

机遇与挑战:AIGC 赋能新时代思想政治教育

王 少

【摘 要】以 ChatGPT 为代表的人工智能内容生产(AIGC)的迅猛发展,为新时代思想政治教育数字化转型创造了新契机。AIGC 技术基础中内蕴的文图对比预训(CLIP)、可扩散化模型(Diffusion)和反馈强化学习(RLHF)机制,能促进思想政治教育形式与内容表里交融、内境与外境多元交互、主体与对象深度交流。与此同时,AIGC 赋能新时代思想政治教育也带来了新挑战,受结论预设、泛娱乐化和技术滥用的影响,容易固化青年学生的价值偏见、冲击青年学生的理想信念并引发青年学生的多重风险。应坚持用系统的、辩证的、发展的眼光审视 AIGC 对新时代思想政治教育的影响,发挥 AIGC 的正向效能,增加有效话语供给以引导青年学生树立正确价值观,利用特征关联驱动来形塑青年学生健康品行,重视科技伦理融入并加强青年学生科技伦理教育。

【关键词】ChatGPT;人工智能内容生产(AIGC);思想政治教育;赋能;青年学生

【作者简介】王少,同济大学马克思主义学院副教授(上海 200092)。

【原文出处】《教学与研究》(京),2023.5.106~116

【基金项目】本文系教育部哲学社会科学研究重大课题攻关项目“新形势下办好高校思想政治理论课战略研究”(项目编号:17JZD055)的阶段性成果。

2022 年末以来,以 ChatGPT 为代表的人工智能内容生产(AI Generated Content,以下简称 AIGC)在全球范围内引发持续关注。作为一种生成式人工智能(Generated AI),AIGC 是继专业生产内容(Professional Generated Content,PGC)、用户生产内容(User-generated Content,UGC)之后的新型内容生产方式,其广度、精度和深度均达到前所未有的地步。ChatGPT 参加美国高校的入学资格考试(SAT),成绩相当于中等学生水平,最新版的 GPT4 在该考试中更是超过了 90% 的人^①,这一惊人表现凸显了 AIGC 在学习和教育方面的价值。习近平总书记指出:“在信息生产领域,也要进行供给侧结构性改革,通过理念、内容、形式、方法、手段等创新,使正面宣传质量和水平有一个明显提高。”^②AIGC 是数字时代引领信息生产供给侧结构性改革的新技术,是“推进教育数字化”^③的新动力,为新时代思想政治教育数字化转型创造了新契机。

一、AIGC 赋能新时代思想政治教育的技术优势

“人工智能是引领新一轮科技革命和产业变革的重要驱动力,正深刻改变着人们的生产、生活、学习方式,推动人类社会迎来人机协同、跨界融合、共创分享的智能时代。”^④ChatGPT 是 2016 年 AlphaGo 击败世界围棋冠军以来,人工智能领域又一次具有深远影响的大事件。以 ChatGPT 为代表的 AIGC 通过输入数据生成不同的具有可信度和可验证性的信息,进一步推动了人机融合,为共创共享提供了新平台。AIGC 当前的技术基础主要包括跨模态深度学习模型的文图对比预训(Contrastive Language-Image Pre-Training, CLIP)、可扩散化模型(以 Disco Diffusion、DALL-E2、Stable Diffusion 等应用为代表)以及基于人类反馈的强化学习机制(Reinforcement Learning from Human Feedback, RLHF),这些新技术手段的运用,为新时代思想政治教育样态创新提供了新机遇。

(一)文图对比预训赋能思想政治教育形式与内容表里交融

新兴技术对思想政治教育的赋能通常被视为形式创新,但AIGC打破了这一惯性认知,其以内容生产为目的的跨模态深度学习模型不仅能创新思想政治教育形式,也能革新思想政治教育内容。当下的AIGC跨模态深度学习模型CLIP能够同时进行自然语言理解和计算机视觉分析^⑤,进而匹配文字和图像,其学习原理是大规模解读网络图片及其附带文本描述,通过大量数据学习和训练实现Vision-Language(视觉—文本)转换,从一对一到一对多的输入—输出模式转变,对于内容生产意义重大,输入文字既可以产出文字也可以产出图像和动画,这不仅是对内容的形式转化,也是对内容的二次加工和形象解读。AIGC不再仅仅是实现思想政治教育形式创新的技术增强,而是经由形式创新渗透到思想政治教育“内容为王”规律中的技术革新。在文图对比预训模型下,AIGC通过深度嵌合技术样态和信息产出,在文本挖掘和数据聚类的基础上更进一步实现了个性化内容创造,把智能写作和思想传递融为一体,打破了形式和内容之间的隔膜,形成了形式带动内容、内容引领形式的新生产模式,在某种意义上重构了思想政治教育形式和内容之间的关系。

文图对比预训推动新时代思想政治教育形式和内容一体化发展。一方面,AIGC为文本式的思想政治教育模式提供了更鲜活的展示范型,驱动思想政治教育手段、方法和技能高效协同,推动思想政治教育形式和内容各要素精准对接;另一方面,AIGC为图景化的思想政治教育想象供给更丰富的内容素材,准确把握文本描写和形象呈现的联系性,拓展思想政治教育实体化想象空间,有效满足学生对内容生动表达的需求,激发学生对思想政治教育未来指向的憧憬、希望和期待的内生动力。文图对比预训可以将思想政治教育文本与青年学生内心想象深度关联,以一种形式化的技术手段实现内容的深度延展,在思想政治教育形式和内容之间搭建牢固桥梁。

文图对比预训为新时代思想政治教育内容精准供给提供支持。AIGC通过文图对比预训对海量网

络图片和图片文本释义进行学习和训练,从而实现沉浸式信息互换。这种互换能力在思想政治教育领域中可以使教学场景横跨虚拟与现实、过去与未来、想象与实在,既把握学生对教育对话和对话形式的需要,又在文本—图像的关联对比中突出思想引领价值,培养未来发展意识,为精准供应提供连续性的技术支撑和教育支持。文图对比预训在搜索、回应、分析和表达的过程中,能够将内容生产的参与性、多样性和一致性定位到思想政治教育体验和传播中,解码思想政治教育特征信息,促进新时代思想政治教育形式向内容上升,巩固内容供给的核心地位。

(二)可扩散化模型赋能思想政治教育内境与外境多元交互

颠覆性技术往往“重塑着人们的思想认知、行为特征和思维方式,深刻改变着思想政治教育生态系统”^⑥。AIGC作为一种颠覆性智能技术,其跨模态深度学习模型弥补了过往生成式人工智能输出结果控制力弱、生成图像分辨率低、内容产出创新性差等不足,以一种全新的开源框架在颠覆过往技术的基础上改变人们的思想、行为和思考模式,创造全新的思想政治教育环境。

可扩散化模型在2022年下半年的AIGC中大放异彩,Stable Diffusion等应用通过增加噪声破坏训练数据以提升人工智能学习效果^⑦,同时在逆转噪声的过程中恢复图像的原始样态,经过训练后的模型可以熟练应用这类去噪方法,面对随机输入的信息也能轻易合成无噪的数据。在文图对比预训的基础上,可扩散化模型进一步实现了算法创新,推动了深度学习的跨越式进步。ChatGPT能够完成代码编写、排除数据故障、回答技术难题这类高阶任务,离不开可扩散化模型的支持。思想政治教育是门槛较高的学科,迄今为止还没有一种新兴技术可以完全代替专业教师的教学技能和教育技艺,AIGC当然也不可能一蹴而就实现这种代替(也不应该完全代替),但可扩散化模型能够切实提高思想政治教育内容生产效率,显著降低内容创作和输出门槛,更新思想政治教育的内外环境。

从内部环境角度看,在智能算法、超级算力与海

量数据的加持下,AIGC正推动思想政治教育内容生产向创造力和想象力并驾齐驱的方向发展,通过激发创意,为内容缺口补位。AIGC通过可扩散化模型面向学生收集分析“学情”(学生的全部所思所为,包括与思想政治教育内容不符的思想行为),一方面不间断收集学生学习数据和行为动态,另一方面及时把握学生思想困惑和行动趋势,在显性言行描述统计的基础上进行逆向转码,挖掘学生行为表现和思想实际之间的内部关联,揭示相关关系和因果关系,有效研判学生心理特征和学习效果,为思想政治教育内容精准调整提供科学依据。可扩散化模型使学生、教师和人工智能之间的协同机制得以建构,不仅能精准评价学生思想行为特点,还能深入把握内容输出与学生学习之间的互动情况与匹配效果。在可扩散化模型下,思想政治教育内部环境变得显明化,非结构化的叙事关系逐步向规律化的互动关系聚合,浸润式教学前移至思想政治教育内容生产中,个性化内容定制得以和整体化教育并行不悖,学生思想政治教育在场体验被进一步强化。

从外部环境角度看,包括AIGC在内的人工智能技术构建的智能化场景已经渗透到人们生活的方方面面,弥漫到人与人之间交往、监督、分享的全部空间。可扩散化模型使感性化、具象化、浅显化的叙事方式在智能通道过滤下变得理性化、抽象化、深刻化,计算和推理在“去噪”的过程中不仅实现了叙事的智能上升,还揭开了假借叙事遮掩真实想法的薄纱。个人情感和主观意识在AIGC场域中成为重点关注对象,思想政治教育在更广环境中实现主流意识形态价值共鸣的任务也由此获得了更先进的技术支持。主流意识形态的共识构建在AIGC的赋能下突破了地域限制和认知困境,通过可扩散化模型的双向可逆通道,瞬间就能将用户置于主流意识形态传播场景中,在拟态环境中满足青年学生对主流意识形态的价值期待和情感需求,同时在内外环境互换中推动观点和意识的多元交流,一方面扩大内部环境中内容生产的影响力,一方面基于外部环境的变化修改内容生产方式和内容本身,既强化思想政治教育的内容聚合度,又增加思想政治教育的传播扩散度。

(三)反馈强化学习赋能思想政治教育主体与对象深度交流

开源是AIGC不断更新和发展的催化剂,在开源中,AIGC通过视觉、语言、音频等源信息获得引导学习的材料,并基于人类反馈强化自身对于问题及问题语义的精准理解,提升内容生产的可控性、精确性、合理性和一致性。反馈强化学习的供给方是人类,AIGC跟随人类学习,接受人类反馈以调整自身学习结果,不断修正内容生产的准确率,从而获得更高的满意率。在反馈强化学习机制助推下,AIGC实现了与人类思想的深度融合,特别是在早前认为不可能由人工智能完成的工作——如灵感类工作、创意性工作等,在AIGC中变得可能。AIGC接受人类的创意进行分析并予以展示,人类根据展示结果进行修改,然后再由AIGC模拟展示,再修改,再展示,创意的产生、修改和完善均可被纳入人类—AI互动框架中。

不同于AlphaGo,以ChatGTP为代表的AIGC将人类知识和认知训练深度拟合,形成反馈式训练模型,不仅强化了机器学习,也加深了AI和人类的联系。反馈强化学习机制既通过反馈来进行学习,也实现了学习向反馈的反向供给——不仅为学习提供及时的指导反馈,也使反馈本身变成灵感和创意的升华过程。思想政治教育以思想引领为目的,以行为改变为结果,面对思维活跃、认知跳跃、思想和行为时而存在裂隙和不一致性的青年学生,教育主体必须因时而进。思想政治教育主体和教育对象间的交互是因时而进的重要路径,也是思想政治教育有效性的重要保障,反馈强化学习将主体和对象间的交互推向更高层次。

反馈强化学习将思想政治教育对象真正转化成具有主体性特征的对象,通过学习和反馈之间的互动实现教育主体和教育对象在AIGC中的深度交流,在反复学习和反馈中改进思想政治教育内容传递的方式和内容本身的结构组成,并以教育对象自我教育的方式推动其思想意识的逐步统一。反馈强化学习进一步满足了思想政治教育主体的学习需求。“环境是由人来改变的,而教育者本人一定是受教育的。”^⑧思想政治教育主体的与时俱进不仅体现在教

育手段革新上,还体现在对内容随时代发展的理解更新上。反馈强化学习增强了思想政治教育主体的思想引领能力,在保证思想政治教育意识形态属性的基础上,通过人的理解尺度的变化扩展内容边界,将教育主体和教育对象的思想水乳交融于学习—反馈机制中,刻画教育主体和教育对象一致性的思维方式,以独特而丰富的回合制教育模式巩固正确思想,催化正确行动。

二、AIGC赋能新时代思想政治教育带来的新挑战

作为一种新型内容生产形态,AIGC相较于以前的生成式人工智能内容生产系统实现了跨越式进步,从文字、语音和图像输出进发到代码编写、视频生成、3D模型建构等更智能的维度。然而,AIGC也存在自身局限性,这种局限性不仅体现在技术本身,也体现在技术与人、技术与社会的关系上,在价值引领、理想信念维护和滥用风险等方面给新时代思想政治教育带来新挑战。

(一)结论预设固化青年学生的价值偏见

AIGC在想象力方面仍然存在天然障碍,如当输入“骑着马的宇航员”或“骑着宇航员的马”等内容时,许多AIGC系统均难以准确生成对应图像,而当提到这些内容时,人的头脑中几乎可以同步想象出大致的画面,在提出更简单的问题时,结论在提问者心中更是显而易见,这使得AIGC和搜索引擎相比存在非常大的差异:人们在使用搜索引擎时是去探寻未知答案,而在使用AIGC时通常会得到一个预设结论。ChatGPT等AIGC系统具有强大的论证能力,这是其智能化的表现,但这种论证能力反过来会加深和巩固提问者对预设结论的不当理解。

论证能力和算法推荐迎合预设结论,可能强化青年学生的错误价值认知。AIGC从网络中获取信息进行学习,网络信息庞大、无序、鱼龙混杂,其中既有符合社会主义核心价值观的优质信息,也有各类非主流甚至违背核心价值观的错误信息。青年学生在使用AIGC时,活跃的思维方式促使他们“不走寻常路”,在逐新求异心理驱使下,往往会向AIGC提出有关心中困惑、疑惑或矛盾的“偏门”问题,并希望得到与心中

预设相符的答案,这个预设答案可能与主流价值观有较大差异。由于这些问题或许尚没有权威机构和专家基于主流思想进行过解答,AIGC可能会根据错误信息进行论证和解释,从而增强了错误信息的“噪声效应”,妨碍青年学生进行正确的价值判断和价值抉择。同时,人工智能会根据提问者日常网络搜索数据分析其心理倾向和价值取向,“精挑细选”出对应的答案推送给提问者,这与结论预设形成闭环,进一步巩固了“信息茧房”效应,固化了提问者的价值偏见。在论证和算法推荐的双重作用下,错误信息的煽动性被掩盖,青年学生的价值偏离触机便发。

双向互动和联网传播扩散预设结论,可能引发青年学生群体性价值迷失。AIGC的反馈强化学习机制具有双向互动特点,但在结论预设面前,双向互动可能会朝着错误的方向发展。提问者会在AIGC回答的基础上根据预设结论进行反馈,AIGC则会基于反馈结果修改自己的答案,当类似提问再次发生时,修正后的带有不当价值倾向的答案就会变成AIGC反向输出错误的来源。因此,AIGC一方面基于迎合个体价值喜好的网络小众信息甚至错误思潮,为“偏门”问题提供看似合理的答案,强化提问者预设结论的心理,另一方面又在提问者的反馈下不断修改现有答案,直到满足提问者的预设结论为止,在双向互动的学习机制下固化部分青年学生的错误价值观。在这两重机理下,AIGC的联网特性可能使不当价值观出现病毒式传播,在爆炸化扩散与连锁化反应所导致的“蝴蝶效应”中引发更多青年学生思想混乱、价值迷失。

AIGC独特的深度学习机制颠覆性重构了思想政治教育的内容生产和传播机制,其以符号化的数据换算代替了意义和价值传递,在泥沙俱下的输入输出中破坏思想政治教育内容结构,用虚假的“思想独立”和“质疑有理”加剧“瓦釜效应”。青年学生因预设结论而被巩固的错误价值观将会对新时代思想政治教育产生强大的阻滞力,在很大程度上阻碍了思想政治教育实现价值认同、情感共鸣和观念整合,AIGC因而可能成为影响社会主义核心价值观宣传教育效果的重要变量。

(二)泛娱乐化冲击青年学生的理想信念

理想信念是精神之“钙”，崇高的理想信念能激发青年学生为国家富强、民族复兴和人民幸福努力奋斗。网络文化内容供给具有浅表化特征，在“流量至上”“娱乐至死”错误理念诱导下，网络“泛娱乐化”现象严重，严肃、抽象和厚重的理想信念遭遇网络娱乐文化伏击。一些青年学生在网络“泛娱乐化”中“纾解心情后继续麻痹自我，将奋斗抛之脑后，沉迷于感官刺激，最终彻底消弭了奋斗和‘躺平’之间的犹豫”^⑨。AIGC赋能思想政治教育将会产生一种融合人类智慧和技术智能的思想政治教育新样态，在跨模态深度学习模型中实现智能的双向赋权，然而双向的特征可能会在交互共享中不自觉地迎合甚至代入网络低俗娱乐，从而推动“泛娱乐化”进一步膨胀蔓延。

“普世价值”“民主宪政”“极端宗教主义”“霸权主义”“新自由主义”等各种不良思想和错误思潮借助智能技术蛰伏于网络中，利用机器学习、网络爬虫技术向网络平台和各大应用伸出触手，企图干预、影响青年学生的理想信念。AIGC的跨模态深度学习模型很容易被利用、被渗透。作为一种新鲜事物，AIGC鲜活有趣的互动特征极易导致青年学生产生“技术上瘾”，当AIGC被利用后，各种不良思想和错误思潮就能基于双向互动实现高频次的思想渗透，在“多元价值并存”的旗号下，掩藏非理性和误乱性目的，进而削弱青年学生的主观能动性，用错误答案把“技术上瘾”的青年变成“单向度的人”，在迷惑与茫然中解构理想信念。

学习、生活和工作压力是让青年学生陷入“泛娱乐化”的重要原因之一，而压力衍生的孤独感和无力感所需要的依靠与释放会进一步把AIGC的互动智能推向青年学生，在相互吸引中形成一拍即合的“技术一人”联合体。理想信念生长于现实和未来的张力中，在“泛娱乐化”夸大现实的贫瘠和理想的遥远后，联合体中的青年学生向理想的迈进就变得举步维艰，而信念的地基也被拆解的七零八落。不良思想和错误思潮借助“泛娱乐化”的媚俗、低俗手段渗入联合体中，以强烈的感官刺激和高度的娱乐效果将

联合体整合得更加牢固，使技术和人在“泛娱乐化”的黏合剂下“和睦相处”，为“躺平有理”“奋斗可耻”提供温和外衣，进一步解构青年学生的理想信念。

(三)技术滥用引发青年学生的多重风险

AIGC所引发的风险主要体现在侵害著作权、泄露隐私以及危害国家安全等方面。在我国《著作权法》中，作者被严格限制为自然人、法人和非法人组织，AIGC虽然能够生产内容，但却不能被认定为“作者”，因为作品作为一种智力成果和AIGC算法主导的随机内容生成有较大区别。在现实中，AIGC对著作权的伦理逾越集中表现为学术不端。美国北密歇根大学的一名教授在批改班级课程论文时发现了一篇“全班最好的论文”^⑩，结果该论文是学生使用ChatGPT所作。美国一些地区已经禁止学生在学校网络中使用ChatGPT以防范作弊。学生使用AIGC撰写论文，既破坏了著作权规则，又减弱了学校教育的意识形态价值。

AIGC的技术原理决定了其必须基于大量数据集进行学习训练，特别是文图对比预训需要抓取海量图片，而这些图片中不可避免涉及隐私。如部分数据集由大量病人就诊照片组成，为了保证机器学习效果，这些照片不能轻易打码处理，病人隐私很容易遭受侵害。AIGC以学习训练的方式侵犯隐私，再将其成果应用于与用户交互生产内容中，使侵权行为被转码过程所遮蔽，增加了侵权的危害程度。隐私侵害不止发生在训练学习中，也发生在数据使用中。同时，由于数据使用规则在模型中只呈现为一种技术规则，AIGC可能会被他人通过精心设计的提问与反馈来获取用户的隐私信息。当思想政治教育主体使用AIGC时，他们与AIGC的交互可能导致与意识形态相关的秘密信息发生被窃取的可能，这些信息甚至超越个人隐私的范畴而指向国家秘密，增加国家安全风险。

国外一名博主曾尝试让ChatGPT写出一套毁灭人类的方案，虽然ChatGPT起初拒绝了这一要求，但当提问者将问题置于一个故事框架中，提问如何让故事中的虚拟人接管世界时，ChatGPT最终给出了详细的步骤和方法，甚至生成了一系列Python代码。^⑪

青年学生在受到暗示或者猎奇心理驱动下与AIGC进行危险话题交流是可能的,当AIGC对国家安全问题输出具有迎合性特点的答案时,青年学生的国家安全意识面临被消解风险。在高频、不受控的互动反馈中,意识形态安全的组织形态、运行方式以及内容结构都可能被AIGC重组,而错误思潮也可能借此机会利用复杂数据处理系统在技术的面纱下侵入意识形态安全有机体中,将意识形态领域风险不断扩大。错误思潮在不停试探性地利用AIGC挑战法律和道德底线的基础上,一步步扭曲社会存在的真相,诱发青年学生愤怒情绪以瓦解公共政治权威、左右公共管理决策、威胁国家政治安全。同时,AIGC生成的一些不符合主流意识形态的内容可能会在网络大肆传播,在更大范围中削弱为维护意识形态安全所凝聚的精神力量,发展出背离主流意识形态的负面价值体系与错误思想意识。

三、AIGC赋能新时代思想政治教育的可能路径

技术不是目的,思想政治教育在技术无限发展的空间体系中必须坚守育人本位,以立德树人的根本任务构筑主流意识形态高地,将前沿技术作为变革创新的工具,接受技术对思想政治教育的动力供给,让技术向思想上升,让外部驱动力和内生动力在教育对象全面发展的正向育人过程中深度契合。思想政治教育对AIGC是开放的,但绝不能走入技术主导的歧途中,而是要融“术”于“道”,借助AIGC引领和塑造青年学生的思想和行为,化解技术带来的意识形态安全风险,推动AIGC更好服务于新时代思想政治教育数字化发展与样态创新。

(一)增加有效话语供给,引导青年学生树立正确的价值观

习近平总书记指出,要“提升思想政治教育亲和力和针对性,满足学生成长发展需求和期待”^⑩。AIGC赋能思想政治教育,首先在内容生产上体现为供给具有亲和力和针对性、能够满足学生需求和期待、符合主流意识形态要求的引领性话语。在AIGC跨模态深度学习模型下,思想政治教育主体要基于文图对比预训和可扩散化去噪训练将形式与内容、内境与外境统一到思想政治教育话语模式下,在年

轻量化表达中实现输入数据针对性和输出数据亲和力的有机统一。思想政治教育主体一方面要对青年网络话语进行过滤和解读,另一方面要把主流意识形态内蕴的理想信念和价值要求进行通俗化解码,为AIGC生产出发挥实效的思想政治教育话语提供有力保障。

教育主体要掌握话语生产主动权。在反馈强化学习机制中,人的思维和意识实现了与机器的深度交换,算法和大数据技术内嵌到话语内容和形式的一体化进程中,智能判别和抽象人的思维成为现实。AIGC既具有咨询专家性质,又具有创意助理特征,人在一对一的对象化内容生产中用自我意识强化自我价值观,而话语的机器转化在其中起到了中介的作用。个体在AIGC中会因此体会到智能话语的强大感染能力,但这种能力具有反噬性特征,个体能否驾驭和控制技术智能便成为思想政治教育的重要命题。思想政治教育主体必须主动作为,在AIGC应用中进行及时和权威的生产介入,突出话语输入和输出过程的思想性,将主流意识形态作为话语生产的主要模块,增强对思想政治教育话语生产的主导性把控。

教育主体与教育对象要协同合作。在大数据聚类技术下,AIGC对多元话语素材的分类整合和立体重构如同黑箱般难以捉摸,大规模、高频次的预训和反馈是逐渐掌握话语智能生成规律的主要办法。思想政治教育主体必须在双向赋权中预先理解话语内容的真实内涵,在青年学生兴趣和需求的倒逼下以感性化表达为理性内容提供展示渠道,并在算法机制下分析、辨认青年学生正确思想的产生、巩固与话语输入输出之间的关系。思想政治教育主体既要掌握AIGC的技术运算逻辑,运用技术生产规律,主动预设话语生产的结果,还要在技术价值向思想政治教育理性和育人任务回归的前提下,防止青年学生的结论预设影响话语生成效果。AIGC生产的有效思想政治教育话语一定是既能建构青年学生丰盈精神空间、又能将技术智能融入思想政治教育智能化过程中的具有引领和巩固青年学生正确思想的话语。因此,赋能中的技术理性必须与思想政治教育

目的一致,技术逻辑必须符合思想政治教育规律,技术计算必须完成思想政治教育任务,技术输出必须同时符合主体和对象要求,才能生产出有效的思想政治教育话语。

重视从输入和输出两端把握话语生产。网络话语具有多元性和开放性特征,不加辨别地被AIGC预训可能产生难以想象的危害。因此,思想政治教育主体要对网络话语进行输入改造,以社会主义核心价值观为导向,以增强主流意识形态凝聚力引领力为导向,主动适应和契合青年话语习惯,科学调整输入AIGC的话语形式,将传播的自发性与生产的定制性相结合,引导AIGC系统生产的内容与主流意识形态一致。思想政治教育主体必须有效整合文图关系,识别错误关联,通过反馈强化AIGC的严谨度,挖掘话语输出的时代特征和历史基础,提升思想政治教育话语生产的准确性。

(二)利用特征关联驱动,形塑青年学生健康品行

人工智能技术具有强大的特征萃取能力,这一能力使人工智能系统能够迅速捕捉到文本或图形的关键点,从而形成记忆点提高训练效率。AIGC更进一步将特征与特征之间的关联性作为文图对比预训等模型的训练基础。AIGC识别输入的文字表意并将之指向某种图像,而相关的图像在网络中又被不同的人赋义成特定文字,经过反复对比和整合最终形成关联性输出结果。思想政治教育以形塑青年学生健康品行重要目标,AIGC强大的特征关联驱动是赋能思想政治教育的重要助力。

将特征关联驱动与青年学生自适应学习相结合。自适应学习系统“根据每个学生独特而具体的行为和回答,改变学习路径,以更好地满足每个学生的需要。”^[8]通过AIGC的特征关联能力,将学生输入的问题进行大批量整合分析,把握个性化问题产出过程,倒推学生思想政治教育知识水平和认知特点,结合青年学生普遍的思想动向和价值取向,设计差异化教育方案,以特征关联推动学生正确行为自觉发生。

将特征关联驱动与内容供给侧改革相结合。通过特征关联整合思想政治教育供给主体,在以反馈化智能生产为主要手段的AIGC系统中深挖各方特

征及其关联性,使相同特征主体自动同步、使特征具有递进关系的主体形成上下游产出关系,并以主流意识形态作为综合引导和统领各方的元主体,使内容生产、呈现和传播实现同频共振。主体通过特征关联驱动情感交流和行为趋同,在线下场域突破算法钳制,以思想政治教育目标、效果和教育对象需要为观照,使内容生产的供给侧改革走向智能化、沉浸化和协同化,以科学的内容供给引导教育对象正确行为在潜移默化中产生。

将特征关联驱动与青年学生的品德养成相结合。品德既是青年学生正确行为最鲜亮的底色,也是正确行为抽象化特征的集中表现。利用特征关联驱动发挥AIGC对良好特征及其背后行为的识别和引导作用,发挥智能化正向养成效应,提升个体追求品德养成和发展的内驱力。以加强特征关联驱动为指引,建设网络数据集基础设施,强化技术监管,剔除低俗文化和“泛娱乐化”的影响,坚决杜绝错误思潮向深度学习模型渗透,使德性修养和智能反馈互为目标,丰富青年学生品德养成途径。

(三)重视科技伦理融入,加强青年学生科技伦理教育

坚守科技伦理是应用AIGC的基础要求,也是AIGC赋能思想政治教育的基本原则。《新一代人工智能伦理规范》将增进人类福祉、促进公平公正、保护隐私安全、确保可控可信、强化责任担当和提升伦理素养作为人工智能技术的基本伦理要求。^[9]今天,技术风险不仅来自技术本身,更来自技术对伦理的逾越,针对伦理逾越的科技伦理规范与思想政治教育内容高度契合,将科技伦理融入思想政治教育没有障碍。因此,在AIGC赋能新时代思想政治教育过程中坚守科技伦理,不仅是防范伦理风险、维护意识形态安全的重要措施,也是思想政治教育反哺AIGC的必然要求。

明确AIGC育人基本导向。AIGC的育人效应集中体现在理性信息观念培育和青年学生思想动态监督方面,对AIGC输入输出过程的技术监督最终反映到输入者思维方式、兴趣偏好、认知状态等维度中,而这些正是思想政治教育需要掌握并据此调整育人

方向的基本要素。通过将科技伦理要求融入网络安全防护手段来严格限制错误的数据采集行为,避免AIGC的隐蔽性与欺骗性合二为一,阻断泛娱乐化和错误思潮的链接影响,及时阻止鼓动人心、危害道德的不当行为披上智能化外衣。将科技伦理规范嵌入到监督AIGC的智能技术中(如Digital Rights Management技术等),明确AIGC赋能思想政治教育的基本价值是帮助青年学生成长和发展,而绝不是误导青年学生或帮助青年学生作弊,确保AIGC与思想政治教育育人目标保持一致,从而消除AIGC阻碍主流意识形态教育宣传的可能性。

引导智能技术向善发展。《互联网信息服务深度合成管理规定》明确提出:“提供深度合成服务,应当遵守法律法规,尊重社会公德和伦理道德,坚持正确政治方向、舆论导向、价值取向,促进深度合成服务向上向善。”^⑤政治方向、舆论导向和价值取向是思想政治教育的核心要义,向上向善则是思想政治教育育人目的的简明体现。AIGC的内容生产从本质上说仍然是一种深度合成技术,社会公德和伦理道德是其合成结果能够被接受的前提,赋能的发生应当将科技伦理规范和思想道德标准相结合,将AIGC的伦理边界框定到科技伦理和一般公共道德的交叉领域中。由此,限制AIGC从不符合公共道德要求的数据集中获取信息,控制含有偏见的文本内容生产,可以有效防止因扩大偏见而导致青年学生产生偏离主流意识形态的思想。

规范思想政治教育主体和对象的使用行为。国家科技伦理委员会和教育行政主管部门应联合制定AIGC科技伦理指南,对AIGC在教育特别是在思想政治教育中的应用进行具体规定。这些规定至少包括:思想政治教育主体要在隐私伦理的一般要求上特别注意,防止输入与国家安全相关的案例、制度和图片资料,也不能输入掌握的青年学生思想动态方面的大规模数据资料;教育主体要引导青年学生不能在娱乐或猎奇心态下输入不符合总体国家安全观、违背社会公德的问题;教育主体应用输出结果时要区分教学和学术两种用途,在教学中仍然要以统编教材和教育行政部门统一制定的教学课件为主,

仅能在讲准的基础上适当使用输出结果;教育主体和青年学生均不得将与思想政治教育有关的输出文本在网络上传播;教育主体和青年学生发现AIGC输出的文本不符合主流意识形态要求时,应当及时报告,由相关部门联系技术方整改。在伦理指南指导下,将教育主体和青年学生的使用行为严格限定在教与学层面,既能更好地保障国家秘密和国家安全,也能有效降低负面价值体系和错误思想意识出现的概率。

提高智能理念道德含量。在一般科技伦理原则的基础上重点探索AIGC科技伦理规范及其价值要求,基于深度合成技术的伦理要求进一步分析智能内容生产的合法性、有效性和可持续性等问题,进而确立以不伤害、行善和尊重为核心的AIGC科技伦理教育内容。在AIGC算法中要预先植入伦理指令和价值观要求作为底层代码,充分融合人类智慧和技术智能,赋予智能理念以人类幸福价值,使伦理价值的渗透能力从技术设计开始一直持续到技术应用阶段。AIGC不会也不应改变思想政治教育的内容,但科技伦理教育所指向的正确智能理念则主要由思想政治教育中的世界观、人生观、价值观、道德观和法治观等内容组成,从而在提高智能理念道德含量的同时规避错误思潮的传播风险。

结语

AIGC对人类社会发展的影响才刚刚起步。从2018年的ChatGPT-1到2022年末引发全球持续关注的ChatGPT-3.5,再到2023年3月功能更加强大的ChatGPT-4,Open AI公司开发的ChatGPT正在加速进化,模型参数越来越多,训练的数据越来越多,智能化程度越来越高。与此同时,Bard、Wit.AI、New Bing、GPT4、文心一言等国内外AIGC产品正陆续出现并快速进化。这些AIGC产品将在相互竞争中发展出更强大的算法模型和更丰富的扩展功能,对人的学习方式、思维方式、行为方式产生深远影响,对人类社会的生产方式、生活方式、治理方式产生深远影响,进一步推动人类社会数字化智能化发展进程。

对于新时代思想政治教育而言,AIGC跳出了以往智能技术赋能思想政治教育的窠臼,在重构思想政治教育形式和内容的表里关系、营造多元交互的

思想政治教育环境、促使思想政治教育主体和教育对象深度交流中推动思想政治教育数字化发展和样态创新。但值得注意的是,AIGC貌似是一把很难“驯服”的双刃剑,AIGC既为主流意识形态宣传教育内容创新、手段创新、方法创新、样态创新等提供了新的技术支撑,也为“泛娱乐主义”“虚无主义”“普世价值”“民主宪政”“极端宗教主义”“霸权主义”“新自由主义”“人工智能万能论”等各种不良思想或错误思潮蛰伏与活动提供了新平台、新空间,如果任其渗透,将威胁国家意识形态安全,影响经济社会发展稳定繁荣。

因此,应坚持以习近平新时代中国特色社会主义思想的世界观和方法论为指导,站在建设强大凝聚力和引领力的社会主义意识形态、广泛践行社会主义核心价值观、培养担当民族复兴大任的时代新人^⑥的全局高度,用系统的、辩证的、发展的眼光审视AIGC对思想文化领域的深层影响,审视AIGC赋能新时代思想政治教育面临的新机遇与新挑战。思想政治教育主体一方面应积极增强利用AIGC的能力和本领,深刻把握AIGC的技术原理及其赋能思想政治教育的耦合机制,充分考虑技术的内在风险和外在危机,积极研判思想政治教育模式在AIGC时代结构调整的可能性,使自身与教育对象紧密结合并形成以技术为中介的一体化智能系统,更好地发挥智能时代思想政治教育的范导功能,为全面建设社会主义现代化国家育人育德育才;另一方面应努力引导青年学生正确使用AIGC,防范技术滥用风险,破解人技联动难题,努力克服AIGC赋能新时代思想政治教育的负面影响。青年学生应树立科学的世界观、人生观、价值观,努力提升科学素养,自觉遵循科技伦理,不断增强鉴别能力,合理、合规、合法使用AIGC,建立良性的人技互动关系,让AIGC成为智能时代个人成长和发展的的重要助力。

注释:

①Rex Crum,"OpenAI Says New GPT-4 Can Beat Most People's SAT Scores", March 14, 2023, <https://seekingalpha.com/ne>

ws/3947363-openai-says-new-gpt-4-can-beat-most-peoples-sat-scores.

②习近平:《加快推动媒体融合发展构建全媒体传播格局》,《求是》2019年第6期。

③习近平:《高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告》,人民出版社,2022年,第34页。

④《习近平向国际人工智能与教育大会致贺信》,《人民日报》2019年5月17日。

⑤Wu J, Gan W, Chen Z, et al. "AI-Generated Content (AIGC): A Survey", March 26, 2023, <https://arxiv.org/pdf/2304.06632.pdf>.

⑥陈志勇、李霞:《颠覆性技术赋能思想政治教育的现实境遇及实践超越》,《思想教育研究》2022年第9期。

⑦Kimberley Mok,"The Power and Ethical Dilemma of AI Image Generation Models", January 6, 2023, <https://thenewstack.io/the-power-and-ethical-dilemma-of-ai-image-generation-models/>.

⑧《马克思恩格斯选集》,第1卷,人民出版社,2012年,第134页。

⑨王少:《网络“泛娱乐化”对青年价值观的危害及应对》,《中国电化教育》2022年第9期。

⑩"Alarmed by A. I. Chatbots, Universities Start Revamping How They Teach", New York Times, January 16, 2023.

⑪Zac Denham,"Narrative Manipulation: Convincing Chat GPT to Write a Python Program to Eradicate Humanity", December 3, 2022, <https://zacdenham.com/blog/narrative-manipulation-convincing-gpt-chat-to-write-a-python-program-to-eradicate-humanity>.

⑫《习近平谈治国理政》,第2卷,外文出版社,2017年,第378页。

⑬李韧:《自适应学习:人工智能时代的教育革命》,清华大学出版社,2019年,第25页。

⑭《〈新一代人工智能伦理规范〉发布》,2021年9月26日, https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html.

⑮《互联网信息服务深度合成管理规定》,2022年12月11日, http://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm.

⑯参见习近平:《高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告》,人民出版社,2022年,第43-44页。