

Doi:10.11835/j.issn.1008-5831.pj.2023.08.006

欢迎按以下格式引用:代金平,覃杨杨. ChatGPT等生成式人工智能的意识形态风险及其应对[J]. 重庆大学学报(社会科学版),2023(5):101-110. Doi:10.11835/j.issn.1008-5831.pj.2023.08.006.



Citation Format: DAI Jinping, QIN Yangyang. The ideological risks of generative artificial intelligence such as ChatGPT and its response[J]. Journal of Chongqing University (Social Science Edition), 2023(5):101-110. Doi:10.11835/j.issn.1008-5831.pj.2023.08.006.

# ChatGPT等生成式人工智能的意识形态风险及其应对

代金平,覃杨杨

(电子科技大学 马克思主义学院,四川 成都 611731)

**摘要:**意识形态风险防范是事关民族复兴、国家主权、文化建设的核心范畴。在当今人工智能技术蓬勃发展的数字时代,以 ChatGPT 为代表的生成式人工智能技术内嵌渗透到社会生活领域之中,伴随着交互深度与辐射范围的不断扩展,裹挟了众多具有新特点的意识形态风险。清晰明辨 ChatGPT 等生成式人工智能产生意识形态风险的内在机理是防范生成式人工智能意识形态风险的前提。ChatGPT 等生成式人工智能凭借内容生成的首要功能对意识形态的价值观念层进行渗透;依靠交互对话的产品形态对意识形态的话语方式产生影响;利用多维应用的功能属性对社会生活实践领域展开强势介入,促使当下意识形态斗争态势风云诡谲,变化万千。理性审视和精准识别 ChatGPT 等生成式人工智能所诱发的意识形态风险是防范生成式人工智能意识形态风险的关键。从意识形态引领角度出发,ChatGPT 等生成式人工智能的运行机制深蕴的资本逻辑解构社会成员已形成的认知和价值,借由“技术-资本”双重优势导致意识形态面临引领乏力风险;从意识形态话语维度审视,ChatGPT 等生成式人工智能带来的“人机协同”话语生产新模式削弱主流意识形态话语深度,多元的生成式话语输送淡化话语权威,导致意识形态话语减蚀风险;从意识形态管理层面反思,ChatGPT 等生成式人工智能引发的信息传播新范式消解意识形态管理的主导权,及其对人的强化宰制,导致意识形态管理弱化风险。ChatGPT 等生成式人工智能在掌握先进技术基础上诱发了一系列辐射广泛、渗透全面、不易掌控的意识形态风险。如何防范与化解 ChatGPT 等生成式人工智能所带来的意识形态风险,成为当前我国意识形态建设无法回避的新课题。首先,贯彻落实人民至上的价值理念,以马克思主义理论指导意识形态防范工作,推进打造人机和谐、风清气正的网络空间;其次,打造富有中国特色的话语途径,在讲好中国故事的话语基础上破解西方话语的现代困境;最后,积极倡导全球治理并贡献中国方案,以共商共建共享的发展原则,依托人类命运共同体

基金项目:国家社会科学基金重点项目“网络社会治理体系研究”(18AKS015)

作者简介:代金平,哲学博士,二级教授,电子科技大学马克思主义学院博士研究生导师,《重庆邮电大学学报(社会科学版)》主编,重庆自然辩证法研究会副理事长,Email:1605573180@qq.com。

构建负责任的人工智能格局,促使 ChatGPT 等生成式人工智能在现代文明面前彰显光明的未来前景。

**关键词:** ChatGPT;生成式人工智能;人工智能;意识形态风险;风险防范

**中图分类号:** D64; TP18 **文献标志码:** A **文章编号:** 1008-5831(2023)05-0101-10

## 一、研究背景与学术认知

2022年11月,美国人工智能公司 OpenAI 发布的生成式人工智能产品 ChatGPT(Chat Generative Pre-Trained Transformer)引发全球瞩目,它的爆火牵引着既往离散的人工智能发展方向倏然聚拢,类似的生成式人工智能如 Bard、文心一言、通义千语等在短时间内纷纷涌现,勾勒出人类与生成式人工智能多维交往的未来图景。以 ChatGPT 为代表的生成式人工智能在深刻形塑与变革人类生产方式、生活方式和思维方式的同时,又以一种“更加脆弱的隐形意识形态”<sup>[1]</sup>面相出场。党的二十大报告指出,全党必须“主动识变应变求变,主动防范化解风险”<sup>[2]</sup>。在当今国际国内各种思潮交织碰撞、东西方意识形态博弈的背景下,ChatGPT 等生成式人工智能极有可能成为意识形态渗透的工具,成为威胁国家意识形态安全的风险增长点。因此,面对 ChatGPT 等生成式人工智能这一“变量”,前瞻性研判其裹挟的意识形态风险,并谋划有效的应对略策,是防范和化解意识形态领域重大风险的战略需求。

意识形态风险研究是 ChatGPT 等生成式人工智能风险研究的重要视角之一,是相关学者基于 ChatGPT 等生成式人工智能风险研究逐渐延伸细化后出现的,并呈现出逐渐增多的趋势。目前学术界关于 ChatGPT 等生成式人工智能的意识形态风险研究多散见于不同研究领域和理论视角,学者们主要进行了以下几方面研究:一是对意识形态风险诱因展开研究。ChatGPT 等生成式人工智能意识形态风险诱因可大致划分为两个层面:技术本身的问题、技术与人的关系问题<sup>[3]</sup>。具体而言,表征为内置于技术的算法偏见、算法霸权,以及人机关系倒置的认知局限<sup>[4]</sup>。二是对意识形态风险样态进行研究。有学者归纳出 ChatGPT 等生成式人工智能所诱发的意识形态撕裂风险、意识形态操纵风险、意识形态解构风险<sup>[5]</sup>。有学者从教育变革角度出发,认为 ChatGPT 隐藏着“教育价值坍塌、新型数字权威崛起、平台资本的意识殖民化”<sup>[6]</sup>等新型意识形态风险。三是探究意识形态风险防范路径。从宏观层面而言,在应对生成式人工智能的意识形态风险时应遵循鼓励发展、保障安全、技术规制、协同监管等原则<sup>[7]</sup>;从微观层面而言,要防范这些风险,就要主动识别智控技术,积极解码 ChatGPT 介入逻辑的底层算法技术,以价值理性为引导推动完善公共向善技术体系建设等<sup>[8]</sup>。

既有的学术成果对 ChatGPT 等生成式人工智能意识形态风险的诱因、样态和应对路径展开了多维度诠释,但缺乏对其意识形态风险的系统性探讨。基于此,本文首先以 ChatGPT 等生成式人工智能的工作原理为主线,结合意识形态的运行框架,探讨其何以产生意识形态风险;其次,以意识形态领导权、话语权、管理权的“三权”结构为底层框架,分析产生何种意识形态风险;最后,针对 ChatGPT 等生成式人工智能意识形态风险特性,尝试构建“价值-话语-治理”的三维分级应对策略,探究意识形态风险防范的整体性方案。

## 二、机制探源:生成式人工智能何以产生意识形态风险

所谓生成式人工智能(Generative AI)是指基于学习大规模数据集而生成新的原创性内容的新

型人工智能。生成式人工智能作为人类社会实践的产物,早已僭越出自身工具属性范畴,与社会系统中的政治、文化、价值等要素交织叠加在一起,正在以内容生成、交互对话、多维应用等方式发挥着意识形态功能。从发生学意义上讲,生成式人工智能的意识形态风险发生于意识形态的作用框架之内,即“价值观念-话语方式-社会实践”这一套由内到外、层层嵌套的互动联结整体。故而,探析生成式人工智能如何产生意识形态风险,在聚焦其自身特性的同时也应复归到意识形态的作用机制之中。

### (一) 内容生成:生成式人工智能对价值观念的渗透

价值观念是意识形态作用机制的内核圈层,凭借其价值“制高点”的势能规约着意识形态话语方式和行为实践的向度与样态,是意识形态保持生命力的内在驱动。生成式人工智能在意识形态价值观念层的嵌入逻辑就是以内容生成为基点,经由大数据的收集检索和对算法推送的系统整合,以内容生产者的角色在意识形态领域实现多方位的价值观念重构。究其技术本质,生成式人工智能以深度学习(Deep Learning)算法为底层逻辑,即计算机模拟人类大脑神经网络的工作原理,学习归纳已有数据后进行模仿式、缝合式创作,生成全新内容又能解决判别问题,辅助人类进行信息内容的生产。生成式人工智能所生成的内容不仅包括文本、图像、音频、代码等外显性内容,同样包括价值、观点、策略、逻辑等内隐性内容,两种明暗交织的内容生成逻辑在反映一定意识形态的同时,也熔铸与重构着原有的意识形态。内容生成的各个环节都负载着价值判断和意识形态倾向。一方面,在内容生成的准备阶段,包括噪音冗余数据的筛选剔除、结构化处理以及关联性建构等环节都内含着一定价值预设和文化选择。譬如,训练 ChatGPT 的原始数据在分布上存在倾斜状况,所收集的数据其中大部分为英文材料,而英文世界所反映的是欧美主流国家的价值导向和观念偏好,产生的是契合资本主义意识形态的内容;另一方面,在生成式人工智能的内容生成的反馈和应用等环节中也负载着利益相关者的利益和好恶。北京理工大学计算机学院研究者经过实验指出,“ChatGPT 生成内容中存在对于中国的大量偏见言论”<sup>[9]</sup>。大语言模型的优化并非自我管理、自我配置的纯粹抽象计算过程,而是一个需要人工不断参与调整修改参数的“微调”过程,ChatGPT 等生成式人工智能通过特定内容输出的长期渗透会加剧偏见价值传播。

### (二) 交互对话:生成式人工智能对话语方式的影响

“传授一种空洞的语言是不可能的,传授语言的过程本质上就是传授意识形态的过程”<sup>[10]</sup>。意识形态价值内核的展示并非自我授权,而是依赖于一套系统的言说范式、规则和策略——话语方式,它对意识形态进行解释、传播和普及。ChatGPT 等生成式人工智能在意识形态话语层面的嵌入逻辑就是以交互对话的形式,以快速生成、高度拟人的话语构建范式出场,借由逻辑缜密、形式完整和表达准确的话语实践争夺话语权威。一是从对话形态看,ChatGPT 等生成式人工智能通常采用一问一答的形式与用户展开多轮对话,展现出富有层次、通顺连贯的对话能力,大部分答案通过“一、二、三”的分点阐述并以总结性的话语收尾。同时又以持续对话的形式占有用户对某一议题的全部认知,而使用者在社会交流实践中又不断地参与对话,与多元社会主体共同建构了生成式人工智能话语方式的传播图景。二是从话语风格看,ChatGPT 等生成式人工智能以高度拟人化的“对话者”身份介入原先机械僵硬的人机交互,同时还有承认错误并纠正和拒绝不合理请求等表现,以看似“退一步”的妥协反向地增添其自身话语可信度,这种媲美真人对话的人机交互更具诱惑力与吸引力,也意味着其对意识形态话语方式的影响将更隐秘、更令人难以察觉。三是从话语规则看,



ChatGPT等生成式人工智能的话语运行规则具有鲜明的话语展开逻辑。譬如在使用ChatGPT时,用中文提问得到的答案信息量要比英文少得多,同时会根据用户反馈一步步调整答案,这种问答式的、不断通过提问者的否定来层层递进完成问题探讨的方式,正是著名的有“苏格拉底助产法”之称的辩证术<sup>[11]</sup>,ChatGPT呈现出的是西方逻各斯框架。

### (三) 多维应用:生成式人工智能对行为实践的介入

从意识形态运行过程来看,行为实践是意识形态完整运行机制的逻辑终点,只有意识形态这一观念上层建筑深入到实践的维度,才能从“解释”世界走向“改造”世界。ChatGPT等生成式人工智能在意识形态实践领域的嵌入逻辑就是以多维应用为契机,通过多元的场景适应性将实践主体包裹在其所营造的场域中,实现对个体的行为范式、生产方式和实践理路的普遍渗透和隐蔽规训,实现对人这一意识形态主体在更大程度上的宰制与介入。一是全景式嵌入。相较于以往只针对单一领域的分析式人工智能产品(如“阿尔法狗”只能适用于棋类游戏),ChatGPT等生成式人工智能展示出可以嵌入泛在应用场景的强悍适应性,既可用于解决诗词编写、教育辅助、代码编译等多重任务,也具备跨文本、音频、图像、视频等多模态生成功能。当前,OpenAI已宣布将与教育、商业、制造等多行业展开合作,ChatGPT如同一张正在延伸的巨网强化着其对社会的掌控能力。这种“全景嵌入”的幻觉依据场景的泛在性植入人的认知,进而形塑了一个人工智能高度介入行为实践的交互性新场景,使得它或将成为一种海德格尔所说的“座驾”控制和支配着现代人的全部生活。二是隐性介入。ChatGPT等生成式人工智能对意识形态主体行为的影响通常不在于以理性逻辑的力量说服人,而是略过理智思考,直接闯入实践情境,将经过处理和筛选的特定内容与人们的意识形态相勾连,以其内在的思维范式和调控原则介入人的行为实践,使人们在无意识中接受生成式人工智能的潜隐规训。

## 三、理性审思:ChatGPT诱发的意识形态风险研判

在人工智能技术普遍应用的现代性背景下,人类社会俨然成为了以技术与意识形态高度糅合的风险社会。习近平总书记指出,“要加强人工智能发展的潜在风险研判和防范,维护人民利益和国家安全”<sup>[12]</sup>。作为前沿人工智能技术,ChatGPT等生成式人工智能在赋予社会发展的前进动力时也诱发一系列意识形态风险,其中包括意识形态引领乏力风险、意识形态话语减蚀风险、意识形态管理弱化风险。

### (一) 意识形态引领乏力风险

党的二十大报告指出,要“建设具有强大凝聚力和引领力的社会主义意识形态”。社会主义意识形态发挥引领作用,将社会个体各种分散的、冲突的、零碎的价值观点凝聚为共同的价值共识,同时不断强化、提炼、传递主流意识形态而彰显出强大引领力量。在数字化时代,伴随着人工智能技术在意识形态领域的变革效应,ChatGPT等生成式人工智能的运行机制深蕴的资本逻辑解构社会个体已形成的认知和价值,使得意识形态面临引领乏力风险。

#### 1. 资本逻辑渗透导致主流意识形态引领乏力

“资产阶级,由于开拓了世界市场,使一切国家的生产和消费都成为世界性的了”<sup>[13]</sup>。资本通过操纵ChatGPT等生成式人工智能追求价值增值,所掀起生产关系的变革正以科技领域为圆心逐渐席卷至经济、政治、军事等各领域,附着于它的资本逻辑也随之渗透扩张,对主流意识形态引领造

成严峻的风险隐忧。互联网巨头在资本逻辑的驱动下凭借自身数据资源和关键技术获得“资本-技术”的双重垄断优势,通过设定一系列迎合资本逻辑的算法规则、安全标准和价值判断,放任 ChatGPT 等生成式人工智能侵犯用户权益,利用算法进行特定“信息投喂”。而大量主流意识形态相关信息无法满足资本增殖的需求,被隐匿于算法背后的资本逻辑重新排列,在资本逻辑的“拣选”机制中被隐蔽、下沉、后置。生成式人工智能只生产内容,却无法判断内容本身的真假,而虚假信息往往触及热点或博取眼球被算法推送至信息前沿,能更快速实现资本增值,在真假难辨的信息场域下,导致主流意识形态面临的噪音更多,引领的难度更大。

## 2. 科技霸权加码威胁主流意识形态领导权

2023年3月,外交部发布的《美国的霸权霸道霸凌及其危害》白皮书指出,“美国在高科技领域大搞垄断打压、技术封锁,遏阻其他国家科技和经济发展”<sup>[14]</sup>。即美国凭借其在科技生产、分配、交换和消费领域的垄断优势,通过一系列不公平竞争手段企图阻滞新型大国的崛起速度。目前,ChatGPT 作为美国当前最先进的人工智能技术,无法排除被美国政府或其他行为主体利用而进行意识形态布势的可能性。美国为维护其霸权地位,企图通过“数字殖民”“算法霸权”,甚至组建垄断的、排华的、小院高墙式的“技术联盟”等手段进行意识形态演变,这对我国意识形态领导权安全的威胁不容小觑。

## (二) 意识形态话语减蚀风险

“话语背后是意识形态的幽灵”<sup>[15]</sup>。话语在形成、传播和发挥作用的过程中必然会深深印刻话语主体所代表的意识形态烙印。ChatGPT 等生成式人工智能展示的多元化的话语选择、个性化的话语生产和精准化的信息推荐的功能,实为一种柔性的、隐而不宣的意识形态话语控制力量,冲击着主流意识形态话语的传播力、引导力、影响力与公信力。

### 1. “人机协同”的话语生产削弱主流意识形态话语深度

在传统意识形态工作模式中,意识形态话语生产者往往由具有鲜明政治立场、良好知识储备的“现实的人”担任。而 ChatGPT 等生成式人工智能的出现延拓了意识形态话语生产的主体,却削弱主流意识形态话语深度。原因在于 ChatGPT 等生成式人工智能的话语生产是一种“从有到有”的生成,就是对海量未经组织的原始话语数据进行排列处理,而难以创造出富有批判性、创造性的话语内容,这使得主流意识形态话语赖以生存的逻辑自洽性和理论阐释力面临解构,“人机协同”的话语生产模式在一定程度上削弱了主流意识形态话语的渗透力。

### 2. 多元化的话语选择淡化主流意识形态话语权威

“意识形态话语权是话语主体引导民众的意识形态导向、调控意识形态教育话语内容以及表达思想观念、政治立场与价值取向的能力与权力”<sup>[16]</sup>。在当前人工智能嵌合话语表达的拟态环境下,意识形态话语权威面临话语主体旁落和把关缺位的双重挑战,意识形态话语的自主选择性与复杂多元性为许多虚假信息和煽动性言论等提供了肆意发散的空间,致使主流意识形态话语权威建构面临前所未有的风险。ChatGPT 等生成式人工智能基于庞大数据量的深度学习,可以根据某一特定的文字风格抽象出算法规则并创作出具有该风格的崭新作品,展示出一幅话语选择的多元化图景。多元的话语选择导致主流意识形态的话语传播生态更趋复杂多样,尤其是广大网民对多元化、趣味化的话语追逐替代了对逻辑性、政治性话语的关注,导致官方话语效力衰弱,导致意识形态话语权威在言说传递的层层过滤中被消解淡化。

### (三) 意识形态管理弱化风险

意识形态领域的有效管理是意识形态领导权与话语权的具体实现方式,对意识形态建设具有持续性的激励与约束作用,也是党执政稳固、国家长治久安的关键影响因素。马克思说:“矛盾和对抗不是从机器本身产生的,而是从机器的资本主义应用产生的!”<sup>[17]</sup>随着 ChatGPT 等生成式人工智能的广泛应用,其自身的运行逻辑与行动框架渗透到社会实践中,在一定程度上冲击着我国意识形态的管理建设。

#### 1. 多元社会思潮借由智能传播新范式消解意识形态管理的主导权

“以 ChatGPT 为代表的智能传播将带来人类信息传播范式的根本转变,也意味着旧有治理范式的失效和缺失”<sup>[18]</sup>。即人类从公开平台获取信息和内容的社交传播,已经开始转向非公开的、由个人从智能机器获取信息内容的智能传播方式。具体而言,个人通过问询生成式人工智能而获得私人化信息内容的过程并不对外公开,但每次对话的信息将被记录融合在生成式人工智能的算法之中并整合至后续的意识形态传播中。换言之,ChatGPT 等生成式人工智能所引发的智能传播革新弱化了官方的“把关人”角色,将使得“新自由主义论”“意识形态中立论”“马克思主义过时论”等多元负面社会思潮更加难以识别和捕捉,也为不良思潮的隐匿传播提供滋生的温床,给我国新时代意识形态管理带来了严峻的挑战。

#### 2. 生成式人工智能技术对人的宰制加大意识形态管理难度

ChatGPT 等生成式人工智能技术在解放人的脑力劳动的同时,也实现了对人的多方位控制。实事求是地说,人类社会目前并未完全实现“对那些异己力量的控制和自觉的驾驭”<sup>[19]</sup>,仍处于以劳动教育为主和凭借劳动时间的付出来获得物质生活资料的历史阶段。因此,当人过度浸淫在对 ChatGPT 等生成式人工智能的崇拜依赖中,人的认知判断力、过程决策力、想象创造力在人工智能构建的信息世界中被收缩、削弱。ChatGPT 等生成式 AI 被赋魅为一种“主人”的新型公共权力,而将人“降格”为技术附庸,人的真实本质被放逐,人作为意识形态主体的能动性被消弭,进一步加剧意识形态领域管理危机。

## 四、防范应对:ChatGPT 等生成式人工智能的意识形态风险规制策略

在百年未有之大变局的时空结构下,如何化解防范 ChatGPT 等生成式人工智能所带来的意识形态风险,正确破解“技术利维坦”带来的技术难题成为了现代国家意识形态阵地建设的关键一步。爱因斯坦曾对人类的未来做出预测:人类未来的光明前景必须是以关注人类自身为主线,保证科学技术的合理运用来造福人类<sup>[20]</sup>。马克思主义的科技观始终高举着发扬人类主体能动性的旗帜,始终重视人的本质的全面显现。实际上,意识形态风险是一种“可能迟滞或中断中华民族伟大复兴进程的全局性风险”<sup>[21]</sup>。ChatGPT 等生成式人工智能诱发的意识形态风险,已成为维护国家意识形态安全、政治安全无法回避的新课题。

### (一) 坚持马克思主义,以“人民至上”超越“资本至上”

资本是社会主义市场经济的重要生产要素,在社会主义市场经济条件下规范和引导资本发展,既是一个重大经济问题,也是一个重大政治问题<sup>[22]</sup>。生成式人工智能的发展离不开资本的驱动,中国是社会主义国家,这要求必须始终坚持资本运行的“社会主义”底色。因此,要溯回到马克思主义这一立党立国的根本指导思想,面对潜藏于生成式人工智能中“资本至上”的意识形态攻势,必须



以马克思主义“人民至上”核心价值超而越之。

### 1. 坚持马克思主义在人工智能产业的指导地位

“如果在某个领域放弃了马克思主义的指导地位,就会导致灾难性后果”<sup>[23]</sup>。生成式人工智能技术与产业的发展存在巨大利益空间,故尤其需要警惕资本利益凌驾于人民利益之上的价值倾向。谨防利益竞逐的“资本至上”逻辑对主流意识形态的蚕食,一方面构建高水平社会主义市场经济体制,依法规范和引导资本,以促进 AI 产业有序健康发展;另一方面,切实贯彻习近平新时代中国特色社会主义思想的世界观和方法论,做到人工智能领域“发展为了人民、发展依靠人民、发展成果由人民共享”<sup>[24]</sup>。

### 2. 坚持马克思主义对生成式人工智能时代人的思想引领

“在坚持马克思主义指导地位这一根本问题上,我们必须坚定不移,任何时候任何情况下都不能有丝毫动摇”<sup>[25]</sup>。在生成式人工智能深度变革人的思维方式的背景下,必须用马克思主义育人化人,加强思想引领。一方面坚持用中国化、时代化的马克思主义教育人民、武装头脑、振奋精神;另一方面,通过宣讲、教育等多种形式增强人民群众使用马克思主义进行辨别、批判错误社会思潮的能力。

### 3. 坚持人文本位与机器辅位相统一的原则

马克思还曾鲜明地指出,“任何解放都是使人的世界即各种关系回归于人本身”<sup>[26]</sup>。诚然,生成式人工智能的发展有助于进一步解放人的脑力劳动,但在资本驱动下也势必会为了攫取高额利润而不断强化着人对它的黏性、依赖性。因此,要将生成式人工智能框定在为“每个人的自由发展”和“自由人联合体”创造条件的范围之内。具体而言,在坚持人文本位与机器辅位的统一中,既要秉持终身学习理念提升人的劳动能力,又要探索以人为主体、更好发挥人类智能和人工智能各自优势的人机协同新形态。

## (二) 坚持讲好中国故事,以中国话语破解西方话语

内置于 ChatGPT 的西方话语底蕴与逻辑在国际科技交往中日趋显现,并逐渐挤压着中国在国际科技竞争中的话语空间。若失去生成式人工智能赛道的话语权,我国就会在全球科技竞争中被严重掣肘,进而削弱意识形态话语影响力。因此,除了提升科技硬实力和坚定主流价值导向外,还必须进一步讲好中国故事,以“中国话语”破解当前生成式人工智能领域盘踞的“西方话语”。

### 1. 立足阵地思维,站好中国立场

“对于社会主义思想体系的任何轻视和任何脱离,都意味着资产阶级思想体系的加强”<sup>[27]</sup>。当前,ChatGPT 掀起全球热议成为当之无愧的“话题明星”,“自由”“人权”“民主”等西方话语杂糅民主宪政等错误思潮以多种形式藏匿于热点话题中,侵蚀主流意识形态话语阵地。一方面,要提高错误思潮辨别力和批判力,对隐匿于生成式 AI 中形形色色的错误思潮敢于亮剑;另一方面,结合中国本土生成式 AI 成果创新主流意识形态议题设置,理直气壮讲出“中国话语”,推动中国故事在国际话语交锋中掌握主动权。

### 2. 立足技术思维,传播好中国故事

伴随生成式 AI 在全球范围的广泛应用,人工智能生成内容(AIGC)在国际话语传播中的参与比重将持续加大,国际话语传播的博弈逐渐表征为智能技术的博弈。一贯通过把控传播渠道维系自身垄断的西方话语体系很可能凭借生成式 AI 等新技术延续全球话语的“再垄断”。对此,一方面要

掌握智能传播规律,打通智能信息处理的技术关隘;另一方面,推动生成式 AI 结合 5G、区块链等新兴技术,探索中国话语在智能传播时代的多元平台。

### 3. 立足受众思维,讲好中国故事

探寻能被国际社会普遍理解的世界性中国话语,是掌握意识形态话语主动的关键所在。一方面要深入认识并结合生成式 AI 背景下话语受众的心理、话语和诉求特征,将中国故事以兼具中国特色又接轨国际的方式予以呈现。另一方面,主流意识形态话语以扬弃态度对生成式 AI 时代话语进行征用与收编,消除受众对话语权威的逆反心理。

### (三) 倡导全球合作治理,推出伦理规范的中国方案

“要深度参与全球科技治理,贡献中国智慧,塑造科技向善的文化理念,让科技更好增进人类福祉”<sup>[28]</sup>。无论是 ChatGPT 等生成式人工智能抑或其他人工智能技术,都是世界性的、时代性的,因而无论是科技互惠共享还是科技风险治理,必须具有全球视野。建构新的共同体意识和采取切实的共同体建设措施深入参与全球人工智能治理<sup>[29]</sup>,以“中国方案”为全球人工智能治理贡献智慧,是应对意识形态管理失效风险的可行之径。

#### 1. 在理念之维,以“共商共建共享”推动构建生成式 AI 全球治理的伦理共识

共商共建共享是中国一贯奉行并持续践行的全球治理观,是多边主义精神真正的彰显。面对生成式 AI 技术可能对意识形态安全产生的威胁,应进行深入沟通,以大局为重、求同存异。应尽快推动组建关于生成式 AI 的伦理委员会,邀请各国教育、科技、伦理、法律等领域专家学者共同论证及风险评估,确保技术更新迭代始终为人类服务。

#### 2. 在主体之维,以“人类命运共同体”推动构建生成式 AI 全球治理的多元协同

命运共同体是解决全球人工智能治理难题的中国方案<sup>[30]</sup>。构建人工智能命运共同体为解决生成式 AI 背后潜藏的跨国治理难题提供了可能性,在世界各方协同治理下,有利于实现人类的共同利益和安全繁荣。进言之,推动生成式 AI 治理,还需加强多主体间的对话交流合作,共同抵御其可能产生的意识形态风险。

#### 3. 在制度之维,以“负责任的人工智能”推动构建生成式 AI 全球治理的透明监管

生成式 AI 发展的快速程度及风险涌现的紧迫程度,呼唤并要求生成式 AI 意识形态风险治理的全球监管机制加速形成。中国新一代人工智能治理委员会提出的“负责任的人工智能”核心概念引导利益攸关方思考生成式 AI 监管准则制定方向。一方面根据内容生成类智能工具的技术特点,明确划定使用准则和伦理界限,尤其注重其在教育、学术、法律等领域的使用限制;另一方面划定生成式人工智能技术责任主体,“由政府规制承担掌舵者的风险防控职责,协调人工智能企业、第三方机构和社会公众参与风险治理,构建框架性的软硬法混合规范体系,形成从人工智能企业内部到认证机构、关联企业以及社会公众的风险合规监督机制”<sup>[31]</sup>。

## 五、研究总结

马克思指出:“问题就是公开的、无畏的、左右一切个人的时代声音。问题就是时代的口号,是它表现自己精神状态的最实际的呼声。”<sup>[32]</sup>人工智能技术给意识形态带来的风险困境是现代社会的必须面对的时代难题,尤其是以 ChatGPT 为代表的生成式人工智能风险造就了技术时代的危机图景,人类文明生活在现代文明的十字路口,既充满了时代机遇又充斥着风险危机。这种风险实际上



就是可能导致马克思主义的异化,而相较于现代社会的科技异化困境,ChatGPT等生成式人工智能以更隐蔽、更虚拟、更多元的形式对意识形态领域造成危机挑战。在风险与收益相伴而生的今天,ChatGPT等生成式人工智能只有真正回归以人为本的科技理念,调和意识形态风险的异化鸿沟,厘清人工智能技术的风险边界,才能实现人机和谐、技术向善的美好愿景。

#### 参考文献:

- [1] 尤尔根·哈贝马斯. 作为“意识形态”的技术与科学[M]. 李黎,郭官义,译. 上海:学林出版社,1999:69.
- [2] 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[M]. 北京:人民出版社,2022:28.
- [3] 王少. ChatGPT介入思想政治教育的技术线路、安全风险及防范[J]. 深圳大学学报(人文社会科学版), 2023(2):153-160.
- [4] 邹开亮,刘祖兵. ChatGPT的伦理风险与中国因应制度安排[J]. 海南大学学报(人文社会科学版), 2023(4):74-84.
- [5] 王延川,赵靖. 生成式人工智能诱发意识形态风险的逻辑机理及其应对策略[J]. 河南师范大学学报(哲学社会科学版), 2023(4):1-7.
- [6] 高奇琦,严文锋. 知识革命还是教育异化? ChatGPT与教育的未来[J]. 新疆师范大学学报(哲学社会科学版), 2023(5):102-112,2.
- [7] 张夏恒. ChatGPT的政治社会动能、风险及防范[J]. 深圳大学学报(人文社会科学版), 2023(3):5-12.
- [8] 李舟,易华勇. ChatGPT介入意识形态安全治理的逻辑、风险及应对[J]. 闽江学刊, 2023(4):67-74,171-172.
- [9] 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):16-25.
- [10] 俞吾金. 意识形态论[M]. 北京:人民出版社,2009:2.
- [11] 张生. ChatGPT: 褶子、词典、逻辑与意识形态功能[J]. 传媒观察, 2023(3):42-47.
- [12] 新华社. 习近平在中共中央政治局第九次集体学习时强调 加强领导做好规划明确任务夯实基础 推动我国新一代人工智能健康发展[J]. 党建, 2018(11):1,19.
- [13] 马克思,恩格斯. 马克思恩格斯选集(第1卷)[M]. 北京:人民出版社,2012:404.
- [14] 美国的霸权霸道霸凌及其危害[N]. 人民日报, 2023-02-21(17).
- [15] 侯惠勤. 意识形态的变革与话语权——再论马克思主义在当代的话语权[J]. 马克思主义研究, 2006(1):45-51.
- [16] 李超民,陆慧. 意识形态“三权”的知识谱系、协同功能与实现方法[J]. 广西社会科学, 2019(2):38-43.
- [17] 马克思,恩格斯. 马克思恩格斯全集(第19卷)[M]. 北京:人民出版社,1972:483.
- [18] 钟祥铭,方兴东,顾焯焯. ChatGPT的治理挑战与对策研究——智能传播的“科林格里奇困境”与突破路径[J]. 传媒观察, 2023(3):25-35.
- [19] 马克思,恩格斯. 马克思恩格斯选集(第1卷),北京:人民出版社,1972:42.
- [20] 爱因斯坦. 爱因斯坦文集[M]. 许良英,译. 北京:商务印书馆,1973:73.
- [21] 中共中央党史和文献研究院. 十九大以来重要文献选编(中)[M]. 北京:中央文献出版社,2021:654.
- [22] 依法规范和引导我国资本健康发展发挥资本作为重要生产要素的积极作用[N]. 人民日报, 2022-05-01(02).
- [23] 朱继东. 科学理解坚持马克思主义在意识形态领域指导地位的根本制度[J]. 马克思主义研究, 2020(10):117-125.
- [24] 习近平. 在省部级主要领导干部学习贯彻党的十八届五中全会精神专题研讨班上的讲话[M]. 北京:人民出版社, 2016:24-25.
- [25] 习近平. 习近平谈治国理政(第2卷)[M]. 北京:外文出版社,2017:33.
- [26] 马克思,恩格斯. 马克思恩格斯文集(第1卷)[M]. 北京:人民出版社,2009:46.
- [27] 列宁. 列宁全集(第6卷)[M]. 北京:人民出版社,2012:327.
- [28] 习近平. 加快建设科技强国 实现高水平科技自立自强[J]. 求知, 2022(5):4-9.
- [29] 高奇琦. 马克思主义视域下的人工智能与未来治理之道[J]. 政治学研究, 2021(3):78-88,162.
- [30] 张冬冬. 人类命运共同体理念下的全球人工智能治理:现实困局与中国方案[J]. 社会主义研究, 2021(6):164-172.
- [31] 王洋,闫海. 生成式人工智能的风险迭代与规制革新——以ChatGPT为例[J]. 理论月刊, 2023(6):14-24.
- [32] 马克思,恩格斯. 马克思恩格斯全集(第40卷)[M]. 北京:人民出版社,1982:289.

## The ideological risks of generative artificial intelligence such as ChatGPT and its response

DAI Jinping, QIN Yangyang

(*School of Marxism, University of Electronic Science and  
Technology of China, Chengdu 611731, P. R. China*)

**Abstract:** Ideological risk prevention is a key area related to national rejuvenation, national sovereignty and cultural construction. Today, as artificial intelligence technology is booming, generative AI technology represented by ChatGPT has been penetrated into the society. And with the expansion of the depth and scope of its interactions, it has wrapped up with many ideological risks with new characteristics. The prerequisite of preventing the ideological risk of generative AI, such as ChatGPT, is to clearly identify its internal mechanism. It makes an infiltration into people's ideological value by virtue of its primary function, which is generating contents. It affects the discourse of people's ideology by virtue of a form of interaction and dialogue with them. It strongly gets involve in people's practices and their social life by virtue of the function of multi-dimensional application, eventually it has led to the current ideological struggle to be extremely controversial and ever-changing. The key to prevent the ideological risk of generative AI, such as ChatGPT, is to take rational examination and precise identification on it. From the perspective of ideological leader, the capital logic embedded in its operation mechanism deconstructs the cognition and values already formed by members of the society, and leads to the risk of ineffective ideological leadership due to the dual advantages of "technology-capital". From the perspective of ideological discourse, it provides a new discourse mode of "human-machine collaboration", which weakens the depth of mainstream ideological discourse. And its multiple generative discourse delivery has abated the authority of discourse, resulting in the risk of eroding ideological discourse. From the perspective of ideological management, the new paradigm of information dissemination triggered by generative AI has eliminated the dominant right of ideology management and strengthened its control over people, which brought about the risk of weakening ideological management. Generative AI, such as ChatGPT, takes technological iteration as an opportunity of the times, and on the material basis of advanced technology, it has led to a series of ideological risks with extensiveness, comprehensiveness and uncertainty. How to prevent the ideological risks brought about by ChatGPT and other generative AI has become a new topic that cannot be avoided in China's ideological construction. First, we must stick to the value of "putting the people first", regard Marxist theory as the guideline of ideological prevention, and promote the building of a cyberspace with human-machine harmony and clean air. Second, we need to build a discourse channel with Chinese characteristics, and break the modern dilemma of Western discourse on the basis of discourse of telling a good Chinese story. Lastly, we need to actively advocate for global governance and contribute to China's solutions. We must uphold the principle of extensive consultation, joint contribution and shared benefits, rely on the carrier of a community with a shared future for mankind, and build a responsible landscape of artificial intelligence. By doing so, generative AI is prompted to show its bright future in the face of modern civilization.

**Key words:** ChatGPT; generative artificial intelligence; artificial intelligence; ideological risk; risk prevention

(责任编辑 彭建国)