未来人工智能赋能教育评价需要关注并解决的核心问题。

(六)技术僭越加剧教育评价风险

机器智能化使其在教育中的作用越发凸显,在一定程度上替代学生和教师进行评价,这也将导致教育评价中人机关系的异化,引发严重的伦理危机。其一,造成教育评价主体的让渡。以往的教育评价多由教师主导,人工智能技术水平的发展,使得机器能够替代教师完成大部分评价工作,实现评价的智能化、自动化。在此背景下,如若教师将复杂繁琐的教育评价工作完全交由机器完成,那么将导致教育评价主体的让渡,由机器决定教师该教什么和怎么教,教师则成了机器评价结果的被动接受者,其教育主导作用将逐步丧失(苏慧丽等,2024)。其二,引发评价主体的技术依赖。随着智能技术对教育评价的渗透,机器在教育评价中发挥着越来越重要的作用,师生会越来越依赖智能评价系统开展教育评价,使机器成为教育评价的权威,长此以往,师生的感知、判断、推理、决策、反思能力将严重退化(吴龙凯等,2023),导致评价主体思维的惰化(余胜泉等,2023)。其三,导致人机责任归属不明。从目前看,由于评价标准不健全、反馈机制缺失,加之算法偏见的存在,智能教育评价的科学性和客观性大打折扣。如果依据机器的评价结果进行教育决策,很可能会

给教育发展带来负面影响,从而引发机器智能评价的责任归属问题。智能教育评价所产生的后果究竟该由机器承担还是由教师和学生承担,是未来智能教育评价需要解决的关键问题。

四、实践路径

从整体看,人工智能赋能教育评价的机遇和风险并存,如何发挥智能教育评价的正向效用并规避其潜在风险,是智能教育研究需要关注和解决的核心问题。未来,人工智能赋能教育评价需要厘清本质内涵、构建指标体系、优化算法设计、健全反馈机制、强化人机互信、完善伦理规约,推动智能教育评价创新发展。

(一)厘清本质内涵,明确智能教育评价核心价值

教育评价是教育改革的指挥棒和方向盘。智能技术的演进,将为“教”和“学”带来颠覆性影响,但对教育评价的影响十分有限。根本原因在于,教育评价反映了人才的培养目标和教育质量的判别标准,其设计与实施需要参考教育改革与实践需求,推动教与学的转型升级。从这一意义上说,智能教育评价应该从教育评价的逻辑审视智能技术应用,而不该过度夸大技术的作用,盲目迷信智能教育评价系统的分析结果(王一岩等,2023b)。智能技术对教育评价的赋能主要体现在通过数据感知和算法分析,优化教育评价的中间环节,并为教育决策提供参考。为防止盲目迷信机器评价结果导致机器对教育主体的“僭越”,教育者需要平衡“人评”和“机评”的关系,明确智能教育评价旨在帮助教师、学生和管理者完成繁琐重复的数据采集和分析工作,为评价主体评判教育现状、优化教育决策提供参考,而不能成为教育评价的唯一标准。学生、教师、管理者等也需明确机器智能评价的核心价值及其局限性,加强对机器评价结果的审视和评判,正确利用智能教育评价的核心效用,并规避其潜在风险。

(二)构建指标体系,规范智能教育评价实践方向

评价体系构建是教育人工智能研究的基础性工作(宋宇等,2023)。科学合理的评价体系有助于将人类能够感知和理解的完整世界转化为机器能够建模和分析的具体对象,帮助机器更好地理解客观事物的存在状态。智能教育评价需要以科学合理的评价

指标体系为指引,而指标体系的构建需要以教育经验为参照,将复杂的评价内容逐层分解为机器可感知、可分析、可解释的细粒度指标体系,为智能教育评价的实施提供方向指引。因此,智能教育评价指标体系的构建,一是要完善智能教育评价的实践逻辑,将复杂的评价内容逐层分解为层次化的指标体系,并构建指标体系和外显数据表征之间的映射关系,在此基础上,通过数据的感知和分类以及指标之间关联关系的复杂建模,揭示评价对象的深层次特征;二是要聚焦智能教育评价的场域特异性,从人类主观经验和机器建模分析的双重视角出发,构建层次清晰、结构合理的智能教育评价指标体系,为教育数据的采集、评价结果的生成、潜在问题的挖掘、改进方案的设计提供标准参照;三是要提升指标体系对教育场景的适用性,构建能够反映评价对象本质特征、表征教育过程整体样态的指标体系,实现对评价对象的深入分析和精准建模。

(三)优化算法设计,推动智能教育评价科学发展

算法偏见严重制约了智能教育评价的科学性和客观性,也可能引发严重的人机互信危机,为智能教育评价的实施带来极大风险(王佑镁等,2023b)。因此,未来智能教育评价需要从数据采集和算法设计层面消除可能存在的偏见,确保人工智能赋能教育评价的公正性和合理性:一是丰富数据来源,提高数据质量。智能教育评价系统应加强对不同地区、不同性别、不同背景学生的关照,从多个角度采集数据,避免依赖单一来源或单一类型的数据,同时,通过数据清洗提高数据质量,降低数据中存在的偏差。二是优化算法设计,提高分析效能。算法设计应遵循公平、透明、可解释的原则,消除偏见对模型性能的影响,确保算法在决策过程中保持中立。同时,算法设计者要提高算法的透明度和可解释性,让利益相关者能够了解算法背后的逻辑和依据(卢宇等,2022),提高对智能教育评价系统的接受度。三是多元主体参与,强化监管机制。评价者应广泛征集学生、家长、教师和管理者的意见,了解不同主体的需求,及时调整算法参数和决策逻辑,确保评价结果的科学性和准确性。

(四)健全反馈机制,提高智能教育评价实践效用

反馈机制是发挥教育评价“指挥棒”作用的关键保障。教育评价的目标不在于评价本身,而在

于通过对评价结果的深层次分析,准确评估评价对象的发展水平,并生成有益于教育实践的改进措施。因此,建立健全智能教育评价的反馈机制对提升教育评价效能具有重要意义。教育评价的设计和指标体系的制定要充分考虑评价结果如何作用于教育实践的改进和优化,构建“评价设计—评价实施—评价反馈”的闭环模型,为智能教育评价提供指导思想。教育评价的实施既要关注评价内容本身,又要探究能够对评价对象产生影响的核心要素(如评价学生的学科能力要充分考虑学习动机、学习态度等影响因素),以此深入挖掘评价对象的关键问题所在。反馈环节要充分利用智能技术手段对评价对象的关键问题进行深入分析,并生成定制化的解决方案(如为学生提供个性化的学习干预,为教师生成定制化的教学方案等),为教育实践的优化提供方向指引。同时,评价者应加强对反馈效果的验证,探究机器生成的反馈方案能否为教育实践的优化提供正向反馈,帮助评价设计者优化反馈方案,提升教育评价的实践效用。

(五)强化人机互信,推动智能教育评价有效实施

人机互信危机是智能教育评价有效实施的关键阻碍。如何提升教育主体对于智能教育评价的接受度、强化人机互信,是智能教育评价需要解决的关键问题。而此问题的解决,一方面需要完善机器智能评价的实践逻辑,确保智能教育评价的科学性和客观性,另一方面要提升师生的智能素养,实现对智能教育评价的合理利用。其中,完善实践逻辑的措施包括:一是筑牢智能教育评价的数据基石,通过多源异构数据的采集分析,对评价对象进行全时空、多维度的数据化建模分析,全方位表征评价对象的核心特征;二是破解智能教育评价的算法偏见,消除智能教育评价中的不公平、不科学、不合理等问题;三是提高机器智能评价的可解释性,不仅要评价的结果呈现给师生,还要清楚阐述评价的逻辑和依据,以及如何利用评价结果对教育实践进行改进和优化(刘桐等,2022)。提升教师素养的举措包括:一要提高师生的人工智能意识和技能,使其了解人工智能赋能教育评价的核心价值,善于利用智能教育评价系统开展教育评价;二要提高师生的风险意识,帮助他们理性辨识人工智能赋能教育评价的潜在风险,并合理规避,避免过度依赖智

能教育评价系统引发自身思维的退化(王佑镁等,2023a);三要对智能教育评价的结果进行理性分析和合理利用,即充分发挥人类经验和机器智能的核心优势,实现人机协同的精准评价(宛平等,2024)。

(六)完善伦理规约,保障智能教育评价良性发展

人工智能技术将使得教育评价的系统性、科学性和可解释性得到极大提升,为教育高质量发展奠定坚实基础。但人工智能的教育应用始终是一把“双刃剑”,在带来诸多利好的同时,也不可避免地引发一系列伦理风险。因此,未来智能教育评价的设计与实施需划定人工智能赋能教育评价的伦理边界,避免因智能技术的误用和滥用引发伦理危机:一是尊重学生隐私,保障数据安全。这就需要构建完善的数据安全防控机制,划定教育数据采集、共享、分析、应用的范畴和边界,避免数据的不合理利用造成学生隐私泄露。二是厘清人机关系,避免过度依赖。学生、教师和管理者要明确自身在智能教育评价中的主体地位,确保智能教育评价的开展能够合理、合规、有效地解决评价中的关键问题,避免对智能教育评价系统的过度依赖引发机器对教育主体的僭越(周洪宇等,2023)。三是构建监管机制,强化责任归属。评价设计者要完善智能教育评价的监管机制,整合人类经验和机器智能的优势,实现人机协同的监管与反馈(王一岩等,2024),全方位监管数据来源、算法质量、评价的效果和改进措施,并对其中的潜在问题进行深入挖掘和系统分析,保证智能教育评价的实施符合法律法规和道德规范。

[参考文献]

- [1] 李芒,余露瑶(2024).“前车之鉴”对教育数字化的启迪[J].电化教育研究,45(4):5-10+17.
- [2] 刘桐,顾小清(2022).走向可解释性:打开教育中人工智能的“黑盒”[J].中国电化教育,(5):82-90.
- [3] 卢宇,章志,王德亮,陈鹏鹤,余胜泉(2022).可解释人工智能在教育中的应用模式研究[J].中国电化教育,(8):9-15+23.
- [4] 宋宇,许昌良,朱佳,柴少明(2023).面向思维培养:基于精准标注技术的智能化课堂教学分析及应用[J].华东师范大学学报(教育科学版),41(8):79-89.
- [5] 苏慧丽,张敬威(2024).机器的“人化”与人的“机器化”:智能时代教育的主体性危机与破解[J].现代远程教育研究,36(1):12-20+28.
- [6] 宛平,顾小清(2024).生成式人工智能支持的人机协同评价:实践模式与解释案例[J].现代远距离教育,(2):33-41.
- [7] 王一岩,刘淇,郑永和(2024).人机协同学习:实践逻辑与典型

模式[J]. 开放教育研究, 30(1): 65-72.

[8] 王一岩, 郑永和(2023a). 可解释学习者建模: 价值意蕴与应用图景[J]. 现代远程教育研究, 35(5): 96-103.

[9] 王一岩, 郑永和(2023b). 智能时代个性化学习的现实困境、意蕴重构与模型构建[J]. 电化教育研究, 44(3): 28-35.

[10] 王佑镁, 王旦, 梁炜怡, 柳晨晨(2023a). “阿拉丁神灯”还是“潘多拉魔盒”: ChatGPT 教育应用的潜能与风险[J]. 现代远程教育研究, 35(2): 48-56.

[11] 王佑镁, 王旦, 王海洁, 柳晨晨(2023b). 算法公平: 教育人工智能算法偏见的逻辑与治理[J]. 开放教育研究, 29(5): 37-46.

[12] 吴砥, 郭庆, 吴龙凯, 程浩(2023). 智能技术赋能教育评价改革[J]. 开放教育研究, 29(4): 4-10.

[13] 吴龙凯, 程浩, 张珊, 宋琰玉(2023). 智能技术赋能教育评价的时代内涵、伦理困境及对策研究[J]. 电化教育研究, 44(9): 19-25.

[14] 徐瑾劼, 申昕(2023). 重塑以学习者为中心的教育评价生态——基于教育评价智能化发展的全球观察[J]. 开放教育研究, 29(3): 40-46.

[15] 余胜泉, 汪凡淙(2023). 人工智能教育应用的认知外包陷阱及其跨越[J]. 电化教育研究, 44(12): 5-13.

[16] 郑永和, 王一岩(2023). 科技赋能教育高质量发展: 价值内涵、表征样态与推进策略[J]. 中国电化教育, (1): 118-126.

[17] 教育部(2020). 中共中央国务院印发《深化新时代教育评价改革总体方案》[EB/OL]. [2020-10-13]. https://www.gov.cn/gongbao/content/2020/content_5554488.htm.

[18] 周洪宇, 李宇阳(2023). ChatGPT 对教育生态的冲击及应对策略[J]. 新疆师范大学学报(哲学社会科学版), 44(4): 102-112.

(编辑: 李学书)

The Value, Challenges, and Paths of Artificial Intelligence Empowering Education Evaluation

ZHENG Yonghe, WANG Yiyan & YANG Shuhao

(Research Institute of Science Education, Beijing Normal University, Beijing 100875, China)

Abstract: *Artificial intelligence becomes a new drive to the innovative development of education, transforming and upgrading educational evaluation. In the era of intelligence, its application in educational evaluation transforms education evaluation from the empirical to digital evaluation, from the singular to comprehensive evaluation, from the outcome to process evaluation, and from the diagnostic to feedback evaluation. At the same time, a series of problems emerge such as: "technology first" obscuring the essence of education evaluation, "lack of standards" misleading the direction of educational evaluation, "algorithmic bias" limiting the validity of educational evaluation, "lack of feedback" weakening the value of educational evaluation, "human-machine trust" hindering the implementation of educational evaluation, and "technology violation" exacerbating the risks of educational evaluation. The future intelligent educational evaluation needs to clarify its essence, improve its indicator system and algorithm design, improve its feedback mechanisms, strengthen human-machine trust, and enhance ethical norms, in order to improve the systematic, scientific, and objective nature of educational evaluation, which provide effective and crucial support for the innovative development of education.*

Key words: *artificial intelligence; educational evaluation; core values; realistic challenges; practical path*