

# 智能技术赋能教育评价改革

吴砥<sup>1</sup> 郭庆<sup>2</sup> 吴龙凯<sup>2</sup> 程浩<sup>2</sup>

(1. 教育部教育信息化战略研究基地(华中), 湖北武汉 430079; 2. 华中师范大学人工智能教育学部, 湖北武汉 430079)

**[摘要]** 以通用人工智能大模型为代表的智能技术与教育评价的融合, 已成为数字时代教育评价改革的重大课题。本文回顾了我国教育评价改革发展历程, 分析了智能技术应用于教育领域的主要阶段和特征, 阐释了智能技术重塑教育评价改革的核心关注点, 包括更加丰富的教育评价内容, 个性化、综合性、伴随式的评价方式, 多元化的教育评价主体, 数据驱动的评价手段等。文章在研判评价改革未来趋势的基础上, 指出智能技术自身的发展进步及其在评价中日益广泛深入的应用, 将助推结果评价更加科学、过程评价更加智慧、增值评价更加便捷、综合评价更加完善。最后, 文章给出智能技术赋能教育评价改革的实践路径建议, 即协同推进评价改革, 突破关键评价技术, 试点探索新评价范式, 注重技术伦理与隐私。

**[关键词]** 人工智能; 大模型技术; 教育评价改革; 综合评价

**[中图分类号]** G521 **[文献标识码]** A **[文章编号]** 1007-2179(2023)04-0004-07

当前, 教育评价改革广受关注。传统教育评价的目的、内容、方法、手段、主体等难以满足教育改革需求(谈松华, 2017)。习近平总书记指出, 要深化新时代教育评价改革, 构建多元主体参与、符合中国实际、具有世界水平的教育评价体系(中华人民共和国教育部, 2023)。以通用人工智能大模型等为代表的智能技术不断被应用于教育领域, 以提供个性化、泛在化的教育服务, 缔造智慧教育新形态(杨宗凯等, 2022), 引发人们对教育评价的新思考。如何顺应智能时代教育改革趋势, 将通用人工智能大模型等技术与教育评价深度融合, 已成为数字时代教育评价改革的契机与趋势。2020年, 中共中央、国务院印发的《深化新时代教育评价改革总体方案》(简称

《总体方案》)明确提出, 发挥教育评价指挥棒作用, 改进结果评价、强化过程评价、探索增值评价、健全综合评价(中华人民共和国教育部, 2020)。本研究旨在回顾我国教育评价改革的发展历程, 探析以通用人工智能大模型为代表的智能技术重塑教育评价改革的发展重点与推进路径, 明确其应对策略, 以促进教育评价改革高质量、可持续发展。

## 一、时代背景

### (一) 教育评价改革走向数字化

我国教育评价改革始于20世纪80年代中后期(涂端午, 2020)。1985年, 中共中央发布《关于教育体制改革的决定》, 明确教育体制改革根本目

**[收稿日期]** 2023-07-21 **[修回日期]** 2023-07-24 **[DOI 编码]** 10.13966/j.cnki.kfjyyj.2023.04.001

**[基金项目]** 国家社会科学基金“十四五”规划2022年度教育学重点课题“智能技术赋能教育评价改革研究”(ACA220026)。

**[作者简介]** 吴砥, 教授, 博士生导师, 教育部教育信息化战略研究基地(华中), 研究方向: 教育数字化政策、数字素养评价(wudi@mail.ccnu.edu.cn); 郭庆, 博士研究生, 华中师范大学人工智能教育学部, 研究方向: 学生数字素养评价、智能教育评价; 吴龙凯(通讯作者), 教授, 博士生导师, 华中师范大学人工智能教育学部, 研究方向: 教育人工智能、智能教育评价、数字化政策与实践等(longkaiwu@ccnu.edu.cn); 程浩, 硕士研究生, 华中师范大学人工智能教育学部, 研究方向: 智能教育评价。

**[引用信息]** 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.

的在于提高民族素质,指出教育管理部门要组织教育界、知识界和用人单位定期评估高校办学水平(何东昌,1998)。1993年,中共中央、国务院发布的《中国教育改革和发展纲要》指出,建立各级各类教育的质量评估标准和评价指标体系,将评估学校教育质量作为一项经常性任务(何东昌,1998)。这些政策文件的颁布,为我国教育评价体系的发展打下了坚实基础。

进入21世纪,我国教育评价工作快速发展。2001年,国务院印发《关于基础教育改革与发展的决定》,提出改革考试评价和招生选拔制度,加强对学生能力和素质的考查,探索综合评价的考试(中华人民共和国教育部,2001)。2002年,教育部发布《关于积极推进中小学评价与考试制度改革的通知》,提出建立促进学生发展、教师职业道德和专业水平提高、学校教育质量提升的评价体系,将形成性评价与终结性评价相结合(中华人民共和国教育部,2002)。2004年教育部成立高等教育教学评估部门,开展高校教学水平专项评估工作(黄连金,2007)。新世纪以来,我国教育评价改革主要围绕素质教育展开,社会机构开始参与教育评价活动,教育评价体系逐渐形成和完善。

经过十年积累,我国教育开启迈向现代化发展的步伐,教育评价走上创新发展的道路。进入21世纪第二个十年,教育部先后颁布多份教育评价文件(2015),对现行教育评价的方式、内容、主体、手段等提出改进意见,并提出充分利用现代信息技术,建立和完善教育质量综合评价数字化管理平台(中华人民共和国教育部,2013),搭建互联互通的信息共享平台,扩大评价效力。然而,彼时的信息技术与教育的融合发展尚处于起步阶段,人工智能、大数据等新技术也尚未成熟,信息技术、智能技术对教育评价的革命性影响未得到充分彰显。

“十三五”规划实施以来,我国教育信息化事业得到跨越式发展,取得显著成效(吴砥等,2022)。图像识别、语音识别、智能扫描、大数据等智能技术逐步成熟,对教育评价影响深远。2017年,教育部发布《国家教育事业发展的第十三个五年规划》,提出积极促进信息技术与教育的融合创新发展,建立科学评价体系(中华人民共和国教育部,2017)。2020年推出的《总体方案》明确指出要充分利用人

工智能、大数据等智能技术,提高教育评价科学性、专业性、客观性(中华人民共和国教育部,2020)。近年来,随着以ChatGPT为代表的通用人工智能大模型技术的发展,我国更加重视教育评价的改革创新。以智能技术深化教育评价体系变革,成为牵引育人方式、办学模式、管理体制、保障机制等领域综合改革的关键环节(中华人民共和国教育部,2022)。

当前,智能技术尤其是通用人工智能大模型技术,已经成为数字时代教育评价改革的关键途径。推进通用人工智能大模型技术支持的教育评价改革,是建立具有中国特色和世界水平的教育评价体系,推动我国教育内涵式发展,服务教育强国建设的重要时代课题。

## (二)通用人工智能大模型技术应用特点

智能技术与教育的融合发展大致经历了三个阶段(吴砥等,2023)。首先是机器学习阶段。它强调算法的通用化,实现能存会算的数据自动演算。基于数据和人工提取的特征构建概率统计模型,并运用模型预测与分析数据。机器学习主要通过采集问卷、日志等数据,构建学习者画像、分析学习风格、预测学业表现等,使教育评价在一定程度上实现自动化。其次是深度学习阶段。它强调架构的通用化,实现了能看会听的数据智能感知。深度学习技术超越人工设计特征,通过设计适用于多种应用的通用神经网络架构,具有自动提取高维语义特征的能力。深度学习通过采集文本、图片、音视频等多模态数据,实现学习情感分析、教学行为分析、智能阅卷、个性化推荐等教育应用,可提高评价的自动化程度。第三是以ChatGPT为代表的通用人工智能大模型技术阶段。它强调模型通用化,具有能理解会创作的数据认知。通用人工智能大模型不再针对每个领域训练不同的模型,而是训练适用于所有领域的通用大模型,具备上下文学习和理解能力以及一定的创作和知识生产能力,通过采集文本、图片、音视频等多模态数据,实现课件创作、教案编辑、程序编码、智能出题、智能助教、智能学伴等教育应用。

通用人工智能大模型等技术在教育中的应用逐步从专用走向通用,即不仅适用于窄领域、单学科、局部学段,还可基于通用智能底座利用少量数据进行适配,满足宽口径、跨学科、全学段的多样

性教育需求。通用人工智能大模型还将促使教育应用流程从分发走向生成。现有的学习资源推送一般按“诊断—定制—分发”的流程运作。这虽然一定程度上解决了个性化资源推送问题,但资源需要人工制作和汇集,难以满足学习者的多样性需求。通用人工智能大模型能够将预制学习资源分发转变为实时加工和生成个性化学习资源,真正实现个性化服务,达成大规模教育与个性化培养的统一。此外,通用人工智能大模型具备内容生产、人机交互能力,将加速教育元宇宙和虚拟空间的构建、打通真实世界与虚拟世界的连接,促进教育场景从单一走向多元。

## 二、重要关注点

以通用人工智能大模型为代表的智能技术将促进教育评价的系统性变革,其影响主要体现在评价内容、评价方式、评价主体、评价手段四个方面。

### (一)创新教育评价内容

通用人工智能大模型的发展改变了学习者获取知识的方式与途径。传统以知识传授、经验学习为主的教育体系难以满足智能时代的人才需求,创新意识、数字素养、协作意识、批判思维等高阶能力变得尤为重要。通用人工智能大模型技术赋能的教育评价,可采集、还原并储存真实教育情境的精准数据,追踪评价对象的全过程成长轨迹,实现评价内容的创新。第一,评价范围更加拓展。通用人工智能大模型促使教育应用场景从单一走向多元,教育教学不再局限于单一的时间和地点,评价内容也逐渐趋向于多元学习情境下的价值判断,除传统教室场景外,还包括线上学习、室外教学等;除学校场景外,也关注家庭场景和社会场景。通用人工智能大模型还可以对多元学习场景的教学过程进行全面记录,进而精准评价和有效解释教学现象。第二,评价内容更加丰富。通用人工智能大模型能够突破以单一标准尺度评价学生总体水平的缺陷,评价内容由以往的学业水平判断转向德、智、体、美、劳“五育”并举,以及学习成效、学习态度、情感价值观、批判性思维等综合性内容的评价(张琪, 2021)。第三,评价结果可视化。通用人工智能大模型结合相关数据分析技术,可以快速、自动生成个体和群体数字画像,促使数字画像更加科

学、完善,同时以海量样本数据为基础,综合比对评价对象的起点水平、个性特征、环境影响等因素,揭示评价结果背后蕴含的潜在影响机制,从而使教育评价更加科学、客观、高效。

### (二)革新教育评价方式

当前,唯分数论导向的评价已不能满足数字时代的人才培养需求,教育评价方式亟需进行全面革新,实现教育全过程的个性化、综合性、伴随式评价。第一,实现个性化评价。传统以考试为主的教育评价往往以一刀切的方式衡量学生的学业表现,忽略学生的个体差异。通用人工智能大模型可根据评价对象的认知特点、能力发展水平和学习需求,提供定制化的评价和反馈。通用人工智能大模型具备实时处理和分析大量数据的能力,可实时评估学生的学习进度和表现,有效实现个性化评价,将评价结果与评价者的个体特征和学习目标相匹配,支持学生个性化发展。第二,促进综合性评价。传统的教育评价方式主要聚焦单一的学科知识掌握情况或工作绩效,忽略评价内容的综合性。通用人工智能大模型支持的评价方式将实现更加综合的评价。通用人工智能大模型可用于构建多元化的评价指标,通过采集教学过程中丰富的行为、学业表现、生理行为、心理状态等多模态数据,利用智能化的数据处理与分析实现对评价对象综合素质的有效评价。第三,开展伴随性评价。传统的教育评价方式由于评价结果反馈不及时,具有滞后性。通用人工智能大模型可以在教学过程中根据实时采集的多模态过程性数据即时开展教与学的诊断和监测,及时向教师、学生、管理者反馈,帮助其开展高效、个性化的教学。

### (三)增值教育评价主体权益

从我国教育评价改革政策可以看出,教育评价参与者包括教师、学生、家长、管理者等多个主体。尽管政策层面强调并认可不同主体的价值与作用,但传统的教育评价多以政府为主体,容易出现教育行政权力越位、缺位等现象(鹿星南, 2023)。评价结果多作为“教育问责”的主要依据(张进良, 2023),学校片面追求升学率、唯分数等问题还未得到根本解决。在通用人工智能大模型等智能技术的推动下,教育评价主体权益将得到极大丰富。一方面,海量的实时性强、真实度高、内容丰富的教育评价

数据得以全面汇聚,可满足社会、家长、教师等不同利益相关者对教育数据的需求,激活不同评价主体对信息的知情权、发言权与监管权。另一方面,通用人工智能大模型可提供有效的教学分析与评价结果,帮助教师设计、生成更具个性化和针对性的评价任务,促进其反思改进教学设计,提高教学水平与职业发展水平;提供家校沟通平台,帮助家长实时跟进学生成长表现,并依据评价结果与智能技术的指导,提供个性化学习辅导与培养计划,更好地支持学生的学习与发展;鼓励学生参与评价和反思,学生可根据自身学习需求与兴趣获得个性化评价和学习建议,提高自主学习能力与元认知能力。此外,对学校和社会而言,智能技术可减少评价过程的主观偏见与不公,提供全面、精准、客观的评价结果,推动学校和社会管理的优化升级。

#### (四) 变革教育评价手段

单纯关注分数的考试不利于数字时代创新人才的培养,借助信息技术实施的上机考试、电子档案等教育评价手段也难以满足当前的教育评价需求。在通用人工智能大模型的支撑下,教育评价手段亟需全面变革,使评价实践真正摆脱唯分数论的囹圄。一方面,教育评价手段将从以结果性数据为主要依据转向结果性数据与过程性数据相结合。传统的试卷、问卷等评价手段只能获得以分数为表征的结果性数据,但以此为依据衡量学生发展、教师教学、学校工作等难免失之片面。智能技术支持的过程性数据能够被无损、伴随性采集。这些过程性数据与结果性数据相结合,将形成教育主体在教与学过程中完整的行为、认知、情感等数据集,以便开展更加全面细致、科学精准的评价。另一方面,教育评价将从以试卷、问卷为主要手段,转向以大数据采集分析系统为主要手段。智能技术的教学应用将使得人机协同、交互的数据被全方位、跨时空地采集,这些数据统一汇聚存储于大数据采集分析系统。同时,通用人工智能大模型可用于详细分析采集到的数据特征,生成合适的数据处理与分析方案,得到精准的评价结果,从而将教育主体从评价中解放出来。

### 三、支撑与赋能四类评价改革

随着我国教育评价的发展,评价类型已不再是

单纯的结果评价,《总体方案》明确指出了四种评价类型,即结果评价、过程评价、增值评价、综合评价。在通用人工智能大模型技术的支持下,这四种评价类型也将得到全面赋能,使结果评价更加科学、过程评价更加智慧、增值评价更加便捷、综合评价更加完善。

#### (一) 促进更加科学的结果评价

传统的结果评价主要采取标准化测验或量化指标计分等方式。这种评价得出的结果不能科学反映评价对象的真实发展水平。一方面,传统的结果评价大多以强制性方式实施,容易引发评价对象的焦虑情绪,造成评价结果失真。另一方面,尽管标准化测验、评价量表等工具的开发主要借助经典测量理论、项目反应理论等心理测量理论,一定程度上能解释评价对象的发展现状,但这种结果评价大多是一刀切的,忽视了不同对象的个性特点。通用人工智能大模型支持的结果评价将更加科学。首先,通用人工智能大模型与大数据技术、虚拟现实等技术结合,可用于构建立体、沉浸、智能化的评价环境,最大程度地激励评价对象积极参与评价活动。许多研究证实,虚拟现实、游戏化等开放性环境开展的结果评价可获得更加科学有效的评价结果。其次,通用人工智能大模型可以根据评价对象的发展现状进行科学命题与智能组卷,自动生成个性化的评价项目。同时,根据评价对象的作答过程和结果,通用人工智能大模型可以指导下一轮命题,提高结果评价的连续性和可追溯性,使结果评价真正促进评价对象的发展。此外,通用智能技术还将提高结果评价的客观性。机器存储有数以亿计的样本数据,评价主体可根据海量数据比对和常模参照分析,给出更加科学的评价结论。例如,通用人工智能大模型可以准确批改主观题(如作文),还可采取实时问答等方式赋能主观评价(如通过对话评价学生的论证能力与批判性思维)。

#### (二) 赋能更加智慧的过程评价

过程评价是教育评价的发展性功能与价值的重要体现,指在教育教学中了解教学的动态效果,及时反馈和调节教学活动,不断完善教学计划,以达到预期教学目标。传统的过程性评价主要采取质性分析方法,如建立学习档案袋,通过人为制定的标准评价教与学的过程性表现,存在主观性强、

实施难度大等问题。通用人工智能大模型可以实现更加客观、精准、智能的过程评价,促使过程评价的普及应用。一方面,通用人工智能大模型赋能的过程评价依托智慧教学平台、智能传感、可穿戴设备等技术工具,伴随性地采集教学过程中多元主体的多模态行为、情感、认知数据,借助多模态数据分析技术对采集的原始数据与汇聚的数据进行实时处理、融合、分析,实现追踪师生教与学的全过程智能化评价(刘邦奇等, 2022a)。另一方面,通用人工智能大模型从分发走向生成的技术特点意味着教师可以借助其在极短时间内完成自动出题、自动组卷、自动批阅,实现即时过程性评价,无需从题库中抽题,大大减轻提前储备和维护题库的负担,也可避免题目重复导致的评价失准等问题。总之,通用人工智能大模型的应用将大幅降低过程性评价的难度,过程性评价将被更多的教育主体所认可、重视和采用,成为更加普及的评价范式。

### (三)探索更加便捷的增值评价

增值评价指以增长水平而不是以横断面的分数作为衡量学生素养发展、教师教学质量、学校办学水平等的依据。增值评价主要关注学生的成长和进步,而不是绝对的成绩水平,由此对教师和学校工作的评价也将基于学生的学业增值。采用这一评价方式,有助于提高教育评价的公平性,激发生源质量差的学校提高教学质量的动力。当前的增值评价还存在诸多挑战,包括传统数据存储与管理方法难以支撑增值评价对数据的稳定性需求,评价模型种类繁多、专业性强,不易使用,评价结果较为抽象(刘邦奇等, 2022b)。在通用人工智能大模型的支撑下,落实增值评价将变得更加便捷。第一,通用人工智能大模型可以建立增值评价追踪数据库,优化数据的收集、存储、分析与处理等流程,提高追踪数据存储的稳定性与可靠性。第二,利用通用人工智能大模型生成性的优势,可根据教育主体的实际情况和需求选择适宜的增值评价模型,同时进行相应的调整与运算,提高增值评价的易用性和科学性。第三,借助通用人工智能大模型可以便捷快速地对不同目的的增值评价的结果进行汇总和分析,形成有指导性意义的评价结论。另外,计算机图形学、图像处理等可视化技术可以对增值评价结果进行交互式可视化呈现,帮助教育主体直

观、形象地理解增值评价结果。

### (四)落实更加完善的综合评价

综合评价是一种对知识掌握、思维发展等智力因素以及情感、态度、价值观等非智力因素进行完整而系统的价值判断。随着我国教育的不断深化,综合评价的重要性凸显。然而,综合评价实践还存在难以解决的难题,包括评价指标体系的科学性有待提升,评价数据的维度模式需要拓展,复杂评价数据的处理技术亟待突破,综合评价结果的客观性亟需确保(刘邦奇等, 2022c)。通用人工智能大模型将为综合评价的有效实施带来机会。第一,通用人工智能大模型可以在科学分析评价数据的基础上提取评价规则,辅助评价指标的筛选和权重计算,不断迭代优化综合评价指标体系。第二,通用人工智能大模型与大数据技术结合可以针对综合评价指标体系的要求,广泛采集和汇聚多场景、全时空、多模态的教学过程数据,为科学开展综合评价奠定数据基础。第三,通用人工智能大模型可以为复杂的多模态数据处理与分析提供可行的思路、方案,实现文字、图像、音频、视频等多模态数据的智能分析,促进对学生情感态度、语言表达、合作能力、创新思维等内容的全面而精准的评价(杨宗凯, 2022)。第四,通用人工智能大模型支持的综合评价具有多元化评价主体,不仅仅包括教师、学生、家长等,还包括一个超级虚拟主体。它可以借助海量的数据样本进行比对,使得评价不会因主体单一导致评价结果的主观、片面。此外,通用人工智能大模型结合区块链等技术可以确保综合评价数据的安全存储和不可篡改,实现综合评价数据应用的透明和公平。

## 四、思考与建议

### (一)强化顶层设计,协同推进评价改革

智能技术支持的教育评价改革是当前教育数字化建设的重要任务。国家已出台智能技术支持评价改革的战略规划。然而,绝大多数地方和基层教育管理机构对国家评价改革相关战略规划的落实是碎片化、不系统的。此外,随着越来越多的企业涉足教育产业,数字教育产品的种类和型号愈发繁杂,数据规模越来越大。然而,不同厂家的设备缺乏统一标准,设备之间相对封闭难以直接通信,造成教育数据之间的流转、共享困难,使得智能技

术支持教育评价改革难以落实。各级教育管理部门应深入领会智能技术的内涵和功能,协同推进教育评价改革。一方面,国家层面应进一步完善智能技术促进教育评价改革的顶层设计。各级各类教育管理机构要结合自身发展实际以及上级规划,制定针对性的教育评价改革战略行动。另一方面,各级教育管理机构或企业可搭建标准化的教育信息和数据共享平台或系统,开发不同系统产品之间的信息交互接口以及不同教育组织之间信息交换授权接口,实现数据采集与存储的统一规划和管理。

### (二)完善数字底座,突破关键评价技术

在大力发展教育数字化的背景下,依托教育新基建,充分运用通用人工智能大模型、大数据、区块链等关键技术,可以构建支撑教育评价改革的“数字底座”。这一方面需要打造教育评价数据大脑,建立教育基本数据库,消除各级各类部门间的数据壁垒,探索高效的教育评价途径,建立智能化、科学化、全方位的教育评价系统;另一方面,需要突破关键技术,实现教育评价的创新发展,如多模态数据融合的知识追踪技术、交互式数字画像技术、自适应评价反馈技术等。此外,通用人工智能大模型与教育元宇宙等关键技术结合,可用于打造中文通用语境和各类专业语境的智能基座,构建公共算力平台,提高算力在评价中的利用率与精度。全场景、多模态教育数据的综合分析结果可在结果分析、决策制定、政策优化等方面为教育主体提供精准、科学的决策支撑。

### (三)加强试点示范,探索新型评价范式

自《总体方案》发布以来,许多学校都在寻求开展智能技术支持的教育评价改革,但大多是“摸着石头过河”,没有形成完善、可推广的经验。通用人工智能大模型技术支持的教育评价不仅简单、实用,还可结合教育实际需求,实现信息、知识、数据与资源的共享,缩减数字鸿沟,促进教育公平。国家和地方教育主管部门可通过制定相关规划,鼓励并支持有条件的地区开展智能技术赋能教育评价改革试点工作,并加强研究教育评价方法与技术规范,确保评价的科学性与公平性。试点地区教育主管部门应联合高校专家,组建评价改革团队,指导并监管试点区的工作,选取区域内合适的学校、学段,采取先易后难、从局部到整体的推进思路开

展多轮试点探索,不断迭代优化评价范式。此外,试点地区应鼓励相关企业积极投入研发资源,开发通用人工智能大模型支持的智能化评价工具、系统和平台,为评价改革提供强有力的技术支撑。

### (四)注重技术伦理,保护评价对象隐私

通用人工智能大模型支持的教育评价改革在监管、研发、使用等环节的伦理问题不容忽视(吴砥等,2023)。监管方面应加强意识形态引导,加快出台相关法律法规解决数据安全风险,培养高水平数据管理人员,有效评估数据质量、及时性、客观性,改善教育数据生命周期管理。研发方面应针对技术伦理与教育评价融合的模式、方法、路径、影响等展开探索,在评价工具和系统研发过程中注重数据匿名化、安全存储、附加使用权限等,以确保数据的安全存储、传输和使用,保护评价对象的个人数据隐私;严格审查并监管涉及敏感数据和个人隐私的评价工具和系统。应用方面要引导教育管理者有意识地使用通用智能技术,总结通用人工智能大模型在评价改革中的技术优势和潜在的伦理问题,保证评价标准、算法设计、评价过程、结果解读的公开透明,避免有失公允的歧视性评价。此外,尊重评价对象的知情权和选择权,确保评价对象知悉评价目的、方法和后果,保护评价对象的合法权益,也是评价过程需注意的事项。

### [参考文献]

- [1] 何东昌(1998). 中华人民共和国重要教育文献(1976-1990)[G]. 海口:海南出版社:2288+3471.
- [2] 黄连金(2007). 略论中国高等教育评估的发展历程和发展方向[J]. 黑龙江教育(高教研究与评估), 729(9): 82-83.
- [3] 刘邦奇, 喻彦琨, 袁婷婷(2022a). 智能技术赋能过程评价:目标、路径与典型场景[J]. 现代教育技术, 32(5): 14-23.
- [4] 刘邦奇, 朱广裘, 张金霞(2022b). 智能技术支持的增值评价模式及典型实践[J]. 中国远程教育, 575(12): 49-57.
- [5] 刘邦奇, 刘碧莹, 胡健, 袁婷婷(2022c). 智能技术赋能新时代综合评价:挑战、路径、场景及技术应用[J]. 中国考试, (6): 6-15.
- [6] 鹿星南, 高雪薇(2023). 人工智能赋能教育评价改革:发展态势、风险检视与消解对策[J]. 中国教育学报, 358(2): 48-54.
- [7] 谈松华(2017). 关于教育评价制度改革的几点思考[J]. 中国教育学报, (4): 7-11.
- [8] 涂端午(2020). 教育评价改革的政策推进、问题与建议——政策文本与实践的“对话”[J]. 复旦教育论坛, 18(2): 79-85.
- [9] 吴砥, 李环, 陈旭(2023). 人工智能通用大模型教育应用影响探析[J]. 开放教育研究, 29(2): 19-25+45.
- [10] 吴砥, 李环, 尉小荣(2022). 教育数字化转型:国际背景、发

展需求与推进路径[J]. 中国远程教育, 570(7): 21-27+58+79.

[11] 杨宗凯(2022). 元宇宙推动教育的全面数字化转型[J]. 教育研究, 43(12): 23-26.

[12] 杨宗凯, 王俊, 吴砥, 王美倩(2022). 发展智能教育学 推动教育可持续发展[J]. 电化教育研究, 43(12): 5-10+17.

[13] 张进良, 杨苗, 谈桂芬(2023). 智能技术赋能基础教育评价改革的实际困境与路径选择[J]. 中国远程教育, 43(2): 18-27.

[14] 张琪, 王丹(2021). 智能时代教育评价的意蕴、作用点与实现路径[J]. 中国远程教育, 553(2): 9-16+76.

[15] 中华人民共和国教育部(2001). 国务院关于基础教育改革和发展的决定[EB/OL]. [2001-05-29]. [http://www.moe.gov.cn/jyb\\_xxgk/moe\\_1777/moe\\_1778/201412/t20141217\\_181775.html](http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/201412/t20141217_181775.html).

[16] 中华人民共和国教育部(2002). 教育部关于积极推进中小学评价与考试制度改革的通知[EB/OL]. [2002-12-18]. [http://www.moe.gov.cn/srcsite/A26/s7054/200212/t20021218\\_78509.html](http://www.moe.gov.cn/srcsite/A26/s7054/200212/t20021218_78509.html).

[17] 中华人民共和国教育部(2013). 教育部关于推进中小学教育质量综合评价改革的意见[EB/OL]. [2013-06-08]. [http://www.moe.gov.cn/srcsite/A06/s3321/201306/t20130608\\_153185.html](http://www.moe.gov.cn/srcsite/A06/s3321/201306/t20130608_153185.html).

[18] 中华人民共和国教育部(2015). 教育部关于深入推进教育管办评分离促进政府职能转变的若干意见[EB/OL]. [2015-05-06]. [http://www.moe.gov.cn/srcsite/A02/s7049/201505/t20150506\\_189460.html](http://www.moe.gov.cn/srcsite/A02/s7049/201505/t20150506_189460.html).

[19] 中华人民共和国教育部(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]. [http://www.moe.gov.cn/jyb\\_xxgk/moe\\_1777/moe\\_1778/202010/t20201013\\_494381.html](http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202010/t20201013_494381.html).

[20] 中华人民共和国教育部(2022). 深入学习贯彻党的二十大精神 加快推进教育评价改革落实落地[EB/OL]. [2022-11-08]. [http://www.moe.gov.cn/jyb\\_xwfb/gzdt\\_gzdt/moe\\_1485/202211/t20221118\\_995844.html](http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/moe_1485/202211/t20221118_995844.html).

[21] 中华人民共和国教育部(2023). 习近平在中共中央政治局第五次集体学习时强调 加快建设教育强国 为中华民族伟大复兴提供有力支撑[EB/OL]. [2023-05-29]. [http://www.moe.gov.cn/jyb\\_xwfb/s6052/moe\\_838/202305/t20230529\\_1061907.html](http://www.moe.gov.cn/jyb_xwfb/s6052/moe_838/202305/t20230529_1061907.html).

[22] 中华人民共和国中央人民政府(2017). 国务院关于印发国家教育事业发展规划“十三五”规划的通知[EB/OL]. [2017-01-29]. [https://www.gov.cn/zhengce/content/2017-01/19/content\\_5161341.html](https://www.gov.cn/zhengce/content/2017-01/19/content_5161341.html).

(编辑: 李学书)

## Artificial General Intelligence to Reshape the Reform of Educational Evaluation

WU Di<sup>1</sup>, GUO Qing<sup>2</sup>, WU Longkai<sup>2</sup> & CHENG Hao<sup>2</sup>

(1. *Educational Informatization Strategy Research Base, Ministry of Education, P.R.C, Wuhan 430079, China*; 2. *Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China*)

**Abstract:** *The current integration of general artificial intelligence, represented by large language model (LLM) technology and with educational evaluation has become an important opportunity and significant trend in the reform of educational evaluation for the digital era. This paper reviews the development process of China's educational evaluation reform, analyzes the main features of applying artificial intelligence powered by LLM technology to education, and further explains the core focus of reshaping the educational evaluation reform with this technology, including the focuses on more and enriched content of educational evaluation, promoting more personalized, comprehensive, and ongoing evaluation methods, using a wider range of educational evaluation subjects, and more targeted data-driven evaluation approaches. Furthermore, the allround support and empowerment provided by LLM technology will facilitate the realization of a more scientific evaluation of outcomes, a more intelligent evaluation of the process, a more practical evaluation of the added value, and a more comprehensive and effective evaluation. Lastly, this paper offers meaningful reflections and suggestions for the practical application of general artificial intelligence of LLM technology for education evaluation, namely, to advance evaluation reform collaboratively, to make breakthroughs in critical evaluation technologies, to pilot and explore new evaluation paradigms, and to prioritize data ethics and privacy.*

**Key words:** *artificial general intelligence; large language model (LLM) technology; the reform of education evaluation; comprehensive evaluation*