人工智能诱发隐性意识形态话语风险的逻辑机理及化解策略*

王海威

【内容提要】人工智能作为一种技术力量，隐匿地发挥着意识形态话语的导向、辩护、凝聚等功能，推动意识形态领域的风险和挑战更趋错综复杂，对国家安全产生影响。西方国家将人工智能与政治深度互嵌，催生了意识形态话语主体裂变式繁殖，重塑了意识形态话语生产传播格局与话语生态，衍生出大国政治与意识形态博弈的文化新边界，诱发了原发性风险、继发性风险和并发性风险等隐性意识形态话语风险。本文基于意识形态话语的认知解释、价值信仰和目标策略三层结构，分别从技术逻辑、资本逻辑和人本逻辑分析人工智能诱发隐性意识形态话语风险的逻辑机理，提出风险化解策略，即以技术匡正技术、以价值引领技术、以规制规范技术等，提升人工智能新场域的主流意识形态话语权。

【关键词】人工智能　隐性意识形态　话语风险　安全治理

作者简介：王海威（1978-），东北财经大学马克思主义学院教授（辽宁大连　116024）。

纵观人类社会发展历程，从地理大发现起历经两次工业革命时期、信息革命时期，到如今的人工智能时代，每一次重大技术革命都给国家安全带来新挑战。人工智能作为一种技术力量，在赋能人类社会发展正能量的同时，也推动意识形态风险挑战更趋错综复杂，成为大国意识形态博弈的“新战场”。相较于蒸汽机技术、电力技术、原子能技术、信息技术等前几次科技革命的核心技术，人工智能具备感知和模拟人类思维的特性，技术本身的不确定性风险更高。而且，人工智能具有强大的信息编码能力，在运行中塑造了“作为意识形态的表达符号和编码工具”① 的意识形态话语，隐匿地发挥着意识形态话语的导向、辩护、凝聚、动员、约束等功能，内蕴的意识形态话语风险更趋隐性化。准确辨识人工智能诱发的隐性意识形态话语风险因子，剖析其形成的内在逻辑机理，防范化解隐性意识形态话语风险，是人工智能时代提升主流意识形态话语权的重要任务。

一、人工智能诱发的隐性意识形态话语风险样态

“意识形态话语是理解包括政治上层建筑以及思想文化在内的意识形态的重要介质，其在一定

* 本文系国家社科基金重点项目“大历史观视域下社会主义意识形态理论话语创新研究”（22AKS008）的阶段性成果。

① 冯冉、陈锡喜：《系统观念视域下新时代社会主义意识形态话语创新的三个着力点》，《湖北社会科学》2023 年第 7 期。

程度上也建构了意识形态。”① 就其横向结构来看，社会主义意识形态话语包括话语主体、话语表达、话语对象和话语场域等要素。以智能算法、算力和大数据为核心的人工智能，是一个大型的语义网络，其语言模型是意识形态的量化生产者，潜移默化中可能诱发更为隐蔽、多元和复杂的风险，主要表现为原发性意识形态话语风险、继发性意识形态话语风险和并发性意识形态话语风险。

1．原发性意识形态话语风险

原发性意识形态话语风险是“一种基础应用场景下，嵌于技术底层的内生性意识形态风险”，“源自生成式人工智能技术的核心”②。西方国家将人工智能技术与政治深度互嵌，使人工智能成为一部“意识形态生产机器”。

一是人工智能技术与政治深度互嵌，衍生科技势能向政治势能转变风险。技术作为社会产物，自诞生起就负载价值。人工智能技术作为科技进步成果，被引入传播场域，成为一种隐性的技术意识形态力量，给我国意识形态安全带来严峻挑战。

一方面，意识形态话语风险内嵌于人工智能技术研发过程。人工智能技术标榜技术价值中立原则，技术中立包括功能、责任和价值三重中立，但在现实中，技术设计国的意识形态决定着人工智能遵循什么样的价值判断规则。在算法的设计和开发过程中，话语权归属、利益考量、价值观认可等问题绝不只是技术问题，也是政治经济问题③，这使人工智能成为有价值观偏向的“话语再生产”工具。算力、算法、数据作为人工智能发展的三要素，是衡量人工智能发展水平的重要参考，人工智能的原发性意识形态话语风险正是发端于海量问题数据。比如，ChatGPT 是一种基于网络大数据的人工智能技术，以海量数据喂养的预训模型为基石，其训练数据绝大部分来自欧美。被西方世界操控的数据基因，决定了 ChatGPT 数据内嵌意识形态价值偏置，在算法技术逻辑操纵下，其输出的内容带有鲜明的政治立场和倾向，成为贯彻西方意识形态的宣传工具。截至 2024 年 3 月，ChatGPT 在全球的活跃用户数已经超过了 10 亿。随着其舆论属性和社会动员能力不断增强，ChatGPT 正以隐性思维方式改变人们的思维结构和理解图式，随之而来的隐性意识形态话语风险不容忽视。

另一方面，意识形态话语风险产生于人工智能技术应用的“异化”。科学技术的应用具有鲜明的意识形态属性。马克思曾在《资本论》中指出，“智力转化为资本支配劳动的权力，是在以机器为基础的大工业中完成的”④。科学技术的风险源于具体应用场景下科学技术应用的“异化”。在基础应用场景中，人工智能技术囿于自身的缺陷，呈现出意识形态传播放大价值偏见等潜在风险。ChatGPT 的介绍文本自称，“我的知识仍然受到人们提供的信息的限制”。联合国教科文组织通过的《人工智能伦理问题建议书》警示：“人工智能算法可能复制和加深现有的偏见，从而加剧已有的各种形式歧视、偏见和成见。”⑤ 当前人工智能技术尚处于西方领先状态之下，其输出内容极有可能强化基于“西方中心论”基础上的文化偏见和歧视，并扩大发达国家与发展中国家的“认知鸿沟”。在恶意应用场景下，出于政治操纵、资本逻辑等目的，人工智能技术呈现出虚假信息和行为诱导等潜在意识形态话语风险。以美国为代表的欧美发达国家，科技与经济的泛政治化和泛意识形态化加剧⑥，以脸书为代表的美国企业正在推动政府加强对元宇宙的认知，帮助政府以“负责任”的态度

① 冯冉、陈锡喜：《系统观念视域下新时代社会主义意识形态话语创新的三个着力点》，《湖北社会科学》2023 年第 7 期。

② 欧阳林洁、张永红：《生成式人工智能应用的意识形态风险：命题由来、生成机制与治理进路》，《学术探索》2023 年第 11 期。

③ 参见杨家明、张萌：《认知基础设施与算法在认知域的制度化》，《青年记者》2024 年第 2 期。

④ 《马克思恩格斯选集》第 2 卷，北京：人民出版社，2012 年，第 227 页。

⑤ 联合国教科文组织：《人工智能伦理问题建议书》，https://unesdoc.unesco.org/ark:/48223/pf0000380455_chi。

⑥ 参见姚远、方文青：《科技革命与全球治理新议题——“南京论坛 2021”国际关系分论坛综述》，《亚太安全与海洋研究》2022 年第 2 期。

构建元宇宙版图，意图通过“元宇宙+虚拟货币”吸纳管辖全球经济金融等资源，实现元宇宙“政治战略”。如不及时加以有效规制，人工智能技术将把其所承载的价值立场从虚拟空间延伸到现实世界，带来现实版“文明的冲突”，衍化为思想殖民和文化霸权，成为新型意识形态机器和新帝国主义工具，推动虚拟空间日益成为大国意识形态博弈的新战场。

二是算法重塑话语传播格局，形成大国政治与意识形态博弈的文化新边界。作为人工智能技术在传播领域的创造性应用，智能算法推荐的普及改变了传统信息传播的运行机制和规则流程。其一，以计划性宣传催生基于算法的国际政治新操纵。计划性宣传是一种数字化、智能化、自动化的宣传手段，其以大数据与智能算法为支撑，能够基于大数据分析目标受众、定制推送内容，并基于智能算法有目的、有计划、有组织地虚构政治同意、制造政治共识、操纵公共舆论，通过控制信息的内容、流量与流向来影响公众的认知、决策与行动，以实现自身的政治目的。以美国为首的西方资本主义发达国家把算法技术“作为维护本国政治安全、输出价值观、实现国家意志的战略手段”①，从“五眼联盟”到“棱镜”计划，从“诚实之声”到“网络魔术师”，再从帕兰提尔到剑桥分析，都显示出以美国为代表的西方国家操纵算法技术为其政治统治服务的真实面目。其二，以权力性宣传催生意识形态话语叙事新格局。数据、算法、算力等数字基础设施与政治资本、权力之间紧密相连，成为一种隐性的社会技术权力。由于算法的技术制式和政治经济逻辑已经嵌入了互联网与数字社会的底层架构，渠道平台化、把关算法化、议题弥散化带来了意识形态传播权力的转移、分化和重组。算法推荐主导的意识形态传播重塑了信息场景和权力空间，算法系统驱动的个性化推送增加了政治选择性接触。算法歧视、暗箱操作等不仅对既有政治秩序和价值体系产生破坏性冲击，还重塑了整个信息权力的生产、流通与分配过程，改变和重组了意识形态话语叙事格局。“当生成式人工智能技术通过已然普及的互联网跃现于国际政治舞台时，核心技术国家的辐射力、渗透力将得到显著增强，由智能机器持续不断生产的各类信息逐渐弥漫于网络空间，使得人工信息被遮蔽和吞没，从而形成大国政治与意识形态博弈的文化新边界。”②

2．继发性意识形态话语风险

继发性意识形态话语风险与原发性意识形态话语风险不同，并非生成于技术路径，而是诞生于应用场景。作为人类社会实践的产物，人工智能早已僭越自身工具属性范畴，与政治、文化等要素交织叠加，消解主流意识形态的话语权威性。

一是话语主体裂变式繁殖，冲击主流意识形态的权威性。意识形态话语是言说主体对原始内容进行思维构建的体现，言说主体的价值取向及阐释方式影响着意识形态话语的表达③。在传统意识形态话语结构中，话语流动主要是单向的自上而下的内容供给，主流意识形态居于主导性权威地位。在人工智能际遇下，信息的传播模式逐渐由“人找信息”转变为“信息找人”，改变了传统由记者、编导、总编辑扮演“把关人”的单向度的信息传播模式，催生出意识形态的虚拟话语主体多元化。政府、资本、知识精英、技术精英、普通公众及数字社群等多元话语主体，持续在数字空间中创造新的思想和新的表达，生成式人工智能也介入新闻生产，引发了传播内容复杂多样性和传播风险不可控性。比如，当下关注度较高的 OpenAI 文生视频模型 Sora 作为虚拟话语主体，能够根据用户的一句话生成长达一分钟的视频，“深度伪造”技术催生“眼见不为实”时代来临，使生成内容存在虚实混淆风险。ChatGPT 囿于专业训练数据的缺失和模型局限，常以“一本正经地胡言乱语”弥补

① 中璋：《操纵：大数据时代的全球舆论战》，北京：中信出版社，2021 年，第 15 页。

② 杨章文：《ChatGPT 类生成式人工智能的意识形态属性及其风险规制》，《内蒙古社会科学》2024 年第 1 期。

③ 参见包崇庆、柏路：《人工智能时代主流意识形态话语权建设的多维审视》，《传媒》2023 年第 13 期。

其不健全的技术缺陷，如同一部意识形态机器一样源源不断地生产真假难辨的信息，使生成内容存在失范性风险。“数字化、符号化”的多元话语主体，推动信息传播呈现双向互动性、多样性和跨时空性等特征，实现意识形态话语权从现实社会向网络虚拟空间转化，引发人工智能技术应用中的多样态风险，冲击主流意识形态的话语主导性。

二是“信息茧房”离散话语内容，导致认知窄化和价值偏化。客观理性认知是主流意识形态认同的逻辑起点。在传统意识形态传播结构中，以广域性的覆盖灌输为主导模式。随着人工智能技术的发展，媒体平台和资讯终端为了获取更高的点击量，通常采用“算法+推荐”模式，即通过大数据，根据受众兴趣偏好和浏览历史有选择地推送信息，引发数字信息传播结构变化及“理性认知剥夺”风险。一方面，“信息茧房”引发认知趋于偏狭与窄化。“你关心的，才是头条”的信息推荐模式使人们在不知不觉中被从多样化、立体化和高效性的多元信息和权威观点中剥离开来，导致个体被包裹进同一价值偏向营造的“气泡性”信息空间，陷入同质化“信息茧房”①。比如，针对同一事件的政治立场截然相反的两份报道在 YouTube 推荐算法中并不直接连通，个性化算法在塑造内容与内容连接、人与内容连接中解构对主流意识形态的理性认知。另一方面，算法“定制化”服务暗含价值陷阱。“信息茧房”中的每个人都是自己的话语中心，自我生产、传播和选择信息，算法固化导致思维固化，人们逐渐成为“数据囚徒”，陷入价值混乱。尤其当前自媒体、短视频等新兴传播平台基于“流量逻辑”发布的娱乐性、低俗化、戏谑化信息，成为算法推送的优先方，催生“价值缺位”“精神空场”“意义虚无”等，影响个体对主流意识形态话语的理性认知能力、理解能力和接受能力，制约社会主义核心价值观凝聚。

3．并发性意识形态话语风险

并发性意识形态话语风险是“生成式人工智能技术在长期应用场景下所产生的附随性风险”②，带来人工智能时代的劳动新异化和智能机器排挤人等“数字异化”现象。

一是人机交互重塑意识形态话语生产，衍生新的意识形态战场。梳理近代人类意识形态战场演变，从“WEB1.0”战场的媒体争夺，到“WEB2.0”战场的社交媒体平台争夺，再到“WEB3.0”战场的高智能对话机器人争夺，多重人工智能技术叠加深化了内容的自动生成，意识形态阵地日趋智能化、隐蔽化。基于人机交互的 ChatGPT 等人工智能技术，改变了原有意识形态生成路径，进而以新的技术路径生成了具有算法技术特征的新型意识形态，拓宽了意识形态话语场域。其一，基于拟态环境供给的“主客异位”风险。人机交互技术快速发展，重新定义了信息的生产、分发和消费方式，促进了新形式意识形态话语的生产和传播，已经成为塑造意识形态的关键力量。在传统的信息传播语境中，单向垂直的内容供应，受众处于客体位置。生成式人工智能依托自身强大的算法、算力和海量数据，在内容生产和信息传播领域展现出前所未有的广度与深度，成为一部“意识形态生产机器”，形成了主流意识形态话语生产主体的“合奏效应”。人机交互重置信息传播从主客二元性结构转变为主体间性的互动性结构，引发主流意识形态话语生产主体二元性结构变革，深度改变意识形态实践的主客体关系和人们的认知及思维方式，引发“主客异位”风险。其二，基于沉浸式交互场域的意识形态柔性塑造风险。人机交互具有模拟性特点，以智能算法技术为基础的 ChatGPT 通过理解和学习人类对话，生成文字、图片、视频等内容，形成了“人机协同”的意识形态话语生产方式。在人机协同构建的沉浸式交互场域中，传统趋于稳定的知识生产模式被动态的智能知识库

① 王琎：《新技术变革下意识形态治理研究——理论检审、现实叩问与治理出路》，《马克思主义研究》2023 年第 2 期。

② 欧阳林洁、张永红：《生成式人工智能应用的意识形态风险：命题由来、生成机制与治理进路》，《学术探索》2023 年第 11 期。

所庖代。ChatGPT 类人工智能通过不断演进、增强工具价值以密切与用户间的联系，并逐步建立用户的工具性信任和依赖，易使用户陷入技术依赖窠臼，放弃对信念的理性辩护，大众的意识形态体认逐渐演化为虚拟化和碎片化的形态，真实信息与虚假信息、网络虚拟空间与现实物理世界的边界被模糊。在“沉浸式”拟态应用场景中，精准隐蔽的“数字利维坦”将其附着的西方“普世价值”等错误社会思潮和带有算法价值偏见的有害信息向外灌输与渗透，使用户自身无意识地接受了“另一端”意识形态的柔性塑造。人工智能技术交互语境拟像化，干扰主流意识形态认同，带来技术所有国隐性意识形态渗透风险，使人工智能成为继互联网平台之后意识形态斗争的新战场和新焦点。

二是话语生态“去中心化”，消解主流意识形态影响力。根据马克思主义的理论框架，意识形态是一定社会经济结构条件下的上层建筑，它通过特定的话语体系体现。在传统意识形态传播结构中，话语主体是权威性信息发布中心，在信息输出结构中居权威地位。随着信息技术的发展，数据存储、处理和传输不再依赖中央控制节点的模式，权威性信息传播模式解体，“去中心化”的传播格局形成，这改变了信息生态的权威性结构。随着人工智能技术在新闻采集、生产、分发、接收和反馈中的广泛应用，算法加持下的人工智能载体为意识形态的实时传播提供了新的可能，也重塑了意识形态“去中心化”的话语生态。当前，生成式 AI 进入应用爆发期，OpenAI 的 ChatGPT、谷歌的 Gemini 等生成式人工智能，以其强大的图像、视频、代码和文本生成等数字内容创作能力，参与意识形态话语建构和话语传播。这导致在生成式人工智能媒介建构的话语场中，资本的意识形态渗透匿影藏形，多模态生成场景的“互联网垃圾”无孔不入，主流媒体构建真实的权威被进一步消解，引发新闻内容深度的削弱和新闻价值的失衡，造成意识形态的全域化、动态化、复杂化风险。

二、人工智能诱发隐性意识形态话语风险的逻辑机理

社会主义意识形态话语是“横向要素与纵向层次有机耦合的立体结构系统”①，其纵向结构包括认知-解释、价值-信仰和目标-策略三个层次。把握人工智能意识形态话语风险的本质，需要从纵向的三层结构出发，分析人工智能隐性意识形态话语风险生成的技术逻辑、资本逻辑和人本逻辑。

1. 从认知解释层面看，科技与意识形态相互塑造

认知解释是意识形态的认识论基础。在这个意义上，“选择什么样的事实，怎么判断形势，怎么诠释社会矛盾，是意识形态认知-解释层的主要工作”②，也是主流意识形态的基本功能。人工智能技术嵌入意识形态的认知-解释层面，已成为一种隐性的意识形态权力。“马克思也指出科学技术与意识形态是相互联系的两个整体，科学技术在一定程度上会影响意识形态的内容，而意识形态对科学技术也有反作用。”③

一方面，科技是塑造意识形态的重要力量。马克思指出，意识形态“归根到底是由人们的物质生活条件决定的”④。科技作为这些物质条件的重要组成部分，也在塑造人们的意识形态。从历史逻辑看，回溯技术革命的历史进程，渔猎文明-农业文明-工业文明-信息文明-智能文明等人类文明形态演进的历史，一定程度上就是科技话语势能转化为意识形态话语霸权的过程。在近代工业革命的背景下，科技进步通过推动生产力的发展，塑造了工人阶级的劳动与生活方式，进而改变社会结构

① 冯冉、陈锡喜：《系统观念视域下新时代社会主义意识形态话语创新的三个着力点》，《湖北社会科学》2023 年第 7 期。
② 卢永欣：《对意识形态的结构功能主义分析》，《思想战线》2012 年第 4 期。
③ 何晓颖、袁芃：《论 ChatGPT 的意识形态属性》，《山西高等学校社会科学学报》2023 年第 12 期。
④ 《马克思恩格斯文集》第 4 卷，北京：人民出版社，2009 年，第 309 页。

和人们的社会关系，间接地影响了人们的思想观念、价值取向和行为模式，成为塑造意识形态的重要力量。西方资本主义国家借助工业革命实现现代化，依靠先发优势跃居全球权力中心，将自己在科技领域的话语权扩散至政治领域。当下，人工智能技术改变了人们获取、处理和传播信息的方式，以数字符号为运行载体，将所承载的技术意识形态载入其中，在算法的操纵下运行。人工智能运用算法推荐实施主流政治话语的有效传达，运用算法规则培育公众对主流价值的认同与坚守，在现实世界具象化为资源配置中的技术权力、公共议题设置中的技术权力、信息传播中的技术权力①，生成一种体系化的隐性意识形态力量。

另一方面，意识形态也制约着科学技术的应用。列宁一针见血地指出："几何公理要是触犯了人们的利益，那也一定会遭到反驳的。"② 科学技术是一种生产力，"在本质上并不属于作为上层建筑的意识形态的范畴，但在科学技术应用过程中，常常难免与意识形态交织纠缠而'难舍难离'"③。政治、文化、道德、法律等意识形态制约着科学技术精神产品的发展和应用，人工智能技术的价值内存及特征决定了其必然会被引入传播与应用的政治场域，参与主流意识形态话语传播和效能的再建构，成为驱动政治建设与国家安全的重要技术力量。在西方国家，以"意识形态算法"生成了相应的"算法的意识形态"，算法黑箱、算法歧视、算法后门等作为"意识形态算法"的技术体现，都在建构自身对原有意识形态的应用，所有的算法都在应用逻辑中体现意识形态性，使科学技术沦为资本增殖的工具和政治操纵的手段。

2．从价值信仰层面看，资本逻辑催生价值失衡

价值-信仰层面构成意识形态的定性系统，用价值标尺对社会事实作出价值判断，为社会提供应然性价值标杆。马克思指出，在资本主义生产方式下，"科学作为一种独立的生产能力与劳动分离开来，并迫使科学为资本服务"④，"科学获得的使命是：成为生产财富的手段，成为致富的手段"⑤。技术异化的根源在于其资本主义应用。人工智能作为技术权力嵌入意识形态的价值-信仰层面，资本逻辑和工具理性遮蔽主流意识形态价值导向，消解主流意识形态价值信仰。

一方面，以资本逻辑为中轴进行话语建构。资本逻辑是指在资本主义社会中资本作为一种特殊的经济力量，对社会生活和意识形态产生的一种逻辑。资本逻辑的核心是追求经济利益和个人利益最大化。马克思正是在批判资本主义"结构化的资本逻辑成为统治一切的力量"⑥ 基础上，形成了对以商品拜物教为核心的资本主义社会意识形态现象的深刻剖析和本质揭秘，揭示了资本逻辑对政治文明的渗透风险。首先，商业目的与资本利益是人工智能技术开发的原初动力，使技术背后隐藏商业意识形态属性。当下，在西方资本主义国家，信息资源日益成为重要生产要素和社会财富，资本成为算法的主体，经济利益和个人利益标准构成了算法选择和判断的标准。如 2023 年初微软向 OpenAI 注资 100 亿美元，这是继 2019 年和 2021 年后第三次扩大与 OpenAI 合作，使其估值在不到 10 个月的时间里从 290 亿美元飙升至 800 亿美元，成为全球最大独角兽公司之一。资本的扩张本性会向每一个新技术点渗透，ChatGPT 爆红助推科技投资市场成为当前资本市场的热门流向。从聊天机器人 ChatGPT 到文本转图像模型 Dall-E，再到近期的文本转视频模型 Sora，人工智能技术与商业资本紧密结合，成为获取更多剩余价值的工具，不可避免地被适应资本增殖的意识形态所渗透。其

① 参见邓伯军：《人工智能的算法权力及其意识形态批判》，《当代世界与社会主义》2023 年第 5 期。
② 《列宁全集》第 17 卷，北京：人民出版社，2017 年，第 11 页。
③ 何茂昌：《ChatGPT 的意识形态风险：样态、肇因及防范》，《西南民族大学学报》（人文社会科学版）2023 年第 12 期。
④ 《马克思恩格斯文集》第 5 卷，北京：人民出版社，2009 年，第 418 页。
⑤ 《马克思恩格斯全集》第 37 卷，北京：人民出版社，2019 年，第 202 页。
⑥ 仰海峰：《〈资本论〉的哲学》，北京：北京师范大学出版社，2017 年，第 307 页。

次，资本逻辑宰制下的算法传播规制，不断挤压主流意识形态话语的传播空间。大数据、大算力运行维护成本高，ChatGPT 推出“Plus”付费升级选项，OpenAI 也推出了语言模型、图像模型、音频模型等 AI 使用付费套餐①。本着“流量逻辑”，人工智能算法推荐以“投其所好”的原则迎合用户，导致各类社会思潮和多元文化充斥于网络空间，使主流意识形态在精准推荐和价值分流中不知不觉被“遮蔽”，削弱主流意识形态凝聚文化认同等功能的发挥。

另一方面，工具理性僭越价值理性话语。工具理性强调效率和经济效益，商业价值是首要考虑因素。实践中工具理性运行方式忽视了人文关怀、道德责任和社会正义等价值观，对当前以人为核心的社会信仰发起了挑战。社会信仰源自特定文化共同体的共同意识，社会信仰的迁移分为两个阶段。在第一阶段，以模型依赖解构主流意识形态权威。进入数字时代，人工智能的“工具化”作用愈发凸显，驱动人工智能发展的核心技术——算法深刻影响人们认识世界和改变世界的能力。工具性的算法媒介，以其强大的工具性价值，使政府、平台企业以及用户个体在长期应用中形成技术性路径依赖。技术本身是无主体意识的工具性存在，但受其所处的复杂社会场域影响。当投资者侧重支持和推广符合自身商业利益、价值倾向的应用场景和功能时，算法的工具理性覆盖甚至取代了价值理性，在数据偏见的包裹下易输出具有意识形态倾向的结果，动摇主流意识形态话语权威。在第二阶段，以模型依赖塑造智能权威。“当科学技术的工具理性张扬到极致，超越其生产力功能，被广泛应用并深入渗透到人们的日常生活及社会政治、文化等各个领域时，科学技术就具备了作为一种隐性意识形态的良好条件”②。智能算法技术内蕴以资本增殖为价值追求的资本逻辑和以工具崇拜为生存目标的工具理性。新技术以效率最大化为目标，作为新变量参与意识形态系统的运行过程，导致工具理性盛行，价值理性式微，甚至以工具理性重塑意识形态的价值理性。如 ChatGPT 等生成式人工智能模型的广泛使用，不仅易使人形成对人工智能的工具性依赖，而且使人对人工智能产生无理由的信任，推动社会信仰向非人类的智能权威迁移，催生“资本崇拜”和“工具崇拜”，消解人们对主流意识形态的价值信仰。

3．从目标策略层面看，主客异位挑战以人为核心的意识形态基础

“目标–策略层面构成意识形态的实践系统，就是要在价值–信仰层面意识形态的指导下，将认知–解释层面的意识形态转换为现实。”③ 人工智能技术作为技术权力嵌入意识形态的目标–策略层面，其运行机制不可避免地聚合为意识形态自觉与价值理性认同之间生态整合的实践形态。

一方面，人工智能技术引发劳动主客异位，催生主体泛在潜藏的主流意识形态认同解构。劳动的发展过程是人类主体性的实现过程。马克思恩格斯强调，“一当人开始生产自己的生活资料，即迈出由他们的肉体组织所决定的这一步的时候，人本身就开始把自己和动物区别开来”④。主体性在劳动过程中逐渐显现并得以确立，劳动方式标志着主体性的具体展开形式。人工智能作为一种生产力手段，在一定程度上实现了通过替代人的非创造性简单劳动来解放人的肢体甚至进一步解放人的大脑，在提高劳动生产效率、促进生产技术革新、拉动生产高质量发展等方面发挥了显著作用，但其所引发的劳动变革预示着劳动者主体地位的进一步下降，引发“劳动主体性悖论”和意识形态话语主体的旁落。首先，人工智能加剧意识形态话语主体泛在。生成式人工智能使意识形态图景逐渐

① 参见姚建军、崔宇航：《从解构到建构：ChatGPT 意识形态性的生成机理及应对策略》，《中共天津市委党校学报》2023 年第 6 期。

② 杨爱华：《人工智能中的意识形态风险与应对》，《求索》2021 年第 1 期。

③ 邓伯军：《人工智能的算法权力及其意识形态批判》，《当代世界与社会主义》2023 年第 5 期。

④ 《马克思恩格斯选集》第 1 卷，北京：人民出版社，2012 年，第 147 页。

呈现主体形态多元化、价值观念多样化的趋势，传统集中统一、自上而下的意识形态传播机制代之以去中心化、自由开放的传播模式。随着人工智能的不断升级，人工智能逐渐展现出超人的技术水平和工作能力，并逐渐成为人们所依赖和使用的一项重要的技术工具。如 ChatGPT 形成的人机互动模式下的融合式主体，是“资本逻辑”与“技术控制”的统一体，在运用中以话语建构、主体塑造和价值判断等方式发挥着意识形态功能，重构意识形态话语传播主体，冲击传统主流意识形态话语权。其次，人工智能加剧了“主客异位”。作为人类体力和智力的高阶延展工具，人与人工智能之间的本质关系仍然是有生命的人与无生命的工具间的主客体关系。“去中心化”是元宇宙的基本规则，任何组织或者个人皆可成为信息发布中心，这种去中心化的特征催生出意识形态的多样化虚拟话语主体。同时，数据的选择、标注以及算法程序的设计所依赖的是一定的价值主体，拥有雄厚资金、人力与技术的科技巨头成为数字空间中潜在的权力主体。在人工智能日益显示出强大的改变世界的功能时，传统的主客体关系异位，将人与人工智能之间的本质关系异化为一种主体间关系，呈现出人的客体化和人工智能的主体化趋势。信息传播主体和客体被冰冷的计算机代码所阻隔，既对原有主流意识形态的主导性和权威性造成冲击，也降低了网络主流意识形态的话语温度。

另一方面，人工智能通过对劳动模式和社会关系的重塑，对人的本质产生影响。意识形态的生产是人的本质的对象化活动的产物，其本质反映出物质资料生产方式的价值诉求。马克思在《关于费尔巴哈的提纲》中将人的本质归结为劳动与“社会关系的总和”①。人工智能技术既丰富了人的本质，又改变了人的本质的具体展开过程。首先，人工智能丰富了人的本质。人工智能是“人的本质力量对象化”的产物。从劳动层面看，人工智能使得人逐渐摆脱动物般的生存机遇而进行创造性劳动，人的劳动内涵更加智能化，丰富和发展了人的本质，人工智能是人的本质的“新确证”。在人工智能社会的三要素中，数据是新生产资料，算法是新生产关系，算力是新生产力，共同构成数字时代的生产基石②。人工智能算法深刻改变了人生命活动的社会性，在一定程度上实现了通过替代人的非创造性简单劳动来解放人的肢体甚至进一步解放人的大脑，通过提高劳动生产率为人类提供了更加丰富的物质和精神文化产品，从而促进人的全面发展，增强了人的社会关系的丰富度、聚合度，生成了精神价值生态。其次，人工智能改变了人的本质具体展开过程。人工智能增加自由时间并彰显人的本质力，其实质是人的劳动的延展形式，人的劳动依然是创造价值的唯一源泉，劳动者的社会生产者主体地位不会改变。人工智能技术为人们的社会交往活动提供便捷的同时，也进一步增强了对人的社会属性的消解。人工智能技术通过资本逻辑与技术逻辑的联姻，形成了贯通“本体世界-生活世界-个体世界”精神链，实现了对人生命活动的社会性的重构，聚合为意识形态自觉与价值理性认同之间生态整合的实践形态。

三、人工智能诱发的隐性意识形态话语风险的化解策略

传统的意识形态管理逻辑主要基于“人-人”“人-媒介-人”“人-大众传媒”“人-互联网”四种交往方式，随着以 ChatGPT 为代表的“人-机”交往方式的广泛性应用，显现出从人际交往向人机交往转变的总体趋势。面对人工智能应用所带来的各种意识形态话语风险，既不能因噎废食而一味地排斥数字技术，也要提升新质战斗力，善用新技术开展意识形态工作。

① 《马克思恩格斯选集》第 1 卷，北京：人民出版社，2012 年，第 135 页。

② 参见中国信息通信研究院：《中国算力发展指数白皮书 2021》，http://www.caict.ac.cn/kxyj/qwfb/bps/202109/P020210918521091309950.pdf。

1. 以技术匡正技术，筑牢人工智能意识形态话语的根基

技术安全是国家意识形态安全的“防火墙”，要坚持“筑坝引流”，推动意识形态防范技术的优化升级，加强技术赋能以抵御意识形态操纵风险。

一要筑技术自主之“坝”，用颠覆性技术保障我国意识形态安全。人工智能技术用“全域式大数据采集+人工编码强化学习”的手段达到观念渗透的目的，需要通过自主创新和技术攻关，推动颠覆性技术与意识形态治理的深度融合，化解人工智能技术给主流意识形态带来的风险。当前，继 ChatGPT 之后，各互联网巨头纷纷计划推出类似产品，如 Meta 推出 BlenderBot，谷歌推出 AI Chatbot，OpenAI 又推出视频生成工具 Sora，苹果也加入生成式 AI 战局。中国企业积极创新研发中国版“ChatGPT”大语言模型技术，百度推出“文心一言”（ERNIE Bot），字节跳动上线“扣子”，阿里巴巴发布生成式 AI 模型 EMO，腾讯、华为也都纷纷有所布局。这场 AI 竞赛没有中场休息，需要努力在人工智能核心技术领域取得颠覆性突破，从人工智能介入意识形态传播的技术路径入手，牵住数字技术特别是算法技术从研发到应用的这个“牛鼻子”，打破“技术霸权”，加快建设国家级人工智能中心，建设一批智能算力中心，建构中国版 ChatGPT 技术标准，设计主流意识形态算法推荐，筑牢人工智能发展的技术根基，规避西方的科技围堵，用颠覆性技术保障我国意识形态安全。

二要引算法向善之“流”，扩大主流意识形态影响力版图。增强技术赋能，优化数据与信息采集，推动全国一体化大数据中心体系建设，在程序研发、设计、运行中嵌入社会主义意识形态，生成式人工智能技术在应用前完成社会主义意识形态数据库的定向训练。通过加强算法的可解释性、可溯源化、可追责化，破解“算法黑箱”，匡正算法技术的内嵌程序，构建符合社会主义主流意识形态价值观的算法推荐。将人工智能嵌入意识形态安全建设的各个环节，建构中国版 ChatGPT 技术标准，加快技术迭代升级，建构中国特色社会主义意识形态传播的新场域。提升智能化监管水平，加快数字技术与网络监督预警、突发舆情危机处理技术等的深度融合，基于智能算法技术建构网络意识形态舆情的人机协同实时监测与智能预警机制，构建算法内容自主生产与算法监控动态平衡的意识形态风险防控策略。

2. 以价值引领技术，巩固规避意识形态话语风险

主流意识形态是智能算法技术建设的精神指引，是网络信息文明发展的基础要素，经主流价值驯化的算法能够推动技术进步和价值体验的共同发展，应从根本上提升主流意识形态对智能算法技术的价值引领力、传导力和辐射力。

一要坚持主流价值与智能算法相融通。在智能算法研发之初将主流价值融入算法推荐的核心技术之中，为算法推荐植入社会主义核心价值观主流价值基因，提高主流价值在算法优先级中的比重，将拥有主流意识形态代码作为节点建立的网络空间结构纳入算法场域，通过智能识别、算法推荐、智能引擎的智能技术充分挖掘主流意识形态的“正能量”，化“流量至上”为“流量向上”，实现“算法推荐×主流价值”的乘数效应，在潜移默化中塑造正确网络舆论价值导向。

二要以技术重置实现对智能算法的内在价值纠偏。在话语内容方面，强化对算法信息筛选的正向价值定位，以社会主义核心价值观丰富数字话语内涵，提升主流意识形态价值相关信息的可见性，防范多元社会思潮渗透。在话语排序方面，改进排序算法，优先供给主流意识形态价值信息，提升主流意识形态优先级序列。在话语推送方面，增添算法“反向推荐”的相关技术设置，打破“信息茧房”的价值偏见，以价值重塑消弭意识形态解构风险。在话语预警方面，积极探索人工智能所蕴含的意识形态属性，善于利用人工智能及时分析、发现、预测网络意识形态领域可能面临的风险并进行有针对性的干预和化解。

3．以规制规范技术，确保人工智能话语表达在法治的轨道上运行

制度是确保安全观念转化为意识形态安全治理效能的根本保障。针对人工智能，我国已形成《中华人民共和国数据安全法》《生成式人工智能服务管理办法》《互联网信息服务算法推荐管理规定》《新一代人工智能发展规划》等制度性规范，为统筹人工智能技术创新与风险防控提供了重要基础性遵循。要完善人工智能技术规约，在充分评估其文本数据局限性、算法推演偏见性、意识形态倾向性基础上，从数据、算法、算力三个维度进一步完善人工智能技术的法律、规章、制度，进一步规范制定行业标准，划定技术适用范围和职责权限，推动将主流意识形态作为技术研发推广的核心融入框架搭建、代码开发、运算排序等方面，确保人工智能话语表达在法治的轨道上健康运行。完善人工智能技术监管，制定算法备案与披露制度，建立人工智能内容监管标识制度，完善意识形态敏感领域应用规范与强制定向训练制度，为 ChatGPT 类生成式人工智能意识形态风险的规制提供全程监管。加强研判智能技术风险，对人工智能可能引发的意识形态风险的样态表征、生成动因、规制策略作出科学研判，为资本设置“红绿灯”，守住意识形态安全底线。

4．以人本消弭技术错位，建设与人工智能发展相匹配的大众话语

人民性原则、人本逻辑是社会主义意识形态建设的根本立场和价值取向。人工智能时代，要坚持工具理性与价值理性相统一，妙用人工智能技术“精准对接”话语体系，增强主流意识形态话语渗透力。

一是打造生活话语体系。运用智能技术优化主流意识形态话语议题设置，整合宏大叙事与细微叙事、精英叙事与大众叙事、文本叙事与生活叙事、学理叙事与通俗叙事，借力算法技术推动宏大理论走向日常生活话语①。

二是善用网络圈层话语体系。借助智能算法技术，对“文艺圈”“学术圈”等不同圈层的年龄、职业、话题等相关数据进行解析并找到话语生成规律，构建起适应不同圈层的话题偏好、语言特色的话语体系，构建起适应不同圈层的主流意识形态话语体系。

三是巧用定制话语体系。在内容方面，通过聚类分析，掌握不同群体对主流意识形态话语的表达方式和表现形式的喜好程度，并有针对性地“投其所好”推送理论性话语和叙事性话语。在形式方面，以智能技术助推叙事呈现，巧用大数据精确匹配话语方式，对受众在文字、图片、视频等不同话语表现形式上的停留时间等数据进行统计，预判出受众更倾向于哪种话语表现形式，从而有针对性地进行话语定制，实现话语内容与话语形式相融通，增强主流意识形态话语感染力。

从大历史观看，从地理大发现到当前的人工智能时代，人类每次技术革命的进步都带来了时代的变迁，意识形态工作的内涵和外延也在不断丰富拓展。人工智能技术存在隐性意识形态话语风险，我们既要筑“坝”引“流”，驾驭人工智能技术，又要化“流量至上”为“流量向上”。从技术赋能、价值导向、制度规制、人民立场层面寻求化解风险的策略方法，实现“人工智能×主流价值”的乘数效应，借助人工智能等技术带来的时代红利，牢牢掌握人工智能时代意识形态工作话语权。

参考文献：

［1］习近平：《加快建设科技强国　实现高水平科技自立自强》，《求是》2022 年第 9 期。

［2］侯惠勤：《意识形态话语权建设方法论研究》，《中共贵州省委党校学报》2016 年第 2 期。

［3］刘皓琰：《数字帝国主义》，北京：中国青年出版社，2023 年。

（编辑：荀寿潇）

① 参见包崇庆、柏路：《人工智能时代主流意识形态话语权建设的多维审视》，《传媒》2023 年第 13 期。