

生成式人工智能应用于教师教育的价值对齐困境与超越

王凯¹ 龙鸿宇² 许心怡³ 国建文⁴

(1. 北京师范大学教师教育研究中心,北京100875; 2. 中央民族大学教育学院,北京100875;
3. 北京师范大学教育管理学院,北京100875; 4. 首都师范大学初等教育学院,北京100875)

[摘要] 本研究立足价值对齐视角,提出生成式人工智能应用于教师教育应与教育主体的价值观保持对齐,即确保赋予人工智能的目的是教育主体要达成的目的。生成式人工智能与教师教育价值对齐困境体现为三方面:一是生成式人工智能与教师教育属性的偏离,二是生成式人工智能与教师教育知识体系的脱节,三是教师与机器深度交互的弱耦合。本研究对上述困境进行宏观、中观、微观的归因分析,发现生成式人工智能应用于教师教育的硬性制度缺位、生成式人工智能与教师教育对接机制缺失、人工智能技术缺陷以及教师数智素养滞后是造成价值对齐困境的深层原因,进而提出破解价值困境的三条路径:推进生成式人工智能应用于教师教育的法制化建设、建构生成式人工智能与教师教育的对接机制和增强人机融合能力。

[关键词] 生成式人工智能;教师教育;价值对齐;数智素养

[中图分类号] G451.2

[文献标识码] A

[文章编号] 1007-2179(2025)03-0084-08

生成式人工智能通过人机对话的方式生成各种关于教师教与学的内容文本,扮演了学习助手、智能教师等角色(刘丙利,2024)。这种新型对话式教育在教师教育领域的应用为教师专业发展提供了新的机遇,包括带来专业理念的革新、专业知识与能力的增长、教师专业情意的深化(吴军其等,2023)、教学和行政工作负担的减轻(宋崔等,2023)等。教育部(2021)明确指出,需推动人工智能与教师培训融合,鼓励地方、高校、机构探索“智能+教

师培训”,对教师实施智能化、个性化、交互性、陪伴性培训,形成人工智能支持教师终身学习、持续发展的机制。由此可见,向智能化转型已成为教师教育的未来发展趋势。然而,生成式人工智能应用于教师教育领域仍存在诸多问题。例如,生成式人工智能在教师教育中的应用使“以机器为师”成为常态,“以人为师”的教育活动正面临合法性危机(郭顺峰等,2023),这反而不利于教师教育的发展;生成式人工智能还可能会生成偏激与“毒”

[收稿日期] 2025-04-14

[修回日期] 2025-04-19

[DOI编码] 10.13966/j.cnki.kfjyyj.2025.03.009

[基金项目] 北京师范大学教师教育研究中心2025年度培育课题(25CTERPY001),中央高校基本科研业务费专项资金资助项目(1243300004),北京师范大学教育学一流学科培优项目(YLXKPY-XBTC202203)。

[作者简介] 王凯,讲师,教育部普通高校人文社会科学重点研究基地北京师范大学教师教育研究中心,研究方向:教师教育、教育技术;龙鸿宇(通讯作者),硕士研究生,中央民族大学教育学院,研究方向:教育技术(2731137113@qq.com);许心怡,硕士研究生,北京师范大学教育管理学院,研究方向:教育社会学、教育政策与管理;国建文,讲师,首都师范大学初等教育学院,研究方向:教师教育。

[引用信息] 王凯,龙鸿宇,许心怡,国建文(2025). 生成式人工智能应用于教师教育的价值对齐困境与超越[J]. 开放教育研究,31(3): 84-91.

性语言, 产生误导信息, 甚至加剧教师教育资源不均等现象(矣晓沅等, 2023)。当应用机器的目的或效果不是教育主体想要的目的或效果时, 这种现象便被称之为“价值不对齐”。教师教育强调尊德性而道问学, 培养教师的德性与增强知识素养是其突出目标。生成式人工智能应用于教师教育, 首先要回应的是主体发展中对“德性”和“问学”的要求, 生成式人工智能的应用存在怎样的价值对齐困境? 什么因素造成的? 应采取怎样的措施? 本文基于价值对齐视角, 深入探析生成式人工智能应用于教师教育的困境, 并探寻破解路径。

一、理论分析框架

习近平总书记强调, 强教必先强师, 要把加强教师队伍建设作为建设教育强国最重要的基础工作来抓(新华社, 2023)。在智能时代, 教师教育转型具有前瞻性长远战略部署的意义, 是推进教育强国建设的必然要求(林敏等, 2024)。在智能技术的发展与应用过程中, 机器的“道德”问题日益突出, 人工智能价值对齐理论的提出反映了人类对机器道德问题的忧思。

人工智能先驱维纳(Wiener, 1960)提出人工智能对齐问题, 并将其描述为“我们必须非常确定赋予机器的目的是我们想要的目的”。之后, 罗素(Russell, 2019)指出, 生成式人工智能要确保与人类价值观对齐, 即价值对齐(value alignment)。事实上, 计算机系统存在着针对部分群体的价值歧视, 而价值对齐是帮助人工智能获得道德智能、改变这种不公平现象的主要途径。这里的“价值”具有双层含义:一是人类主观认为更可取的东西即偏好, 二是合理的道德原则(Kim et al., 2021)。实现生成式人工智能价值对齐的方法是帮助机器习得人类偏好与机器道德。强调价值对齐旨在使人类关注人工智能应用的目的而非能力, 使人工智能的应用更加符合人类的意图与价值(Ji et al., 2023)。

加布里埃尔(Gabriel, 2020)提出三种实现人工智能价值对齐的路径:一是寻找一种具有全球性共识的道德信仰;二是利用“无知之幕”理念帮助人工智能建立公平原则, 这要求人工智能决策者完全理性且绝对公平;三是基于社会选择理论, 选择大

多数人认同或倾向的价值取向。鉴于人类在道德价值观上不存在共识, 社会选择是解决价值对齐问题的必要工具(Prasad, 2018)。因此, 生成式人工智能应用于教师教育的价值对齐问题可以通过社会选择实现。从技术视角看, 价值对齐可分内部对齐与外部对齐:技术系统内部对齐即内部对齐, 人自身、人与技术的对齐即外部对齐(闫宏秀, 2024)。

反思生成式人工智能在我国教师教育的应用现状发现, 价值对齐困境主要发生在技术自身、教师教育自身以及技术与教师教育之间, 包括生成式人工智能与教师教育属性的偏离、生成式人工智能与教师教育知识体系的脱节、教师与机器深度交互的弱耦合(见图1)。从宏观层面看, 主要原因在于缺乏规范智能技术应用行为以及教师运用智能技术行为的硬性制度;从中观层面看, 生成式人工智能与教师教育知识生成范式存在偏差, 同时教师教育知识体系革新速度落后于智能时代的要求, 导致二者对接机制缺失;从微观层面看, 智能技术在短时间内快速发展但仍然存在技术缺陷, 同时教师教育主体的数智素养发展滞后, 以致二者未实现能力对齐。在国家层面, 政府应推进生成式人工智能应用于教师教育的法制化建设, 规范生成式人工智能与教师教育主体的行为;在知识体系方面, 两者之间应建构生成式人工智能和教师教育的对话机制;而在未来发展中, 智能技术与教师教育应携手并进, 增强人机之间的交互融合能力。

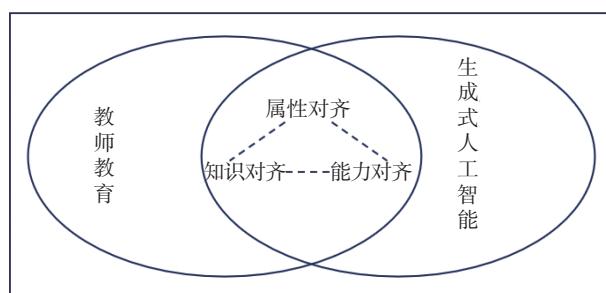


图1 理论分析框架

二、价值对齐困境

(一) 生成式人工智能偏离教师教育属性

当前我国教师队伍建设的阶段性目标是到2035年实现教师教育的中国式现代化, 造就党和人民满意的高素质专业化创新型教师队伍(中共中

央国务院, 2018)。习近平(2023)总书记论述具有中国特色的教育家精神时指出, 新时代人民教师应具备心有大我、至诚报国的理想信念, 言为士则、行为世范的道德情操, 启智润心、因材施教的育人智慧, 勤学笃行、求是创新的躬耕态度, 乐教爱生、甘于奉献的仁爱之心, 胸怀天下、以文化人的弘道追求。这些论述为我国建设新时代教师队伍明晰了价值取向, 也为建设教育强国指明了方向, 提供了根本遵循, 具有重要的理论价值与实践意义(李陈锋等, 2024)。因此, 培养高质量、具有教育家精神并服务于党和人民的现代化教师队伍是教师教育的应然目标。

教师教育具有主体多样性、对象自主性、过程互动性和活动人文性等特点(赵海亮, 2021)。不同教育阶段的教师培养标准存在差异, 中小学教师重视基础教育学科素养、知识的提升以及推动基础教育学科教学变革能力的养成; 高等教育阶段的教师队伍不仅重视学科能力与素养提升, 其学术属性更加突出, 对科研能力、跨学科学习与研究能力要求也更高。当前应用于教师教育领域的生成式人工智能以通用型为主, 虽然其具有普适性, 但缺乏共情教师教育独特场景的能力, 因此不能产生相应的效果。在对话理论视角下, 教育是个对话的过程, 教师的天然知识伴随着人际互动随时更新。然而, 生成式人工智能这种预训练式模型受技术局限, 其内容是“预先封装起来的”, 从而造成其人文互动性质的丢失, 知识就不再随对话的更新而更新, 容易陷入教学内容获取主渠道悖论(苗逢春, 2023)。

在我国, 国家和社会对教师道德寄予更高的期望。师范的“范”被赋予更深刻的含义, 教师要承担起“行为世范”“匡扶天下”的责任使命。但通用大模型训练数据跨度广泛, 知识泛化垂直深耕不足, 无法透彻把握我国师范教育的独特意蕴。ChatGPT等西方科技可能还存在潜在的意识形态偏离风险。人工智能生成内容被应用于教师教育, 培养的教师队伍很可能失去中国特色, 甚至为我国意识形态安全埋下隐患。

(二) 生成式人工智能与教师教育知识体系脱节

生成式人工智能主要依赖算法和模型分析处理已有知识, 这决定了生成式人工智能是“知识的

重组而非创造”(李慧慧, 2024)。然而, 教师教育需要不断进行知识的生产和创造性革新。因此, 生成式人工智能集成的知识体系和教师教育知识体系之间存在脱节。这种脱节是双向叠加的因素造成的, 既体现在生成式人工智能知识体系集成的显性化与预制性, 也体现在教师教育知识体系集成的隐性化与生成性等方面。具体而言, 一方面, 合格的教师需要掌握“教给学生的知识”以及“教学的知识”, 即学科知识和教师专业知识。这些知识是对前人经验的总结, 是最终呈现在书本上具有普遍意义且系统化的知识, 即理论知识。但优秀教师的丰富经验和实践性知识是在大量教学活动中形成的, 它与课堂教学情境密切相关。从课程教学论看, 教师须具备备课、说课、上课、听课和评课等技能。“五课”技能的培养需要依托专门课程奠定知识基础, 借助微格教学使教师在可控的模拟教学环境中掌握特定的教学技能, 通过真实的课堂教学实战巩固提升学习效果(屠锦红等, 2013)。这些经验性实践知识来源于教师个体的感悟、思考, 常常难以用语言表达, 属隐性知识。训练生成式人工智能的数据恰恰是语言、文字等客观有形的显性知识, 其输出的知识大部分是理论知识, 缺乏合格教师必备的实践性知识。

另一方面, 生成式人工智能具有强人机交互性, 可以根据用户需求生成内容, 其输出的内容质量取决于训练数据的准确性与安全性。作为服务人类的工具, 安全、稳定、透明、公平、保护隐私是其必备的特点。然而, 自生成式人工智能投入使用以来, 训练数据是否安全、算法运行的透明程度、生成内容是否有害等问题越来越突出。尤其是当训练数据质量不高或者领域训练数据不足时, 人工智能产品会产生“模型幻觉”, 即生成的内容可能与事实、用户指令或上下文不一致(Huang et al., 2023)。如此一来, 通过生成内容的方式向教师教育主体传递的“知识”是否值得我们信任和学习? 此外, 生成式人工智能输出的内容是“预制”的, 知识不会随人机互动而更新, 智能机器不具备为教师主体提供课堂教学场景的能力。因此, 将生成式人工智能应用于教师教育, 为教师教育主体提供的知识只是一成不变、真实性与专业性存疑的部分理论知识, 教师仍需来自微格教学场景的实践性知识。

(三) 教师与机器深度交互的弱耦合

布鲁姆将学习者的认知思维发展分为“记忆、理解、应用、分析、评价与创造”六个阶段。其中，“记忆、理解”属浅层学习。当学习者的认知思维处于这两个层次时，其对知识点的认知仅停留于表层理解与识记，习惯于单向的知识传授和通过独立探究来获取认知能力的进步(董庆华等, 2024)。后四个阶段属深度学习，学习者需激发内在动机，通过积极思考与探索，掌握知识的深层含义，实现众多知识间的融会贯通，促进思维向更高阶段发展。教师作为重要的文化传承与创新者，应以深度学习为目标，达到高阶认知思维水平才能发挥“启智润心”的作用。一般来说，深度学习教学法具有意义性学习、内容关联性、任务真实性、活动情境性、社会性交互、逆向性评价等特点(杨南昌等, 2024)。在深度学习层次下，生成式人工智能需为教师学习提供真实的学习情景与任务，并能与教师情感交融，即教师和生成式人工智能对话时，二者能达到思维共鸣，让生成式人工智能的能力成为自身能力的延续(Lesh et al., 2004)。

教师教育不仅是传递知识、创造知识的过程，更要以情感滋养教师，晓之以理，动之以情，培养“经师”与“人师”相统一的大先生。其培养体系应包含技术价值与人文价值。生成式人工智能虽能通过数据训练等方式，模拟人类交流口吻输出有一定感情色彩的内容，但难以理解人类的“言外之意”，实现人文关怀，无法扮演“人师”。同时，肢体动作、面部表情也是个体向外界传递信息的重要途径，生成式人工智能尚不能捕捉这些非言语信息，社会性交互效果弱，其能力不能成为教师自身能力的延续。此外，教师教育主体在应用生成式人工智能的过程中，“主体性危机”愈发突出。例如，教师依赖智能技术完成学习与教学任务，在教育活动中“反主为客”，主观能动性缺失，形成思维惰性。

三、成因分析

(一) 宏观层面：生成式人工智能应用于教师教育的制度缺位

制度缺位是指在生成式人工智能赋能教师教育的情境中缺乏必要且有效的制度规范与保障。

其体现在生成式人工智能应用教师教育的全过程以及教师教育主体的规范两个方面。我国2023年发布《生成式人工智能服务管理暂行办法》(以下简称《办法》)，对生成式人工智能的应用作出总体规定，规范智能技术的开发者、监管者以及使用者的伦理道德。但《办法》是针对生成式人工智能在短期内快速渗透各行业而采取的“应急措施”，其本身仍有许多待改进之处(陈万球, 2023)。教师教育对国家具有战略意义，因此对生成式人工智能要求更高。在这一原则下，完善的技术准入标准、内容生产标准、质量评估和监督标准等，可为生成式人工智能产品提供道德标准与法律规范，帮助产品开发者明晰党和人民对教师队伍培养的需求，为人工智能应用于教师教育营造井然有序的学科环境。

在教师教育中，人类作为道德主体也对智能技术的应用存在认知失范，如将智能技术视为解决教育问题的“万能之物”或冲击教育对象主体性的“洪水猛兽”，抑或认为智能技术带来的影响完全取决于使用者的使用方式(庞茗月等, 2022)。教师要么丧失教育活动“以人为本”的初心，将智能技术置于绝对地位并把教育活动的重心从“人”转到“技术”，使教育沦为证明智能技术优势的工具，要么全盘否定智能技术的优势，担忧“机师”终将取代“人师”、产生思维惰性等问题。教师还忽视了重要的一点，即智能技术在生产过程中早被灌输了开发者及监管者的价值取向与意识形态，智能技术应用于教育场景的过程实际上受多方主体价值观的影响而不完全由教师主导。人类作为道德主体，必须明晰使用生成式人工智能的道德规范，才能真正掌握并使用好人工智能技术。然而，我国以往师德师风建设的指导意见并未考虑生成式人工智能的影响。生成式人工智能时代对教师道德的建设要求显然提出了新要求，教育管理者需重新审视师德师风建设的目标与内容，帮助教师明晰正确的教师教育价值观，健全生成式人工智能应用于教师教育应遵守的道德规范，如伦理操作标准、能力指导细则和监督体系。

(二) 中观层面：生成式人工智能与教师教育对接机制的缺失

生成式人工智能嵌入的知识系统与教师教育自身知识体系之间的对接交流存在错位，尚未建立

人-机有效沟通协作的机制。1)生成式人工智能与教师教育存在知识生产范式的差异。大部分生成式人工智能在数据训练中,接受了中外数据的喂养,形成的话语框架可能存在语言偏差。例如,使用不同语言向 ChatGPT 提问,ChatGPT 会迎合不同的语言体系及价值偏好给出不同答案(周茂君等, 2024)。实际上,生成式人工智能将生成内容的选择权交给了用户,它不能基于当下语境和多元文化视角,提供适用于特定文化环境的概念与材料(van den Berg & du Plessis, 2023)。生成式人工智能的知识输入仅仅是对符号的加工处理,忽视身体活动对认知的影响,形成的是非具身性认知,产出的是显性知识。教师的隐性知识具有不可编码性,依赖具身经验与情境互动。身体活动产生的经验对教师成长的重要性不言而喻。生成式人工智能知识生产的不确定性与非具身性同教师教育需要的确定性知识、具身性知识产生矛盾,阻碍二者的有效沟通。2)教师教育知识体系有待完善。教师教育知识体系的理论知识、实践知识均源自实践,“科学知识是镶嵌于实践当中的,具有地方性、语境性及动态性的特征……从最初形态到最终的普遍知识,事实上经历了去地方性和去语境化的过程。”(王娜, 2010)。目前教师教育知识体系存在具有普遍意义的科学知识,是地方性知识的去语境化,这种去语境化的普遍性知识抽离了时空语境、情景语境乃至主体语境(李长吉, 2019)。在当下,科学技术已强势介入教师教育活动,“类人智能”成为教师教育活动除人类外的全新主体,新的智能知识理应作为新型知识观念(即地方性知识)纳入教师教育知识体系。然而,当前教师教育知识体系对这种新型知识并不敏感,知识体系的调整速度缓慢,教师教育知识体系与智能知识并未融合发展,缺少智能知识的参与,既影响教师队伍人才质量,更阻碍教师教育领域的智能化、现代化进程。

(三)微观层面:生成式人工智能技术的“局限”与教师数智素养的“缺乏”

生成式人工智能训练数据在挑选阶段会受研发人员主观偏好、操作偏差等因素的影响,难以保证智能产品的公平性及面向特定用户的客观的“价值取向”。在生成式人工智能应用阶段,数据驱动算法的运算过程较复杂、运行机制难以干预,具有

很强的不可控性,甚至出现“算法黑箱”。“算法黑箱”使大模型的可解释性降低,逃避公众的道德监督。此外,教师教育需要在具体情境中以真实任务为目标,通过对话互动,使主体获得理论性和实践性知识。而生成式人工智能受技术限制,不具备营造真实教学情境、开展实践性教与学的能力,仅通过内容输出远不能达到教师教育的要求。

人工智能教育、计算教育学等前沿学科蓬勃发展,但教师教育体系与前沿学科发展成果、一线教育改革成果的结合并不理想,教师人才培养体系的发展落后于前沿学科的发展(朱旭东, 2023),导致教师教育主体数智素养发展滞后:一是教师教育主体的人工智能伦理道德观念落后,主体对生成式人工智能存在认知失范;二是教师教育主体智能技术的基本概念、理论欠缺,信息的了解与积累有限,不能客观看待智能技术的功能与作用;三是教师教育主体运用智能技术存在技能缺失,如存在智能技术应用的思维定式、功能固着等现象;四是教师教育主体评估智能技术和创造性应用智能技术能力欠缺。培养教师评估和创新智能技术应用能力是教师数字素养培育的最终目标。提高这种能力可激发教师应用智能技术的主动性,促使教师更积极地应用智能技术,实现教师与机器的深度交互。

四、未来路径

(一)推进生成式人工智能应用于教师教育的法制化建设

教师教育包含教师的培养、任用、研修三大要素(陈永明等, 2009),涉及教师培养者、政策制定者、实践管理者等主体。教师教育过程的长周期性、多场域性和多主体性特点,对生成式人工智能产品提出了严格要求,完善生成式人工智能应用于教师教育的规章制度对解决现存的价值对齐问题具有重要作用。

1)构建生成式人工智能在教师教育应用的法律治理体系。在智能时代,我们应当实行法律与科技共治模式,以促进制度优势与科技优势的纵深融合(张文显, 2020),使智能技术在教师教育领域产生巨大效能:通过现代化法律治理体系的建设与完善,促进生成式人工智能的输出对齐,为智能技术的健康发展提供法律基础,保护智能技术使用者的

合法权益。

2)建立专用型生成式人工智能应用于教师教育的监督与评价机制:一是明确各监督主体在生产和使用智能技术的职责,实现层层追责;建立智能技术信息共享制度,包括公开训练数据来源、算法运算机制,实现生成式人工智能在教师教育领域应用的透明化。二是形成动态评价反馈机制,以便生成式人工智能技术发展与教师教育主体的动态需求相适应。

3)制定针对教师教育主体的生成式人工智能技术应用标准。标准制定应汲取专家、教师、技术人员和公众等的智慧,内容应包括生成式人工智能技术使用的范围、方法、道德规范等,确保教师教育主体正确高效地运用智能技术。

(二)建构生成式人工智能与教师教育的对接机制

生成式人工智能的诞生是科技进步与人类文明发展的表现,智能技术应用于教育的过程打破了传统教育活动的时空限制,缓解了教育资源分配不均的矛盾,缩小了教育中的数字鸿沟。但目前生成式人工智能与教师教育的结合生硬粗暴,非“你中有我,我中有你”的有机结合,故破除二者沟通障碍十分重要。

1)优化教师教育人工智能数据库。教师教育的质量改进以证据为基础,证据源于数据,建立完善的教师教育大数据系统是推进我国高质量教师教育体系发展的迫切要求(赵英等,2021)。该数据库需包含我国意识形态和文化因子的数据,帮助智能产品获得理解语言多样性和多义性的能力,修正语言偏差,提高智能产品的意识形态安全;加强生成式人工智能的隐性知识训练,开发可将隐性知识转化为可编码显性数据的工具,帮助智能产品进行教育实践反思,破解经验性知识难以通过“类人”主体传递的难题。

2)实现生成式人工智能与我国现有教师教育平台的深度链接。我国现有教师教育平台(如国家中小学智慧教育平台)资源丰富,具有训练大语言模型数据需要的本土性、动态性、规范性等特点,同时智能技术提供的个性化学习、智能交互能力可以提高教师教育平台的资源动态调配性能,推进智能技术和教师教育专业的协同发展。

3)重构教师教育知识体系,嵌入智能知识图谱。这包括通过与计算机科学、心理学等的跨学科合作,探索生成式人工智能教师教育的应用场景,深化教师教育学科与全球新兴、前沿学科的联系,推动教师教育学科的智能化、现代化变革。教师教育还要总结和吸收相关经验,形成具有教师教育特点的智能知识图谱,建设“人工智能+教师教育”育人模式。

(三)增强人机之间的交互融合能力

生成式人工智能存在的算法黑箱、教学实践短板等问题及教师教育主体智能素养滞后,会影响智能技术赋能教师教育的效果。智能时代已然降临,类人智能与人类智能共同参与教师教育实践的双主体性已是大势所趋。未来,智能技术与教师教育主体应共同提高人机融合能力。

一是突破生成式人工智能技术瓶颈。1)走向可解释的生成式人工智能。已有研究证明,人工智能教育决策的解释对学习者的学习效果具有正向反馈作用,使用可解释性人工智能的教育应用可显著促进学习者的认知与元认知发展(陈昂轩等,2024)。增强人工智能算法的透明度,不仅可以促进教师教育主体对智能技术运行过程的理解,增强“人—机”信任,还能促进教师思维能力的发展,真正实现智能技术赋能教师教育。2)走向多模态融合的生成式人工智能。多模态技术可以帮助智能产品准确载入教师教育活动中人体传递的言语和非言语信息(如面部微表情、肢体动作等),并模拟真实的教学情景,供教师在虚拟课堂中学习理论知识、积累实践经验。3)走向更“现实”的生成式人工智能。虚拟现实和增强现实等技术可为教师教育主体提供“真实”的教学场景,丰富教师教学场景的内容与体验。

二是深化教师教育主体的认识。1)深化教师主体对“人—机”关系主客体地位的认识。教育过程是人类有目的地发起、有计划且有组织地实施的过程。“人”的主体地位意味着在教育中“人”这一要素始终是第一位的,人不应让位于机器,人类需始终占据教师教育过程的主体地位。2)重视教师教育主体关于智能技术伦理道德的培养。智能教育时代,社会对教师提出了全新的道德要求,正确应对智能技术带来的伦理挑战、规避智

能技术应用的风险, 已成为教师教育主体新的必修课。3)重视职前教师的跨学科教育, 帮助其建构智能知识框架。高校作为培养职前教师的主要场所, 应积极开展跨学科教育, 提升职前教师的数字胜任力, 提升教师队伍的整体数智素养。4)重视在职教师的数智素养培养, 提升其智能技术应用能力。政府与学校等机构既是培养在职教师的重要主体, 更是重要载体, 应共同发力, 通过搭建交流学习平台、提供专业课程等方式, 为在职教师的学习发展提供职业成长环境。

未来, 人们应当完善规章制度, 促使生成式人工智能向善发展、教师教育主体向上发展; 积极建构生成式人工智能和教师教育的对接机制, 形成教师教育与智能知识的深度融合; 突破生成式人工智能教师教育应用的技术缺陷以及教师教育主体的数智缺陷, 实现二者的能力对齐。

[参考文献]

- [1] 陈昂轩, 贾积有(2024). 可解释性人工智能有助于提升自适应学习的学习效果吗?——基于 29 项实验与准实验的元分析 [J]. 现代教育技术, 34(10): 92-102.
- [2] 陈万球(2023). 中国生成式人工智能顶层设计的伦理视界 [J]. 伦理学研究, 117(5): 100-107.
- [3] 陈永明, 王健(2009). “教师教育学”学科建立之思考 [J]. 教育研究, 348(1): 53-59.
- [4] 董庆华, 郭广生(2024). 走向深度学习的师生学习共同体建构研究——以北京市某高校创立跨学科师生学习共同体为例 [J]. 清华大学教育研究, 45(2): 144-151.
- [5] Gabriel, I. (2020). Artificial intelligence, values, and alignment [J]. Minds and Machines, 30(3): 411-437.
- [6] 郭顺峰, 李光, 邹红军(2023). ChatGPT 引发的“以人为本”合法性危机与应对——基于技术批判理论视角 [J]. 电化教育研究, (11): 28-35.
- [7] Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., & Qin, B. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions[J]. ACM Transactions on Information Systems, 43(2): 1-55.
- [8] Ji, J., Qiu, T., Chen, B., Zhang, B., Lou, H., Wang, K., Duan, Y., He, Z., Zhou, J., & Zhang, Z. (2023). Ai alignment: A comprehensive survey[EB/OL]. <https://arxiv.org/abs/2310.19852>.
- [9] 教育部(2021). 教育部财政部关于实施中小学幼儿园教师国家级培训计划(2021—2025 年)的通知 [EB/OL]. [2021-05-19]. http://www.moe.gov.cn/srcsite/A10/s7034/202105/t20210519_532221.html.
- [10] Kim, T. W., Hooker, J., & Donaldson, T. (2021). Taking principles seriously: A hybrid approach to value alignment in artificial intelligence[J]. Journal of Artificial Intelligence Research, 70: 871-890.
- [11] Lesh, N., Marks, J., Rich, C., & Sidner, C. L. (2004). "Man-computer symbiosis" revisited: Achieving natural communication and collaboration with computers[J]. IEICE Transactions on Information Systems, 87(6): 1290-1298.
- [12] 李陈锋, 宋东霞(2024). 以教育家精神引领教师队伍高质量发展 [EB/OL]. [2024-02-20]. http://www.qstheory.cn/qshyjx/2024/02/20/c_1130079044.htm.
- [13] 李长吉(2019). 论课程知识的地方性质 [J]. 课程·教材·教法, 39(1): 11-17.
- [14] 李慧慧, 和学新(2024). 生成式人工智能赋能中国教育学知识生产的困境及应对 [J]. 中国教育科学(中英文), 7(5): 51-60.
- [15] 林敏, 吴雨宸, 宋萑(2024). 人工智能时代教师教育转型: 理论立场、转型方式和潜在挑战 [J]. 开放教育研究, 30(4): 28-36.
- [16] 刘丙利(2024). ChatGPT 对话教育: 意涵价值、风险挑战与实践进路 [J]. 开放教育研究, 30(4): 53-61.
- [17] 苗逢春(2023). 生成式人工智能技术原理及其教育适用性考证 [J]. 现代教育技术, 33(11): 5-18.
- [18] 庞茗月, 戚万学(2022). 从解蔽主体到补偿“代具性”: 智能时代教育技术价值新论 [J]. 远程教育杂志, 40(3): 56-66.
- [19] Prasad, M. (2018). Social choice and the value alignment problem[J]. Artificial Intelligence Safety and Security: 291-314.
- [20] Russell, S. (2019). Human compatible: AI and the problem of control[M]. London: Penguin Uk: 3-23.
- [21] 宋萑, 林敏(2023). ChatGPT/生成式人工智能时代下教师的工作变革: 机遇、挑战与应对 [J]. 华东师范大学学报(教育科学版), 41(7): 78-90.
- [22] 屠锦红, 李如密(2013). 高师生“五课”技能综合培育的理论与实践 [J]. 现代大学教育, (3): 96-101.
- [23] van den Berg, G., & du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education[J]. Education Sciences, 13(10): 998.
- [24] 王娜(2010). 语境主义知识观: 一种新的可能 [J]. 哲学研究, (5): 89-95+128.
- [25] Wiener, N. (1960). Some Moral and Technical Consequences of Automation: As machines learn they may develop unforeseen strategies at rates that baffle their programmers[J]. Science, 131(3410): 1355-1358.
- [26] 吴军其, 吴飞燕, 文思娇, 张萌萌, 王嘉桐(2023). ChatGPT 赋能教师专业发展: 机遇、挑战和路径 [J]. 中国电化教育, (5): 15-23+33.
- [27] 新华社(2023). 习近平主持中央政治局第五次集体学习并发表重要讲话 [EB/OL]. [2023-05-29]. https://www.gov.cn/yaowen/liebiao/202305/content_6883632.htm.
- [28] 习近平(2023). 习近平致全国优秀教师代表的信 [EB/OL]. [2023-09-09]. https://www.news.cn/2023-09/09/c_1129854340.htm.
- [29] 闫宏秀(2024). 基于信任视角的价值对齐探究 [J]. 浙江社会科学, 334(6): 39-48+157.
- [30] 杨南昌, 梁慧芳(2024). 智能时代深度学习教学法的核心要义与行动框架 [J]. 电化教育研究, 45(5): 5-12.
- [31] 矫晓沅, 谢幸(2023). 大模型道德价值观对齐问题剖析 [J]. 计算机研究与发展, 60(9): 1926-1945.
- [32] 张文显(2020). 构建法科共治、法德共治、多元共治的法律秩序——共治: 补齐治理短板. 京报电子报 [EB/OL]. [2020-08-03]. <https://wap.bjd.com.cn/common/epaper.html>.
- [33] 赵海亮(2021). 教师教育的本质、特征与发展逻辑 [J]. 教师教育学报, 8(5): 28-35.
- [34] 赵英, 朱旭东(2021). 论高质量教师教育体系建构 [J]. 中国高教研究, 338(10): 52-57.

- [35] 周茂君, 郭斌(2024). 生成式人工智能传播中的偏向与规制——以 ChatGPT 为例 [J]. 学习与实践, 479(1): 33-41+32.
- [36] 中共中央国务院(2018). 中共中央国务院关于全面深化新时代教师队伍建设改革的意见 [EB/OL]. [2018-01-31]. http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/201801/t20180131_326144.html.
- [37] 朱旭东(2023). 健全中国特色教师教育体系, 补齐教师教育体系短板 [J]. 教育科学, 39(4): 1-4+18.

(编辑:赵晓丽)

The Value Alignment Dilemma and Transcendence of Applying Generative Artificial Intelligence in Teacher Education

WANG Kai¹, LONG Hongyu², XU Xinyi³ & GUO Jianwen⁴

(1. Center for Teacher Education Research, Beijing Normal University, Beijing 100875, China;
2. School of Education, Minzu University of China, Beijing 100875, China; 3. School of Educational Management, Beijing Normal University, Beijing 100875, China; 4. School of Primary Education, Capital Normal University, Beijing 100875, China)

Abstract: The widespread application of generative artificial intelligence (GAI) has promoted the digital-intelligent transformation of teacher education. However, GAI applications still faces significant challenges. Based on the perspective of value alignment, this study proposes that the application of artificial intelligence in teacher education must maintain alignment with the core values of the subject of teacher education, requiring that the objectives embedded in artificial intelligence, and reflect human-intended purposes. The value alignment challenge of GAI in teacher education encompasses three dimensions: first, the discrepancy between GAI and the core attributes of teacher education, the mismatch between GAI and the established knowledge framework of teacher education, and the weak coupling of deep interaction between teachers and machines. This study conducts an attribution analysis of the above challenges at the macro, meso and micro levels, and finds that the absence of rigid system for the application of GAI to teacher education, the lack of docking mechanism between GAI and teacher education, and the defects of artificial intelligence technology and the lagging of teachers' numerical intelligence literacy are the root causes. It proposes three major paths to meet the value alignment challenge, namely, promoting the legalization construction of the application of GAI in teacher education, constructing the connection mechanism between GAI and teacher education, and enhancing the ability of human-machine integration are effective ways to solve the value predicament.

Key words: generative artificial intelligence; teacher education; value alignment; digital-intelligent literacy