

生成式人工智能赋能思想政治教育的伦理风险与化解路径

——基于教育对象主体性视角的考察

闫瑞峰

(北京工业大学 马克思主义学院, 北京 100124)

[摘要]数智时代,生成式人工智能与思想政治教育的深度融合能够大大拓展教育创新的边界,但技术宰制逻辑也会对教育对象的主体性产生消解作用,使得数据化规训引发价值主体性的技术性解构,算法认知替代导致思维自主性的技术性剥夺,技术中介化造成主体间性的结构性转换。教育对象的主体性异化又极易诱发系列伦理风险:主体信念弱化引发价值信仰失序;主体思维惰化导致创新精神遮蔽;主体感知虚化造成人文情感稀释;主体责任意识淡化催生道德行为失范。实践中,有效化解生成式人工智能赋能思想政治教育带来的伦理风险,应坚持“以人为本、智能向善”的基本原则,着力以意识形态安全建设夯实思想政治教育的价值引领基石,加强以创新能力为导向的新型教育机制建设,改善智能辅助思想政治教育的人文环境,以及建立和完善基于智能场景的道德责任教育机制。

[关键词]思想政治教育;人工智能;科技;伦理;道德

[中图分类号]G641

[文献标识码]A

[文章编号]1002-5707(2025)6-0034-09

当前,以大语言模型为代表的生成式人工智能的广泛应用,极大地丰富和拓展了“人工智能+教育”的实践场景,促使传统教育生态不断向数字化和智能化方向演进。近年来,我国深入实施教育数字化战略,积极探索智能技术与教育实践的深度融合,教育数字化、智能化的顶层设计和政策文件不断出台并有序落地。教育技术化生存的必然性促使教育快速向虚实融合的新样态转变,思想政治教育在生成式人工智能的赋能下获得强大的变革性力量。生成式人工智能凭借自然语言处理、大数据分析、深度学习、多模态融合等综合性技术优势,拓宽了思想政治教育在个性化学习、知识图谱构建和沉浸式交互等方面的创新空间,进而对受教育者的价值塑造、知识习得和能力培养产生重要影响。但是,“科技是发展的利器,也可

能成为风险的源头。”^[1]智能技术工具理性的单向度扩张也会消解思想政治教育的价值理性根基,造成技术逻辑对教育过程的深度“殖民化”,促使教育对象的主体性发生应激性异化并诱发一系列伦理风险,亟须从理论和现实层面积极探寻应对之策。

一、生成式人工智能赋能思想政治教育 引致主体性消解的技术逻辑

马克思指出:“人的类特性恰恰就是自由的自觉的活动”^{[2][96]}。思想政治教育对教育对象主体性的培养,是实现人的自由全面发展的重要途径。主体性在实践中体现在多个维度,包括自主性、自为性、能动性、自觉性、自由性等方面。^[3]生成式人工智能

本文为2024年度北京高校思想政治工作研究战略课题“生成式人工智能在高校思想政治教育的应用状况与工作对策研究”(项目编号:BJSZ2024ZL01)阶段成果

技术与思想政治教育之间的矛盾,很大程度上内嵌于技术对主体性的消解功能之中,而通过构建基于数据化规训、算法认知替代、技术中介化的技术解释框架,能够为解构数智时代思想政治教育的伦理风险提供必要的理论支点。

(一) 数据化规训引发价值主体性的技术性解构

价值观教育是维护教育对象主体性地位和实现主流意识形态认同的重要途径,并通过基于主体间性的交互机制在主体价值认知、判断和选择中形成价值共识,从而实现价值理性的培育目标。在哲学本体论层面,价值主体性将客体价值与主体存在联结起来,价值形态的历时性演进始终围绕主体的历史性存在展开,从而在教育实践维度构建出“主体觉醒—价值认同—实践自觉”的转化机制,使个体价值选择与社会核心价值导向保持高度一致。然而,在生成式人工智能深度介入思想政治教育的情境下,数据化规训通过重塑教育对象的价值结构,逐渐形成一种新型权力运作模式,并利用数据分析、算法推荐、行为建模、数字画像等技术装置,对个体认知与行为进行精细调控,从而将教育对象的行为轨迹转化为可量化、可预测、可控制的数据流。这使得价值主体性所承载的独立性遭遇前所未有的技术性解构,个体独立进行价值反思与选择的能力被技术逻辑深度干预乃至控制。

一是价值认知框架的数据化锁定。在教学实践中,当信息推送、知识获取、自主学习均受算法支配时,个体的价值认知广度就会被“信息茧房”限定、认知的底层逻辑被数据化重塑,主体批判性思维被即时反馈机制消解,价值选择自由也被数据所裹挟。二是价值反思的时间性消逝。人工智能系统的实时数据矫正功能催生“微粒化生存”,个体在持续的行为调适中失去价值沉淀所需的时间纵深,内在价值尺度让渡于外部数据指标,导致思想政治教育面临价值稀释的人文困境。此时,教育平台使用匹配算法定义情感价值,教学评价系统将道德进行程序式量化,价值的主体间性建构被异化为技术参数的调试过程。三是价值观被数据化的算法形塑。数据化规训通过技术手段将人的行为、情感和思想转化为可计算、可预测、可

干预的数据对象,情感计算通过多模态识别机制(如表情、声音和呼吸)将复杂价值判断转化为可量化的数据符号,并以心理、行为和生理等测量方式进行感知和分析。^[4]例如,系统通过面部表情强度系数动态评估学生对特定政治议题的认同度并据此调整教学策略,实质上是生物信号与价值判断的强相关性假设。数据化规训对价值主体性的解构揭示了现代性危机的技术面相,但解构效应并不是技术发展的必然宿命,而是人类将价值裁判权让渡给技术系统的后果。

(二) 算法认知替代导致思维自主性的技术性剥夺

在思想政治教育场域中,保持思维自主性与独立性构成现代人格培育的内在要求,也是实现人的自由全面发展的基石。在马克思主义技术批判视域下,算法认知替代的本质是智能时代劳动异化的认知论延伸,由此形成基于数字异化的新型认知论困境。当生成式人工智能以异化的技术逻辑深度侵蚀思想政治教育本质时,工具理性对价值理性的僭越性事实使得主体思维的自主性被技术性剥夺,并被动地向算法操纵的认知范式趋同,由此技术对主体性形成宰制效应——算法认知替代导致思维自主性的技术性剥夺。

一是认知架构的技术化重构。一方面,数字界面凭借注意力经济机制实施认知捕获,算法推荐系统通过“信息茧房—认知惯性—算法依赖”的闭环驯化机制,使教育对象逐渐丧失自主信息筛选能力,进而形成界面认知主导的规训效应以及产生算法权威主导的认知驯化效应。另一方面,智能技术深度介入人类思维过程,导致认知主体与自身思维能力之间产生结构性疏离,其本质是数字时代马克思“劳动异化”理论在认知维度的延伸。当人工智能承担知识检索、数据分析、价值判断等高阶认知劳动时,技术工具从认知辅助手段异化为支配性力量,极易造成思维过程的外包化现象。此时,知识建构、矛盾分析、系统反思等核心认知环节被迫让渡给智能系统,致使主体元认知能力的系统性退化;算法推荐系统通过闭环驯化机制,使主体的自主认知沦为算法认知体系的附庸,造成认知模式的算法化。二是思维过程的算法化殖民。智能技术主导下的批判性思维的算法替代,促使人的思维从

辩证逻辑推演向模式匹配响应转变,从价值反思循环向即时反馈依赖转变,从历史情境化理解向碎片化信息处理转变。当教育对象依赖智能设备完成高阶思维任务时,其元认知循环因被技术代理取代而陷入认知外包陷阱之中不能自拔,算法提供的即时答案反馈会削弱主体自主性思考的动力,从而在问题辨析中容易呈现“算法即真理”和“非此即彼”的思维定势。算法逻辑对人类思维产生的隐性规训效应,能够将复杂的人类思考过程简化为可计算和可量化的标准化路径,使人在信息茧房中逐渐弱化自主思考能力,从而导致主体认知自主性在算法驯化中悄然瓦解。生成式人工智能的推理模式本质上是对人类非线性思维的过程性模拟,其技术性的拟态特质遮蔽了马克思主义否定之否定规律,促使主体的矛盾分析能力趋向平面化。

(三)技术中介化造成主体间性的结构性转换

处理好教育主客体关系是思想政治工作的重要方面,而以主体间性为代表的交往范式则是实现价值传递和情感联结的关键纽带。一般来讲,主体性和交互性是主体间性的前提条件,“主体间性把教育看作是主体间的交流,从而确证了教育是人本真的生存方式”^[5]。传统语境下,技术人工物仅作为客体接受人类的改造,二者之间的主客体关系较为清晰,人与物之间尚不存在基于主体交互的主体间性。然而,在人与人工智能的交互过程中,智能系统所具有的类主体性特质使其逐渐具备自主改造客观世界的实践能力,进而使传统人技之间的“主—客”关系转变“主—主(类主体)”关系,主体间性由此发生结构性转换。人工智能驱动思想政治教育场景下,原本“主(师)—主(生)”的二元主体结构转变为“主(师)—主(类主体)—主(生)”三元主体结构,交互主体的变化同样会使主体间性发生结构性转换。一是教育关系的算法化重构引发主体间性异化。传统思想政治教育中“师一生”的双向度具身交互,被异化为“人—机—人”的三向度技术中介模式,交互沦为以数据流为核心的符号交换游戏,人际交往中至关重要的多模态具身交互(如情绪波动和语调变化等情感传递)在虚拟界面被削弱而产生“交互贫困化”困境。二是主体性结构

的双重迁移危机。智能系统的准主体化发展催生主体性分配的零和博弈,致使教育者的专业权威向算法系统迁移(如教学评估权让渡给学习分析系统),受教育者的价值建构向数据模型迁移(如思想动态被简化为情感分析曲线),致使主体性呈现“空心化”与“碎片化”现象。

在智能技术介入教育场景中,主体间性的结构性转换大致存在两种形式。一方面,在弱人工智能作用下,技术中介化拓展了人的主体性空间,并能够使其在跨时空的虚实之间自由转换。加拿大学者麦克卢汉指出,技术对人类认知过程的延伸作用,“正像我们的感觉器官和神经系统凭借各种媒介而得以延伸一样”^[6]。技术作为媒介性存在,不仅重构了物理空间与数字空间的交互界面,也消解了生物体与人工物之间的本质区隔,拓展了人类主体性的外延。另一方面,强人工智能的本体论中介效应,驱动智能系统突破传统技术人工物的客体边界,进而生成具有准主体性的认知实体。这种范式跃迁不仅引发人类主体性向算法系统的认知让渡(如决策让渡),而且导致主体性结构发生根本性改变。强人工智能通过深度学习与自主决策能力的不断进化,逐渐突破传统技术人工物的客体定位,催生出一种兼具工具客体性与认知主体性的双结构存在体,形成一种打破传统主客二元框架的复合交互系统。尽管可将其视为“披着主体性外衣的高阶客体”“伪主体间性”^[7],但“具有自主意识的强人工智能不再是一个执行程序的工具,它的意志和行动也不再是代表着主人或人类的意识与行动。”^[8]

二、生成式人工智能赋能思想政治教育 带来的主体性伦理风险

技术的两面性特征给教育技术化带来的重要启示在于,我们在实践中既要充分发掘技术推动教育发展的正向效能,也要时刻警醒其所带来的各种风险和挑战。现实中,教育技术化的异化极易造成机器的“人化”和人的“机器化”的主体性危机,^[9]产生符号化身份、人际关系疏离、共识凝聚阻碍等问题。^[10]不仅如此,在生成式人工智能赋能思想政治教育情境下,技术的单向度工具理性的过度张扬会使教育对象的主体性受到技术逻辑宰制而发生应

激性异化,从而在技术与伦理关系失调状态下衍生多重伦理风险。

(一) 主体信念弱化引发价值信仰失序

理想信念教育是引导学生树立正确世界观、人生观、价值观的重要途径。信念主要指主体对某种思想、观念、文化的深信不疑,其形成和变化深受意识形态状况的影响。价值信仰主要是指主体基于共同的价值目标期待所共同选择的价值理想或承诺,^[11]其形成主要依赖于特定的理想信念,或者说信念是信仰的基础、信仰是信念的高阶表现形式。思想政治教育作为阶级统治的必然性存在,^[12]通过社会主流意识形态的价值整合功能,系统塑造特定的价值信仰体系。“理想信念的坚定,来自思想理论的坚定。”^{[13][50]}思想政治教育中的思想理论具有浓厚的意识形态属性,是系统形塑主体理想信念和价值根基的重要源头。生成式人工智能的强大知识生产能力能够为思想理论学习创造良好条件,但其生成的内容具有高度的自主性、创造性及不确定性,当算法模型与主流意识形态产生偏差或者背离时,智能生成的内容就有可能会把异质性价值观以“软性灌输”的方式传递给受教育者,进而动摇其理想信念根基并使其弱化对社会主流价值观的信仰。

生成式人工智能所产生的意识形态风险,与训练数据来源和技术缺陷密不可分。一方面,源自负面价值形态的数据信息引发的意识形态风险。现实中,算法设计实践会受到功利论、德性论、义务论等价值立场的深度影响,^[14]负载着多元伦理价值、文化形态、主观偏见等数据的人工智能体,能“以一种看似控制自然的技术逻辑控制着社会权力,塑造人们的价值观念和行为模式”^[15]。例如,ChatGPT大语言模型的训练语料主要来自西方价值体系,其生成内容所负载的价值取向则以西方主流价值观为主。^[16]如果人工智能采用的训练数据中存在政治立场偏见、种族偏见、历史虚无主义、拜金主义、极端自由主义等违反社会主义核心价值观的内容,那么智能生成或算法推荐的内容也极有可能与之相匹配,人工智能体便会成为大量意识形态风险的智能“发源地”。一旦这种有害内容在思想政治教育实践中广泛扩散并深度影响受教育者的世界观、人生观、价值观,那么社会

主流意识形态和价值观将有被边缘化的风险,由此在学生群体中就有可能造成价值信仰失序并弱化其正向理想信念。另一方面,技术缺陷引发的意识形态风险。其一,算法推荐系统通过用户画像构建的“过滤气泡”,促使信息接触面呈现超几何收缩和信息同质化现象,导致认知多样性衰减与价值批判力钝化。其二,AI幻觉的技术性缺陷引致“技术无意识”性的说谎和虚构事实,制造本体信任危机和动摇价值判断的理性基础,对价值观教育构成潜在威胁。其三,大模型训练数据中的文化偏见可能通过智能教学系统进行隐性渗透,并将相关的隐性偏见传导给学生,致使价值观从主体间性的意义网络蜕变为算法可操作的参数空间,从而为价值扭曲埋下隐患。

(二) 主体思维惰化导致创新精神遮蔽

创新能力的培育不仅是教育现代化的本质要求,更是思想政治教育实现铸魂育人使命的重要着力点。思想政治教育以人的思想观念重构为价值旨归,而提升主体的创新能力从根本上依赖于以创新思维为核心的创新精神的培养。创新精神、创新个性、创新才能是思想政治教育的重要内容,思想政治工作者应当扮演启发学生创造性思维、创造性想象力和培养创造性人才的灵魂工程师角色。^[17]正因如此,思想政治教育培养学生创造性思维的重要性,集中体现在马克思主义哲学指导、树立正确的价值导向、塑造创造性人格、提供科学思维方法等方面。^[18]然而,学生在创新实践过程中对智能技术的过度依赖,使得问题发现、思辨过程、原创表达等创新要素被算法推荐、数据建模、自动生成的技术取代,导致创新活动从主体性创造退化为技术性操作,创新实践在“技术增强”与“主体性守卫”之间形成一对张力性矛盾。从知识创新角度来讲,目前生成式人工智能尚难以完成高阶创新任务,只能模拟人类思维从事低阶创造工作,其所具备的创新能力属于“有中生有”的类型,还无法像人类那样完成“从无到有”的新知识生产。^[19]进一步讲,生成式人工智能虽能完成基于数据关联的组合式创新,但主要以“低阶创新”或“伪创新”为主,其在突破性或原创性创新层面仍存在本质局限,尚不能满足高阶创新需要。

面对智能技术与思想政治教育深度融合带来

的便利性和高效率,如若教育对象的辨识力不够或自控力不足,则容易对人工智能体产生过度依赖,从而遮蔽和弱化马克思主义科学思维和创造性人格的积极作用,催生和放大主体的思维惰性并导致浅层思考、弱思考泛滥,由此产生普遍性的创新精神遮蔽危机。一方面,过度性技术依赖引发主体认知降维,以及批判性思维出现结构性萎缩。无底线地依赖生成式人工智能进行知识生产,易催生不加思索的知识“拿来主义”之风,使主体的思考力、判断力与创造力渐趋弱化,最终导致主体思维惰化、批判精神消减。算法路径依赖会使主体认知过程出现异化,人工智能的即时答案供给会使主体跨越“问题悬置—辩证反思”的必要性认知阶段,导致其思维惰性不断固化;持续依赖人工智能完成知识整合与价值判断,使自我监控与自我调节机制逐渐退化,造成主体元认知能力不断萎缩。机器主导的知识生产模式本质上是对辩证认知过程的算法化重塑,诱使使用者从能动的实践主体沦为技术产品的被动消费者而产生主体客体化。另一方面,泛滥的“智能创造”对主体创新能动性的深度侵蚀。主体的创新能动性被泛滥的“智能创造”所取代,“劣币驱逐良币”的创新循环编织出低阶性“创新茧房”,进而导致主体創造力匮乏、创新能力弱化和创新意识淡化。由此,低阶创新不断强化和创新同质化危机并存,创新蜕变为参数调整的技术游戏,创新能级被锁定在有限空间内,造成知识生产的创造性贫困。

(三) 主体感知虚化造成人文情感稀释

人文情感本质上是主体间性的具身化表征,是维系社会伦理关系和实现个体发展的重要纽带,其中主体感知能力则是生成情感能力的认识论前提,在思想政治教育中发挥着不可替代的情感传递作用。思想政治教育是一项充满思想性和人文性的教育工作,尤为注重基于实际感知的亲身体验带来的情感交流和人文关怀,以此不断促进教育对象的情感发展和道德培养。传统思想政治教育的具身化情感培育范式依托物理在场的“主体间对话”,能够以情境化感知和具象化互动塑造出明显的主客交融的在场性优势,实现从感官刺激和情绪唤醒到价值反思和行为内化的合理转换。这种主客交融让感知既真实又立体,有利

于师生之间人文情感的表达和传递,以及实现具有教育温度的情感交流。然而,人工智能体本身并不具备情感体验能力,仅是一种基于理性的交互主体。^[20]人工智能体对情感的认知与感知是建立在理性的客观数据之上,其情感模拟本质是算法对情感外显行为在统计学意义上的复现过程,没有喜怒哀乐、共情、感情、美感等人类所具有的真实情感能力。一方面,人工智能系统生成的虚拟教师通过数据分析和情感计算生成的“标准化共情”,容易割裂情感表达与内在价值的有机联系,在“拟像替代真相”和“拟像与真相错位”的幻象中诱发情感交流的符号化危机。另一方面,现有人工智能技术虽然在一定程度上能够模拟感官刺激,但无法复现实际感知的温度传导和情绪唤起等具身性要素,从而导致具身认知的消解。

虚拟教师缺乏真实情感反馈和情感共情能力,对其过度依赖会严重阻滞对学生的人文情感的具身化培养,泛滥的虚拟场景极易导致情感认知断层、共情体验“去身体化”等现象,由此在面对真实社会问题时容易出现“情感钝化”。这在一定程度上会消解思想政治教育的人本性,原本师生之间所具有的丰富人文情感交流和人格教育有可能会被冷漠、呆滞、无感的符号所代替,主体的存在意义和价值逐渐被机器消解和同化,并朝着主体感知虚化的极端方向蔓延。尽管人工智能技术的类人能力不断增强,但缺乏自由意志、灵感、情感、伦理等人所独有的人伦要素,^[21]使其在情感理解、心灵共情、精神感受等方面的作用依旧有限。人机交互仅仅通过数据、算法、代码、模型等判断推理形式来模拟人类行为,终究无法理解人类的精神世界和情感本质。人工智能对情感教育的过度操控,使得“本该拥有的价值与意义被过滤,思维、价值观的碰撞被简化为程序化、单一化的算法运行,原本充满智慧、富有温度的人际关系变为冰冷、机械的信息符号交互过程,这必然会导致个体共情能力下降以及体验感的降低。”^[22]人工智能无法代替人类的情感力、想象力、道德力,人机关系异化导致人机同质化或人依附于机器的情景,会让本应充满“人情味”的思想政治教育成为冷冰冰的“机器教育”。

(四) 主体责任意识淡化催生道德行为失范

责任感对个人道德行为规范具有重要的形塑

作用,而对主体责任意识的引导和培养是思想政治教育的重要方面。智能技术的工具理性对主体性的僭越,既通过数据、算法等手段增强了教育过程的可控性,又在责任伦理层面制造了主体性消解的困境,从而在技术赋权与责任伦理之间形成一种结构性张力。当技术理性突破工具性边界并演变为支配性力量时,智能技术对人的自主性的技术性剥夺会使主体深陷数字化的虚拟镜像之中,主体责任意识被算法逻辑重构而沦为外部规训的被动反应并丧失“自我立法”的实践根基,由此引发责任伦理危机。在生成式人工智能赋能的教育场域中,系统开发者、教师、管理者、学生等主体构成的多维责任网络,呈现出责任主体离散化、责任边界模糊化和责任认知碎片化的特征,从而产生“责任盲区”或“责任真空”,如系统开发者关注产品性能和设计缺陷责任、教师关注课堂应用和使用责任。在多元责任主体共生的网络生态中,责任悬置与责任稀释现象相互交织,形成连锁反应的责任分散机制(即“旁观者效应”),从而导致主体责任意识在被技术性解构的过程中衍生出系列道德失范行为。

一是平台匿名化带来的道德失范行为。平台通过技术手段剥离用户真实身份标识,构建起看似自由平等的虚拟空间,却无形中为教育对象提供了逃避问责的“数字面具”。匿名化智能平台可能削弱道德责任感,诱使教育对象以“虚拟行为不具现实后果”为由逃避问责,如学生在平台讨论区的语言暴力等失范行为发生率远高于实体课堂,在一定程度上印证了身份隐匿或模糊对主体责任悬置所产生的负面影响。二是道德计算化导致的伦理危害。在智能环境下将道德评价转化为数据指标时,也有可能陷入“德性计算化”的误区,导致学生将学习效果简化为“数据标识”。当道德行为被简化为可量化的“数据标识”(如互动频率和情感波动等)时,学生就容易陷入“数据主义”认知陷阱,从而将道德实践异化为“数据刷分”的技术操作。例如,为追求平台积分或算法推荐,刻意制造高频互动、伪造情感反馈,甚至将道德反思简化为“符合标准答案”的应试策略。三是学术诚信危机。当技术中介突破工具性边界来重构知识生产逻辑时,学术创作的主体性根基就会发生动摇并

使创作劳动趋于异化,导致责任认知从“作者中心”向“人机混合”乃至“机器主导”的范式迁移。当前,生成式人工智能在学术领域的广泛应用,催生了论文撰写、设计创作、报告写作等场景的效率革命,但同时也引发了抄袭、作弊、造假、版权争议等学术失范问题,尤其是智能生成、智能改写、智能润色等功能模糊了学术创作的原创性边界,催生直接复制、隐性拼贴、AI协同造假等学术失范行为。这些现象根源于主体责任被技术黑箱操作所造成责任认知漂移效应,使得学术主体在面对事实核查、版权归属、数据真实性等问题时,表现出明显的否定性态度和漠视性倾向。四是主体人格矮化危机。在人工智能技术主导的数字化教育场域中,主体虚拟身份的增殖性建构与现实身份的持续性萎缩形成张力,促使数字人格对现实人格形成强烈的挤压效应,现实人格在技术规训中逐步弱化或丧失道德担当能力而被矮化或异化,由此导致人格发展的畸形化与精神世界的贫瘠化,从而与人的自由全面发展背道而驰。一方面,数字孪生致使自我人格在虚实之间出现分裂,主体的多重人设并存导致人格碎片化及自我反思的断裂,容易使线上线下的同一主体人格处于严重的割裂状态并难以形成稳定的道德原则体系。另一方面,在主体人格实现“脱实向虚”的数字化重构过程中,虚拟身份的无序扩张导致现实人格的空心化。

三、生成式人工智能赋能思想政治教育 衍生的伦理风险的化解路径

生成式人工智能赋能思想政治教育实践,是一个正面效能与负面风险并存的辩证性过程,我们应根据思想政治教育规律和人工智能技术特性找寻科学的伦理治理之道。人工智能赋能思想政治教育必须突出人的主体性地位,既要以人本主义理念推动技术发展,也要以负责任的态度将技术融入教育实践。新时期,防范化解生成式人工智能对思想政治教育带来的伦理风险,要立足落实立德树人根本任务,坚持“以人为本、智能向善”的基本原则,将人工智能的技术理性纳入思想政治教育规律框架之下,使技术效能按照教育的价值理性进行有序释放。

(一)以意识形态安全建设夯实思想政治教育的价值引领基石

一个社会的主流价值观深刻形塑着社会主义主流意识形态的伦理特性,而社会主流意识形态的伦理特质反过来又深度影响着社会主流价值观的状态。思想政治教育肩负着传播社会主流价值观的艰巨任务,在实践中要以意识形态安全建设为抓手,驱使人工智能的技术理性服务于思想政治教育的价值理性,“提高网络育人能力,扎实做好互联网时代的学校思想政治工作和意识形态工作。”^[23]一是要不断增强主流意识形态的引领力,确保人工智能的价值输出符合社会主流价值观。坚持党对教育系统意识形态工作的绝对领导,建立和完善符合我国主流意识形态和价值观的语料库,特别是要把马克思主义价值理念、社会主义核心价值观以及中华优秀传统文化深度嵌入人工智能系统当中,以及提升算法的价值敏感性设计能力;加强对网络空间的正本清源,净化不良思想和错误思潮的传播土壤,强化基于社会主流价值观的主动性数据“投喂”和内容推送,使生成式人工智能在赋能思想政治教育过程中成为巩固主流意识形态阵地的助推器和倍增器。同时,相关部门也可以尝试探索建设专门针对思想政治工作的智能平台,并以公共责任的形式将其嵌入现有其他主要大语言模型之中,为思想政治教育营造良好的平台环境。二是探索建立全流程覆盖的意识形态监管机制,确保人工智能赋能思想政治教育的各个环节符合社会主流价值观的要求。人工智能技术的价值特性具有明显的可塑性,很大程度上能够实现与特定价值对齐的现实可行性。因此,在人工智能体的设计过程中,要坚持基于整体观和系统观的价值对齐原则,将社会主流价值内容和相关价值原则有效嵌入人工智能的各个环节,包括数据采集和处理、机器学习、算法设计、内容生成等诸多方面,并在事前、事中和事后建立系统敏捷的监督、反馈和纠错机制。同时,建立针对思想政治教育的智能教育技术伦理审查委员会,制定包含数据采集、算法应用、人机交互等环节的负面清单制度,定期开展技术应用的伦理影响评估和实践指导工作。

(二)加强以创新能力为导向的新型教育机制建设

创新能力越来越成为衡量人才价值的重要标准,人类独有的高阶创新能力是判定人机本质差异的重要方面,而人工智能的无限知识生产能力和有限的创新能力形成鲜明对比。在人工智能技术深度重塑思想政治教育的时代,需要确立辩证的技术认知框架,既要善用智能技术的工具理性提升创新效率,更要警惕对技术过度依赖导致的主体性消解风险。因此,传统创新方式在技术的冲击下面临转型要求,必须加强基于人机交互的新型创新能力建设,特别是要坚持以创新精神为导向,在认知和实践层面强化马克思主义创新观教育以及完善相应的创新评价标准体系,并通过一系列矫正手段来纠偏技术成瘾带来的主体创新能力退化现象和风险。一是回归马克思主义技术批判的根本立场,既要善用智能技术增强教育效能,又要通过制度设计守护认知自主性,在技术系统中嵌入必要的反思性设计,保持主体批判性思维的持续性再生产,构建人本导向的数智化教育生态。二是开展基于通识性人工智能技术运用的教育培训工作,着力提升教育对象的智能素养。人工智能素养主要由选择运用人工智能的意识、灵活操作人工智能的能力和规范使用人工智能的责任三大部分组成。^[24]在实践中,应将塑造新型创新精神的目标贯穿于“思想政治教育+人工智能”的教育培训工作当中,把责任伦理教育放在重要位置,培养学生运用智能技术进行创新的综合能力,特别是智能工具辅助下的高阶创新能力。三是强化以创新能力为核心价值导向的教育考核评价机制。教育考核评价作为思想政治教育培养创新人才的重要手段,在教育过程中起着关键的导向功能和激励作用。人工智能的低阶创新特性决定了激发主体创新活力和提升创新层级的必要性,现实中应该注重创新在教育考核评价中的权重,围绕创新思维和创新实践设置关键性评价目标、指标、方法、记录和反馈,由此在创新能力培养与考核评价之间实现良性循环。四是构建基于智能技术辅助主体创新的新型文化生态,围绕创新原则和创新规范凝聚伦理向心力,着重培育主体的创新意识、创新能力和创新责任。在智能技术深度赋能教育的时代背景下,教育机制的创新正在从单一

的知识传授转向以创新能力为核心的综合素养培育,构建以创新为核心的新型文化生态已成为教育现代化的关键命题。实践中,应以培养教育对象的高阶创新能力为核心要务,加快推进基于人机协同创新能力的创新文化建设,通过重塑创新理念、创新原则及相应的伦理规范,在教育领域乃至全社会形成广泛性的创新文化氛围和创新风尚,以此推动主体创新意识、能力与责任等创新素养的全面跃升。

(三)改善智能辅助思想政治教育的人文环境

习近平指出:“思想政治工作从根本上说是做人的工作,必须围绕学生、关照学生、服务学生,不断提高学生思想水平、政治觉悟、道德品质、文化素养,让学生成为全面发展的人才。”^{[25][46]}思想政治教育承载着思想政治工作的核心任务,在教育实践中专注于“做人的精神世界塑造和精气神提振的工作”^[26],强调对人的精神世界和灵魂的改造,从而突出人文性特点。在教育实践中,必须坚持以人为本的人文价值理念,强化数字化和智能化的人文关怀和人本设计,合理融合智能技术的工具理性和思想政治教育的价值理性,让生成式人工智能在赋能思想政治教育的过程中,既有精度、广度和效度,又能体现温度、情感和情怀。一是探索将基于人文价值的情感智能嵌入智能技术的有效路径,着力提升情感计算、情感感知、情感识别、情感模拟与表达的综合技术能力,注重虚实结合、人机互补、场景沉浸、情理融通、话语创新和知行合一的跨模态和多介质手段的综合运用,不断完善人机之间基于高效智能化共情的情感交互功能,推动实现思想政治教育智能化的人文价值和人文关怀。根据思想政治教育的育人规律,将虚拟现实(VR)、增强现实(AR)、混合现实(MR)、扩展现实(XR)等技术融入智能设计之中,打造全景再现和情景交融的教育叙事场景。例如,通过构建智能教育人文关怀机制,将多模态情感识别和调节的技术设计有效嵌入人机互动的全过程之中,防止主体对技术过度依赖导致的情感异化。二是探索思想政治教育场景下良性的人机交互模式,确保教育者和受教育者的教育主体地位。生成式人工智能技术赋能思想政治教育,应当是人类主导下的良性人机协同

教育模式,而非机器越位下的人际关系异化的教育模式。在实践中,要积极倡导人的主体性的回归,充分发挥教师作为人文情感教育的主力军作用,构建基于“师—机—生”三元主体的新型良性交互模式,让教育和教师的人文温度与机器的工作效度有机融合,用人的智慧和情感引领智能技术更好地服务于思想政治教育事业。

(四)建立和完善基于智能场景的道德责任教育机制

人工智能技术深度嵌入思想政治教育引发的道德失范现象,本质上是技术理性对价值理性的僭越所导致的人文价值危机,建立与智能场景相适配的道德责任教育机制成为破解这一难题的有效途径。一是健全适应数智时代的道德责任教育体系。实施人工智能素养教育工程,构建符合数智化时代特征的道德责任培养模式,专门设置人工智能责任伦理课程,培养和强化学生运用人工智能的责任伦理意识。运用数字孪生技术开展镜像训练,通过虚拟场景中的道德决策模拟与现实行为反馈的对照分析,强化虚拟和现实责任的同一性认知,同时增强人机协同决策训练,设计具有道德困境特征的智能交互场景,培养学生在技术辅助下的自主责任控制能力。依托虚拟现实、增强现实等前沿技术,开发基于多模态技术的沉浸式和高仿真的思想政治教育教学系统,使受教育者能够在逼真的智能交互环境中体验道德困境,通过反复实践不断强化主体的道德判断力和责任意识。二是构建分布式责任网络体系。一方面,建立算法责任溯源制度,在算法开发阶段嵌入伦理影响评估模块,实现从代码编写到教育应用的全流程责任标记。另一方面,构建多主体责任共担机制和完善道德责任评估体系,以分布式的责任网络体系明确开发者、教育者、管理者和学习者的责任边界,并按此开发包含技术伦理、教育伦理、学术伦理在内的多维度评估指标,将评估结果纳入教育质量监测体系。三是建立基于智能场景的新型诚信教育机制。充分发挥思想政治教育的引领性作用,从伦理教育、制度规范、监督手段和文化建设等方面对诚信危机进行有效纾解,培养学生使用人工智能技术的诚信意识和责任感,强化道德判断能力和自律能力;完善科研规范机制,明确人工智能在知识生产和应用中的伦

理边界,建立有针对性的学术不端惩罚机制;提升以技术反制智能技术作恶的能力,开发针对人工智能的新式学术不端检测技术推广应用;加强智能教育的诚信文化建设,营造诚信伦理教育新风尚。

参考文献:

- [1] 习近平.在中国科学院第二十次院士大会、中国工程院第十五次院士大会、中国科协第十次全国代表大会上的讲话[N].人民日报,2021-05-29.
- [2] 马克思恩格斯全集(第42卷)[M].北京:人民出版社,1979.
- [3] 和学新.主体性的内涵、结构及其存在形态与主体性教育[J].西南师范大学学报(人文社会科学版),2005,(1).
- [4] 叶俊民,周进,李超.情感计算教育应用的多维透视[J].开放教育研究,2020,(6).
- [5] 冯建军.主体教育理论:从主体性到主体间性[J].华中师范大学学报(人文社会科学版),2006,(1).
- [6] [加拿大]马歇尔·麦克卢汉.理解媒介:论人的延伸[M].何道宽,译.南京:译林出版社,2019.
- [7] 巫娜.“生成式人工智能+教育”的伪主体间性及其风险[J].现代大学教育,2024,(4).
- [8] 韩敏,赵海明.智能时代身体主体性的颠覆与重构——兼论人类与人工智能的主体间性[J].西南民族大学学报(人文社科版),2020,(5).
- [9] 苏慧丽,张敬威.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,2024,(1).
- [10] 轩宣.论思想政治教育技术化生存[J].思想理论教育,2024,(8).
- [11] 靳凤林.领导干部的政治品德[N].光明日报,2018-04-23.
- [12] 李基礼.基于功能系统分析法的思想政治教育现代化探究[J].思想教育研究,2022,(6).
- [13] 习近平谈治国理政(第2卷)[M].北京:外文出版社,2017.
- [14] 闫瑞峰.算法设计伦理治理的立场、争论与对策[J].自然辩证法通讯,2023,(6).
- [15] 谢俊,吴阳琴.算法推荐下网络主流意识形态面临的风险及防范策略[J].自然辩证法研究,2023,(10).
- [16] 张华平,李林翰,李春锦.ChatGPT中文性能测评与风险应对[J].数据分析与知识发现,2023,(3).
- [17] 蔡峋.创造性人才培养与思想政治教育[J].教育发展研究,2001,(5).
- [18] 张晓平.大学生创造性思维培养离不开思想政治教育[J].当代教育论坛(宏观教育研究),2008,(9).
- [19] 肖峰.生成式人工智能与知识生产新形态——关于三阶知识生产的探析[J].学术研究,2023,(10).
- [20] 殷杰.人工智能如何影响“人性”[N].光明日报,2024-10-30.
- [21] 董春雨.从机器认识的不透明性看人工智能的本质及其限度[J].中国社会科学,2023,(5).
- [22] 冯雨奂.ChatGPT在教育领域的应用价值、潜在伦理风险与治理路径[J].思想理论教育,2023,(4).
- [23] 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[N].人民日报,2023-05-30.
- [24] 程琼,刘宏达.基于生成式人工智能的思想政治教育场景构建及其风险防范[J].国家教育行政学院学报,2024,(8).
- [25] 习近平.论教育[M].北京:中央文献出版社,2024.
- [26] 骆郁廷.论思想政治教育的普遍贯通[J].马克思主义研究,2022,(9).

[责任编辑 顾良艳]