

人工智能时代意识形态风险的生成逻辑 与科学防范

刘伟兵

(复旦大学马克思主义学院,上海 200433)

摘要:人工智能之所以具有智能,很大程度上依赖于机器学习技术。机器学习技术通过数据的训练、算法的设计及模型的运行,能够将问题转化为数学统计意义的求解,进而赋予人工智能技术逻辑以数理逻辑。在社会化协作的精神生产基础上,人工智能技术生成了“意识”现象,并在资本应用、技术滥用等应用逻辑的推动下生成与之相适应的意识形态,由此建立在人工智能技术新特性、新规律和技术逻辑基础上的意识形态也面临着由人工智能技术规定带来的新的意识形态风险可能。人工智能时代的意识形态风险在意识形态生成领域、传播领域和认同领域分别呈现出“资本逻辑、算法黑箱与算法伦理”“算法推荐、过滤气泡与信息茧房”“亚文化空间、价值观渗透与数字拜物教”等具体风险表象,不同阶段风险在现实运行中交错杂糅,构筑出复杂的意识形态风险网。基于对人工智能时代意识形态风险生成逻辑的全面把握和风险现象的系统梳理,可有针对性地通过发挥社会主义应用逻辑根治人工智能时代意识形态风险根源,通过推进系统规制和技术革新达到人工智能时代意识形态风险的技术防范,通过营造良善生态等多元治理达到人工智能时代意识形态风险的系统性防范,这也是中国式现代化在人工智能时代创造数智文明的意识形态领域尝试。

关键词:人工智能;算法;意识形态;意识形态风险;资本逻辑;

中图分类号:D64

文献标志码:A

文章编号:1671-4970(2024)06-0010-10

当前,生成式人工智能的出现和快速发展使人们看到通用人工智能出现的可能。其中,ChatGPT、Midjourney、Auto-GPT等生成式人工智能以自动生成图像、文字、音乐、视频等方式,改变了意识形态生成路径,并进而具有引发新的意识形态风险的可能。目前,学界对这一问题研究存在多种视野:包括关注人工智能时代意识形态风险现象、人工智能意识形态风险危害及其防范等。学界围绕人工智能时代意识形态风险问题,在多维的共建研究中也逐渐形成

共识,那就是要先回应人工智能时代意识形态生成路径,才能有针对性地分析意识形态风险及其防范研究。因此,系统厘清人工智能的技术逻辑、全面把握人工智能的应用逻辑,进而深入研究人工智能生成意识形态的路径及其带来的意识形态风险,并进行针对性、科学性防范就具有重要的理论意义和实践价值。

一、人工智能时代的意识形态生成

以往对人工智能与意识形态的探讨往往依

引用本文:刘伟兵.人工智能时代意识形态风险的生成逻辑与科学防范[J].河海大学学报(哲学社会科学版),2024,26(6):10-19.

基金项目:国家社会科学基金青年项目(23CKS053)

作者简介:刘伟兵(1994—),男,副教授,博士,主要从事人工智能时代马克思主义时代化研究。E-mail:wblu595@163.com

循唯物史观的研究范式,即从人工智能带来的生产生活的智能化实践变革入手,把握作为上层建筑的意识形态在其实践基础和经济基础智能化变革后的生成规律。然而,随着生成式人工智能的出现,人工智能直接参与并促进了精神生产、精神交往的智能化变革,因此,对人工智能与意识形态的研究要立足于对人工智能技术的把握,明晰何为人工智能以及人工智能的技术路径与技术逻辑,从而才能从技术出发,客观、全面地把握人工智能与意识形态的关系。

1. 人工智能的技术路径及逻辑

人工智能已经成为当前引领第四次科技革命的主要技术突破。“人工智能是引领这一轮科技革命和产业变革的战略性技术,具有溢出带动性很强的‘头雁’效应。在移动互联网、大数据、超级计算、传感网、脑科学等新理论新技术的驱动下,人工智能加速发展,呈现深度学习、跨界融合、人机协同、群智开放、自主操控等新特征,正在对经济发展、社会进步、国际政治经济格局等方面产生重大而深远的影响。”^[1]当下的人工智能虽然还处于弱人工智能阶段,但是已经日益凸显出其技术特征,并明晰了发展趋向。

人工智能技术图景主要包括机器学习、大数据、计算机视觉、语音识别、自然语言处理等5个部分,其中的后4种技术主要依赖于机器学习技术,机器学习技术构成了人工智能之所以智能的“人工”关键。理解机器学习技术,需要掌握其中最为关键的三组核心概念——数据、算法、模型:第一,数据是人工智能所有技术最为基础的单元。人工智能是一种数字智能,而智能文明本质是一种数字文明,因此,数据被称为人工智能时代的石油。“浩瀚的数据海洋就如同工业社会的石油资源。”^[2]数据不是数字,而是以数字形式承载的信息,包括图像、音乐、视频等。数据在原始数据阶段并不能直接用于机器学习,而是需要对其进行标注、分类、洗练,生成结构化或者半结构化的数据才能被机器识别,只有如此,数据才能成为构成人工智能的前提。然而,训练后的数据本质是一种“感受性关系”^[3],即能指和所指统一的符号结构。此时,数据是一种静态的结构,需要算法手

段才能生成智能表现。第二,算法处理数据的技术路径包括“监督学习”“无监督学习”“半监督学习”“强化学习”共4种类型,本质是通过代码设计一个处理数据的程序,是对数据的数字劳动,这构成了人工智能之所以智能的关键。所以说,“计算是人工智能的本质”^[4]。第三,模型是机器学习算法输出的结果,是由模型数据和预测算法组成的特定数据结构,代表机器学习算法“学到”的东西,构成了人工智能的智能形态。所以,计算机视觉是对图像数据的采集和训练,语音识别则是对声音信息的数据采集和训练,是通过智能算法设计出一个数据模型,进而实现多模态数据融合的现实交互。可见,人工智能的技术路径,就是通过数据训练和算法设计生成一个模型,并在模型的求解过程中表现出解决问题的智能。

由于人工智能的本质是计算,核心关键是包含数据、算法、模型的机器学习技术,使得人工智能技术逻辑的本质是一种数理逻辑,就是把某个问题以数字的方式在模型论、证明论、集合论、递归论等方式中获得确定性。这就造成了人工智能作为技术展开的过程,受形式逻辑的同一律、矛盾律、排中律和论证要充足理由等规律限定,所以,人工智能的“智能”是一种统计学意义上的确定性智能,重形式而不重内容,得出的所有结果必然要有依据,且不存在模糊空间。

2. 人工智能生成“意识”现象

在机器学习技术的快速发展下,人工智能的语音识别、自然语言处理等技术得到了快速发展。生成式人工智能的出现,使得人工智能技术直接参与精神生产和精神交往,可以自动生成文字、图像、视频、音乐等文化产品,带来了意识生成路径的革命性变革。历史唯物主义认为,个体意识是社会结构的产物,是在个体与社会实践交往的过程中逐渐生成的。“‘我’已经成为文化体系所改造了的‘他者’了。”^[5]以往社会意识对个体的意识赋予,是通过个体的“精神劳动”^[6]⁵⁵¹实现的,而随着人工智能技术的发展,尤其是生成式人工智能的出现,使得意识生成的精神生产由个体劳动走向了社会化协作的阶段。“许多人在同一生产过程中,或在

不同的但互相联系的生产过程中,有计划地一起协同劳动,这种劳动形式叫做协作。”^{[7]378}

随着人工智能技术的发展,生产生活的智能化会出现一批专门从事数据采集、标记、洗练的劳动者,也会出现一批专门从事算法研发的人员,这就意味着人工智能技术看似“无人化”“自动化”生成“意识”现象的背后,是社会化生产的合力。而且受人工智能内在逻辑的限定,人工智能生成的意识也要满足形式逻辑的同一律、矛盾律、排中律和论证要充足理由等规律,本质是数理统计求解的“确定性”。正因如此,人工智能生成的意识是一种“确定性”的意识,具有必然性的意识,也就不存在人们自身个体才具有的灵感、直觉等偶然性意识。在此意义上,人工智能生成的意识虽然是人的意识通过社会协作方式的延伸,却少了人的意识的创造性,只是经验性事物的统计体现。但是,人工智能技术在追求一种“确定性”意识的过程中,为何会具有算法偏见、算法偏好等意识形态现象?作为本应中立、客观的技术本身,为何会具有意识形态立场?人工智能具有自由意志“人格”吗?事实上,人工智能的意识形态命题,应该要深入到人工智能的应用逻辑去找寻答案。

3. 人工智能应用与意识形态生成

马克思曾看到了技术逻辑和技术应用逻辑的区别,认为“工人要学会把机器和机器的资本主义应用区别开来,从而学会把自己的攻击从物质生产资料本身转向物质生产资料的社会使用形式”^{[7]493}。人工智能之所以会生成意识形态,需要从人工智能应用逻辑入手,明确人工智能如何在应用过程中生成意识形态,既包含了生产生活智能化变革,促进意识形态上层建筑变化的唯物史观回答,也包含了精神生产和精神交往智能化变革的技术应用路径探讨。

一方面,人工智能技术的资本应用生成符合资本增值需要的意识形态。资本逻辑就是资本增值的逻辑,表现为将一切生产纳入资本的总体生产环节,转化为资本增值。人工智能在资本应用下的技术生成,表现为数据、算法、模型的资本化。数据的资本化,就是数据的注释、选择、洗练都以资本增值为目的和标准。“大数据的抓取、清洗与凝练等过程,背后的主体有

自身依据和目的。”^[8]这就造成了人工智能在存在前提上具有符合资本增值的意识形态性,如此一来,人工智能生成“意识”现象的最基础元素都具有意识形态的偏见。在此基础上,则是进一步的算法资本化。算法本身就是对数据选择、处理的程序,这就意味着算法包含选择、处理的标准和目的,本身就具有一定的价值性和目的性。“任何算法都不是凭空产生的,不是‘价值中立’的,都内蕴着一定的价值负荷和价值选择。”^[9]这种价值性和目的性原本是由算法工程师的主体性所决定的。算法工程师往往执行的是资本公司的意愿,受资本增值的考核,这必然使其意识呈现为满足资本增值的意识。换言之,只要算法工程师不掌握生产资料,受资本雇佣劳动,那么算法工程师的数字劳动必然会纳入资本生产环节,贯穿资本增值的逻辑。在数据、算法的资本化基础上,就形成了一个属于人工智能技术的意识形态。模型的资本化就是生成一个属于资本增值的意识形态数字结构,所有在此模型中得到的结果必然都具有资本的意识形态偏好,并在周而复始的算法求解过程中实现资本意识形态的生产与再生产。

另一方面,人工智能技术的技术滥用生成人工智能的科技意识形态。人工智能在资本逻辑的推动下,推进了生产生活方式的智能化变革,也造成了人工智能技术的滥用,从而在人工智能技术逻辑的限定下生成了新的科技意识形态。资本为了更快增值,必然要赋予和塑造人工智能新的合法性,让人工智能成为人们日常生活中的唯一确定性。人工智能技术滥用的展开过程也是人工智能自身存在的数理逻辑展开的过程,是理性遮蔽感性、思维形式代替思维内容的过程。随着人工智能技术的滥用,会造成价值理性的衰落与工具理性的崛起,出现“社会系统的发展似乎由科技进步的逻辑来决定”^[10]的意识形态假象。而思维形式代替思维内容,则是量化、科学、是非等形式逻辑的范畴成了人们思维的主要选择,如此一来,生产生活智能化普及程度越高,人们就越会关注立场却不关注观点,注重形式而不注重内容,在乎结果而不在乎过程,从而呈现思维单一化、片面化的意识形态问题。所以,人工智能技术的滥用使

得人工智能技术成为一种新的合法性来源,人们将人工智能等同于科学、等同于权威,事实上就是生成了一个可遮蔽其他事物的新科技意识形态。

二、人工智能时代的意识形态风险

明确人工智能会生成“意识”现象,并在应用逻辑下会生成意识形态,这就带来了一个全新命题,即人工智能在生成意识形态中带来何种意识形态风险。对这一问题的回答,要放置于意识形态生成、传播、认同的全链条中去把握。

1. 生成领域:资本逻辑、算法黑箱与算法伦理

在意识形态生成方面,人工智能技术发展背后潜藏的资本逻辑和算法权力会直接影响人工智能技术的发展走向与未来运用,算法的黑箱特性也增加了意识形态走向的不确定性,并在此过程中引发了一系列新的算法伦理问题。

第一,资本增值逻辑是算法权力的最初动因。互联网平台在追求资本最大化的过程中,不可避免地会对公共利益产生影响,乃至对社会权力结构造成冲击。马克思认为,不能“总是仅仅从外在的有用性这种关系来理解”^[6]¹⁹²技术的本质。韦伯说:“权力意味着在一种社会关系里哪怕是遇到反对也能拥有贯彻自己意志的任何机会。”^[11]算法权力亦如此。算法权力以技术发展为基础,对社会数据进行控制,从而构建起一种“公权力”,体现出自有的特征:一方面是工具理性的技术统治。在技术导向之下,理性变成了纯粹的目的,由算法对人的行为动态进行判别,界定个体的意义价值,使算法从一种知识体系演化为意识形态、统治形式,并消弭人在现实活动中的能动性和主体性。纯粹的技术理性为实现社会治理现代化所追求的“有温度的社会”增加了难度。另一方面是资本逻辑的秩序安排。算法的研发与运营需要各类资本的巨大投入,资本的提供者和持有者成为算法运行的实际控制者,而依附于市场的算法技术能够帮助资本实现特定的意愿和诉求。资本逻辑逐渐超越治理逻辑,追求剩余价值最大化成了实际过程中的重要方向,必然会导致社会

利益的分化以及人们对于算法存在合理性的担忧。

第二,算法是人工智能系列程序编程的幕后基础。在完成大数据搜集整理后,深度学习算法进入数据处理、训练过程,进而生成更加熟练的学习能力,再输出相应的决策运算结果。但是在这一过程中,是无数计算机语言代码组成的自然语言,夹杂着烦琐的数据材料和运算规则,对于绝大多数人而言,这是一个无法洞悉的“算法黑箱”。在“黑箱”之下,人们面对的只是单纯的结果输出,而对于解释过程却无法得到清晰的认知,其幕后可操控的未知风险陡然增加。所选择的数据、编写的代码都是算法工程师个人为人工智能技术灌输进去的,背后体现出的价值判断和行为取向也归属于他个人。

第三,算法技术构建的黑箱困境为社会秩序规则带来新的伦理风险。在算法的推演下,人群固化和社会偏见会进一步加剧,由此可能引发更为激烈的现实矛盾。社会不同领域产生的数据种类繁多、信息量庞杂,算法能够通过系列程序掌握每一个体的个人偏好乃至隐私数据,对于信息数据的过度使用乃至泄露、曝光,无疑是对社会伦理关系的一颗“定时炸弹”。算法程序的设计是社会协作的过程,由社会化的数字劳动完成,但因此产生的新的理念和方式并不完全属于整个社会,而只属于其中生产资料的所有者及资本,即社会化生产与个人占有矛盾的人工智能形式。人与技术逐步迈入了“人机共生”的双主体关系,人类作为“元设计者”,技术则是“设计者”^[12],在此关系之下,人类必须提防技术奇点的到来。人工智能的技术奇点意味着智能技术能够根据人类的需求和现实对操作系统进行自主选择与优化,从而出现智能技术对于人类的自主教育。由此,人机伦理关系将产生较为彻底的改变,在未来很长一段时间都需要对“人机共生”议题的是否可能发生、应当如何应对等进行持续分析。

2. 传播领域:算法推荐、过滤气泡与信息茧房

面对算法,人们不仅处在“算法黑箱”之内,还会逐渐被“信息茧房”包裹。在算法程序的背后是基于用户偏好的推荐机制,从而加强

用户对于信息的黏性。算法反复推荐用户感兴趣的信息,用户所接受的信息呈现单一趋势,获取信息路线呈现窄化趋势,从而被困在智能算法编织的“过滤气泡”。当“气泡”不断强化,则会形成“茧房”般的束缚,使得用户观念在算法放大之下不断被固化。观念固化会让人们在看似开放、自由的智能环境下产生偏激、片面的认知。一方面,人工智能算法以隐蔽、微观的方式在社会生活系统中发挥控制力量,潜移默化地对个体的思想观念、情感表达和思维模式进行有效渗透和长久规训。在日常生活中,算法主要通过基于内容、基于知识、协同过滤和混合推荐等类型将用户的浏览历史、搜索记录等数据记忆贴上不同标签归类,为用户制作独一无二的立体画像,进而可以测算出用户的需求、兴趣和偏好,为其持续性推荐同类型或高度相似的内容。另一方面,一旦存在偏见的数据被用于机器学习,加之人工智能技术具有自我训练的倾向,算法运行的结果也将带有偏见,从而在循环往复的算法运行过程中,人工智能将把人类社会的原始偏见不断放大、持续扩散,成为具有强影响力的社会偏见。所以说,算法虽然是人工智能运行的灵魂,但引导算法在正确轨道上运行是更为关键的难题,不能让算法具有明显意识形态倾向,从而对社会发展产生消极影响。特别值得关注的是,人工智能的算法推荐虽然智能化水平很高,与此同时也将信息更为碎片化、微小化地,在社会网络中进行传播,这种传播方式虽然有助于整个社会信息配置效率的提高,但也在一定程度上削弱了主流意识形态的优势传播地位,降低了主流意识形态的实际作用效果。

身处人工智能时代,除了现实的身份认证,人们还在互联网等场域以“网民”乃至“互联网原住民”的形象进行社交,社交过程中会有大量碎片化信息充斥于社交圈群。在这些圈群里,时刻都在上演着——“多元化的网络意识形态在大众传播形成的信息环境即‘拟态环境’中交流、交融、交锋以至交战”^[13]。现如今,微博、微信、抖音、小红书等社交媒体平台和资讯聚合终端“走红”,人工智能时代下的信息分发已由“人寻找合适的信息”转变为“信息寻

找合适的人”,用户偏好与海量信息的精准匹配是现代普遍化的传播模式。进一步地,依托这些发展起来的新模式和新业态为社会主流意识形态的传播带来诸多挑战。以用户偏好为导向的算法推荐造成意识形态传播链条的复杂化和不确定性增强,“过滤气泡”强化效应加剧了人们的价值观分化和社会交往的隔离,这都对主流意识形态和传统主流媒介的生存空间形成挤压,整体上表现出较为明显的意识形态传播风险。

3. 认同领域:亚文化空间、价值观渗透与数字拜物教

经过生成与传播,个性化推送和微型叙事对主流意识形态认同产生影响。身处于信息加速、速度至上的现实,人们对于信息内容所蕴含的意义价值呈现浅表式的关注和掠影式的印象,越来越注重个体本身是否与信息之间存在联系,由此形成的是一种短暂的主观感觉,而不是对社会公共价值的长久关注和追求。

一方面,一旦受众接受了偏离主流的意识形态观点,在算法推荐“投其所好”式“溺爱”下,很可能会在“亚文化”群体中的错误轨道上越走越远。“当筛选的力量没有限制时,人们能够进一步精确地决定,什么是他们想要的,什么是他们不想要的。他们设计了一个能让他们自己选择的传播世界。”^[14]那些被信息窄化和裹挟进入“亚文化”圈子的人们,平常在自我的“小众”天地中可以无所顾忌地发表看法和言论,以获得存在感,但在公众场合,却往往“暗淡无光”地低下头去。这种“恶性循环”会使被算法“捕捉”的人们难以融入社会生活,并可能引发心理健康和个人成长等方面的问题。另一方面,人工智能技术在释放出巨大“技术红利”的同时受到资本逻辑和工具理性的作用,在资本增值“永远的不安定和变动”^{[7]560}的驱使下,智能算法技术不断侵占主流意识形态的生存空间,主流意识形态的价值导向效果也被削弱,这些因素将对人们的信仰观念产生不利影响。如,一些媚俗低级、粗制滥造的价值导向依赖用户黏性、争夺流量和多元思潮的方式“出圈”,反而获得了广泛吹捧,导致公众价值取向的偏向性越来越明显,不断蚕食和消解着主流意识

形态的权威地位。特别是,凭借强大的技术支持和资本优势,一些西方国家通过构建高效隐蔽、灵活多变的“黑手”来对我国进行意识形态的技术渗透,企图以这种方式达到规训受众意识形态认同的目标。在此背景之下,部分意识形态认同尚待建立或者处于不坚定状态的人们,就容易在认同上产生动摇,可能会对社会主义意识形态产生偏见和误读,对马克思主义意识形态的领导权和话语权构成威胁。

更深层次的风险是,在资本与技术的“合谋”之下,意识形态认同层面会出现数字拜物教的风险。一系列数字技术的发展、运用和传播让人们陷入了前所未有的数字拜物教的幻境当中,眼前所见和心中所想都可以通过数字的形式进行定义和解构,长此以往,数据与数据之间的关系会重塑人与人之间的关系,人们不由自主地就会臣服在无所不能的“数字”脚下。“数字拜物教并非人对数字的主观崇拜,而是资本的权力机制统治和驯服人的意识形态工具”^[15]。在数字资本主义的条件下,以“数据”为生产资料,以数字技术为中介,社会整体对数字技术和数字价值膜拜与追崇形成了一整套数字逻辑建构起来的社会规则。这套社会规则实现了对人们更为精准和全面的控制,隐蔽且有序地促进了资本权力的发展。

综上所述,人工智能时代意识形态风险的生成因算法黑箱的不确定性而更具不确定性和多变性:在传播环节,算法的智能推荐使得意识形态风险更加隐蔽;在认同环节,甚至出现了数字拜物教的新型意识形态风险形式。此外,各环节虽依自身规律形成独特风险,但在人工智能技术的整体运用中,这些风险会交叉、并联、整合,所生成的风险也会随技术的推广而扩散。这就意味着,相较于以往,人工智能时代的意识形态风险在技术上更易相互贯通、交织,进而编织成一张复杂的意识形态风险网络。因此,必须综合施策,才能够提高人工智能时代意识形态风险防范与治理的实效。

三、人工智能时代意识形态风险的科学防范

基于意识形态生成、传播、认同全链条在人

工智能时代里的新规律、新现象及新风险,必须把握“航向”、守住“内核”、构筑“屏障”,实现对人工智能时代意识形态风险的科学防范。

1. 把握“航向”:发挥社会主义应用逻辑,根治风险之源

人工智能在资本应用、技术滥用的应用逻辑规定中,受新技术的特征影响,会在生成、传播、认同全链条中生成各种意识形态风险。因此,对人工智能时代意识形态风险防范与应对的“治本之策”就是建构一个全新的人工智能应用逻辑,让人工智能充分应用在向善的实践之中。那么,这个新的应用逻辑是什么?马克思在深刻批判资本主义应用逻辑基础上,早已指出了一个通往未来社会的新的应用逻辑,那就是社会主义应用逻辑,表现为坚持党的领导、马克思主义中国化的逻辑规定和社会主义核心价值观的践行导引。

坚持党的领导是防范人工智能时代意识形态风险的根本保证。“加强党的领导作用,就是通过党的领导来规制算法的技术风险和资本应用风险。”^[16]中国共产党是使命型政党,是接受马克思主义思想指导的无产阶级政党,具有纯洁性、先进性。中国共产党既具有共产主义远大理想,又有中国特色社会主义共同理想,既有为中国人民谋幸福、为民族谋复兴的初心和使命,又有严密和先进的组织体系。如此一来,中国共产党就具有了追寻和实现人的自由全面发展的自觉性的能力保证。正是如此,只有坚持党的领导,才能有一个更加强有力的力量驾驭资本逻辑,规避资本逻辑的侵蚀,保证人工智能的“向善”发展。

马克思主义中国化是防范人工智能时代意识形态风险的发展逻辑规定。马克思主义本身就是资本现代文明内在对抗性的产物,是建立在对资本主义现代社会充分批判与科学揭示基础上的科学理论,其目的指向就是实现人的自由全面发展。马克思主义中国化是通过“两个结合”,实现科学理论在中国具体实践中的再阐释、再建构。与之伴随的是马克思主义中国化与马克思主义化中国的历史进程,在历史实践过程中发展马克思主义、践行马克思主义。正是在此基础上,中国才能探索出社会主义基本经济制度,才能实现“社会主义+市场”的创

新结合,才能在世界历史进程的现代化普遍规律中,探索出既是现代化又超越现代化的中国式现代化。人工智能作为社会化生产的产物,只有在社会主义制度里才能规避或者减少与私人占有的矛盾。在此意义上,只要在马克思主义中国化的发展逻辑规定下,人工智能技术的发展逻辑指向就是为了人的发展,而不是为了资本增值。一旦资本逻辑控制人工智能为了增值而制造特殊利益与普遍利益矛盾的意识形态风险时,在马克思主义中国化逻辑中就具有干预科技发展方向的合法性与正当性。

践行社会主义核心价值观是防范人工智能时代意识形态风险的价值导引。“超越舆论场的资本逻辑悖论与局限,理应以社会主义核心价值观规约资本在舆论场的投资方向、运作过程等,引导资本实现对人的价值关怀,实现资本的良性循环,引导资本逻辑服务于舆论。”^[17]相较于资本应用,社会主义应用逻辑具有历史自觉的价值导向,是合规律性与合目的性统一的体现。作为抽象思想观念的价值引领并不具有约束力,因此在社会主义应用中,抽象的理论是通过具象的方式得以转化和践行的,其中,法治、公正、诚信、和谐等核心价值观应该通过制度化的建设与践行方式融入人工智能时代意识形态风险防范。如,开发人工智能应用之前应该要先进行立法规范,以及新型人机关系伦理冲突应该诉诸法律等法治理念;人工智能的数据、算法、模型等技术的设计过程应该在设计原则中体现公正、诚信、和谐等价值理念,进而引导人工智能技术研发与运用步骤的价值规范性等,由此避免算法黑箱等风险。

2. 守住“内核”:推进系统规制,推动技术革新

人工智能之所以在当下产生如此广泛的作用范围和现实影响,本质上就是科学技术创新的结果,是在技术层面实现了不断突破和磨合的产物。归根结底,对人工智能时代意识形态风险进行防范,要在技术上做文章,从数据、算法,到建模、运行,系统规制人工智能技术全过程。

第一,规范数据收益及其风险界定。数据在进入算法程序之前就已经需要对其可能存在

的风险、产生的收益及应对的方案进行界定和明确,即在数据输入前需要对价值进行纠偏。诚然,算法的复杂性让数据在实际导入应用时具有突出的难以预料性,但是依旧需要尽最大可能判定相关数据参与人工智能技术可能出现的意识形态风险的范围、概率及影响程度。如果这些数据难以预料可能的实际风险或预料到会有较大风险,应该加倍谨慎地、有限度地在算法当中使用,甚至应该一定程度上对这些数据进行剔除,或者待相关应用算法成熟之后再逐步导入这些数据。如果目前预测这些数据本身在意识形态上的风险较小或已经形成了完备的风险应对方案,可以考虑进入算法进一步使用,但需要在推行过程中加强监测和调研,避免意外风险和特殊情况的出现。因此,对于数据收益与风险的解释和测试,是数据真正导入算法之前的必要环节。

第二,优化算法过程及其程序设计。算法设计是对算法运算过程、模型机制的规定过程,这是智能技术得以“动起来”的内核。对算法设计进行优化需要基于对算法权益与风险的评估,进而能够更加有效地在“运动过程中”避免意识形态风险。一方面,算法设计会受到设计者个人体验、知识水平、社会文化等的影响,也需要注入使用者的现实使用需求,需将“以人为本”的理念贯穿算法设计。以人为本、人民至上都是在强调尊重人的发展规律、切实考虑人的现实感受,能够根据不同人群的服务需要提供与之匹配的路径和方式,这是算法设计过程中要遵循的要求。另一方面,可以设计多重链条结构的算法程序互相加以抗衡。从技术角度来看,人工智能时代意识形态风险问题就是一种技术问题,技术问题可以通过技术性策略来加以对抗。“以算法对抗算法是借助算法从技术之维来实现算法治理的程序化牵制,防止算法失控对社会的异化与吞噬。”^[18]通过打造互相牵制的算法运行机制,如算法设计过程中的价值引领机制、算法测试过程中的错误纠偏机制、算法审核过程中的筛查核验机制、算法推广过程中的监测叫停机制等,从而形成完整贯通的算法制衡格局。实现算法制衡,能够在很大程度上使程序设计过程中不断降低信任危

机、算法偏见、社会歧视等意识形态风险出现的可能,也能够有利于打破信息茧房,实现算法的公正运行。

第三,推进技术审查及其运行问责。通过智能算法设计出数据模型后,面对模型运转可能导致的意识形态风险,需要在有关部门主导下建立完善的规章制度条文和道德约束体系,加强对人工智能技术的问责。对于人工智能技术的问责,要求数据提供者、算法设计者、模型所有者等有相关责任的技术开发人员提供相关的解释和说明,如数据从何而来、算法如何设计、模型应用何处、责任归属于谁等关键问题。并且,这些人员有必要对可能或已经造成的风险或损害负责,及时补救相关问题,从而逐步构建起一套“负责任的人工智能”制度体系。政府部门在此过程中需要积极担当作为,在问责过程中建立起总体的问责框架,并让开发、应用算法程序的企业或团体在框架之下参与给出拟定具体准则内容的建议,将责任嵌入人工智能技术推广使用的各个环节,进而不断明确各个主体的责任边界,把稳健性、安全性、公正性和公开性等要素真正纳入整个系统规制过程,推动人工智能时代主流意识形态良好秩序的构建。只有如此,才能打破“算法黑洞”等技术特性带来的意识形态风险。但需要注意,持续的问责、追责不是破解人工智能技术风险困境的“灵丹妙药”,而是更需要在过程中基于主流意识形态形成自发的价值认同和道德自律。

3. 构筑“屏障”:多元综合治理,营造良善生态

人工智能终归要在人们的现实生活环境中被使用,才能发挥其开发设计的实际价值。人工智能也需要社会治理主体的协同参与,共同下好“一盘棋”,这样对于意识形态风险防范才有更佳的普遍性效果。

一方面,发挥法律制度的硬性规范作用。目前,我国已经出台了《互联网信息服务管理办法》《新一代人工智能伦理规范》《新一代人工智能发展规划》《互联网信息服务算法推荐管理规定》等条例规范,《生成式人工智能服务管理暂行办法》也在 Chat GPT 问世之后不久出台,对加强意识形态安全工作提出了明确要求。

这些法律法规和条款规章对人工智能相关平台终端的开发和用户保障进行了规定,为人工智能发展设立起“红色警戒线”。如,《互联网信息服务算法推荐管理规定》要求“算法推荐服务提供者应当落实算法安全主体责任,建立健全算法机制机理审核、科技伦理审查、用户注册、信息发布审核、数据安全和个人信息保护、反电信网络诈骗、安全评估监测、安全事件应急处置等管理制度和技术措施”^[19]。党内也颁布了《党委(党组)意识形态工作责任制实施办法》《中国共产党巡视工作条例》等涉及意识形态的党内法规,不断压实意识形态工作主体责任,各部门汇聚力,构建“共商共管,共防共治”的风险防范治理格局。不仅如此,还需继续根据人工智能的技术特性和具体实践应用,及时更新掌握可能出现或已经出现的问题,动态健全并持续优化相关的法律法规,全面推进和完善人工智能和意识形态的立法、执法、司法、守法工作。与此同时,结合近年来相关互联网平台和应用程序的使用人数陡增的现象,将其作为普法的新途径和新方式。如,在网络直播间打造普法连麦互动课堂、在网页头条等显眼位置设置相关法律内容宣传语等,进一步创新网络普法的话语表达方式,用人们在网络空间中喜闻乐见的方式进行体验式、参与式的生动传播。这些都能够对人工智能时代公众的思想认知看法、内在价值理解、外在表现方式等进行正确引导,扬弃人工智能时代的“数字拜物教”的异化现象。

另一方面,发挥数字素养的柔性教化作用。伴随人工智能技术的推广普及,数字素养不应再是少数人所具备的能力,而已经日趋成为世界各国对于民众素质教育的重要内容。也就是说,不能囿于传统视角和思维方式,而是要充分融入人工智能时代,提升对其相关技术应用的适应能力乃至驾驭能力。具备良好的数字素养有利于生成对人工智能时代意识形态的正确判别,并安全、合法、符合道德规范地使用信息技术。《关于加强科技伦理治理的意见》《数字中国建设整体布局规划》《提升全民数字素养与技能行动纲要》等文件的出台显示出数字伦理与道德已成为数字素养教育的重要内容,如“高等学校、科研机构、医疗卫生机构、企业等

单位要履行科技伦理管理主体责任,建立常态化工作机制,加强科技伦理日常管理,主动研判、及时化解本单位科技活动中存在的伦理风险”^[20],以及“提高全民数字获取、制作、使用、交互、分享、创新等过程中的道德伦理意识,引导全民遵守数字社会规则,形成良好行为规范”^[21]。综上所述,提升全民数字素养与技能的过程,也是加强主流意识形态建设的过程。为此,要营造良好的社会学习氛围和良善的价值环境,打造具有高素质的人工智能创新人才队伍,加强对科技人才技术研发过程中的向善伦理教育和价值塑造,以此带动提升全民的整体数字素养。如,一是加强人工智能时代的数字文化建设,推动相关技术资源的联动配合,打好技术矩阵的“组合拳”,设计打造反映“主旋律”“正能量”的数字文化产品,提升主流意识形态的话语权和竞争力。二是通过鲜活的日常素材,用更易于理解的熟悉内容转化阐释相关理念的具体内涵,调动人们提升数字素养和参与相关建设的积极性,不断提高人们对伦理道德的实践能力和意识形态的甄别能力,以此激活民众维护意识形态安全的责任意识,以高度的数字素养自觉抵御不良信息的侵蚀。

总之,人工智能技术带来了生产生活方式的全域性变革,形成了意识形态的新现象、新规律,进而在不同应用逻辑和技术逻辑的交叉杂糅下出现了诸多意识形态新风险。针对意识形态领域的新风险,只有立足于党的领导、法治、道德规范、社会治理等各方面力量的整合,才能进行既有针对性又有整体性的科学防范,而这本身也是中国式现代化在人工智能时代创造数智文明的意识形态领域尝试。

参考文献:

[1] 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[N]. 人民日报, 2018-11-01(01).

[2] 中共中央文献研究室编. 习近平关于科技创新论述摘编[M]. 北京:中央文献出版社, 2016:76.

[3] 王天恩. 信息及其基本特性的当代开显[J]. 中国社会科学,2022(1):90-113.

[4] 陈钟. 从人工智能本质看未来的发展[J]. 探索与争鸣,2017(10):4-7.

[5] 王晓升. 西方马克思主义意识形态理论[M]. 北京:社会科学文献出版社,2009:6.

[6] 马克思,恩格斯. 马克思恩格斯文集:第一卷[M]. 中共中央马克思恩格斯列宁斯大林著作编译局,译. 北京:人民出版社,2009.

[7] 马克思,恩格斯. 马克思恩格斯文集:第五卷[M]. 中共中央马克思恩格斯列宁斯大林著作编译局,译. 北京:人民出版社,2009.

[8] 王文臣,马梦雪. 论大数据作为生产要素的资本化及其双重影响[J]. 上海财经大学学报,2022, 24(6):108-122.

[9] 孙伟平. 价值哲学视域中的算法歧视与社会公正[J]. 哲学研究,2023(3):46-55.

[10] 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,1999:63.

[11] 马克斯·韦伯. 经济与社会:上卷[M]. 林荣远,译. 北京:商务印书馆, 1997:81.

[12] 涂良川. 马克思历史唯物主义视阈中的人工智能奇点论[J]. 东北师大学报(哲学社会科学版), 2020(1):78-84.

[13] 洪晓楠,刘媛媛. 人工智能时代网络意识形态安全建设的发展契机、潜在风险与调适进路[J]. 思想教育研究,2022(10):138-144.

[14] 凯斯·桑斯坦. 网络共和国:网络社会中的民主问题[M]. 黄维明,译. 上海:上海人民出版社, 2003:2.

[15] 徐艳如. 数字拜物教的秘密及其背后的权力机制[J]. 马克思主义研究,2022(6):105-113.

[16] 刘伟兵. 算法的意识形态与意识形态的算法[J]. 深圳大学学报(人文社会科学版), 2024,41(1):152-160.

[17] 徐菲. 风险与治理:资本与舆论的博弈模型与现实解释[J]. 河海大学学报(哲学社会科学版), 2024,26(2):60-69.

[18] 张爱军. “算法利维坦”的风险及其规制[J]. 探索与争鸣,2021(1):95-102.

[19] 国家互联网信息办公室. 互联网信息服务算法推荐管理规定[R]. 2022.

[20] 中共中央办公厅 国务院办公厅. 关于加强科技伦理治理的意见[EB/OL]. (2022-03-20). https://www.gov.cn/gongbao/content/2022/content_5683838.htm,2022-03-20.

[21] 中央网络安全和信息化委员会办公室. 提升全民数字素养与技能行动纲要[EB/OL]. (2021-11-05). http://www.cac.gov.cn/2021-11/05/c_1637708867754305.htm,2021-11-05.

The Generation Logic and Scientific Prevention of Ideological Risks in the Age of Artificial Intelligence /LIU Weibing (School of Marxism,Fudan University,Shanghai 200433, China)

Abstract: Artificial intelligence relies heavily on machine learning techniques for its intelligence. Through the training of data, the design of algorithms and the operation of models, machine learning technology is able to transform problems into mathematical and statistical solutions, thus giving the logic of artificial intelligence technology a mathematical logic. Artificial intelligence technology generates the phenomenon of “consciousness” on the basis of socialized and collaborative mental production, and generates the ideology that corresponds to it, driven by the logic of capital application and technology abuse. The ideology based on the new characteristics, new laws and technological logic of AI technology is also facing the possibility of new ideological risks brought about by the provisions of AI technology. Ideological risks in the era of artificial intelligence in the field of ideology generation, dissemination and identity respectively present “capital logic, algorithmic black box and algorithmic ethics”, “algorithmic recommendation, filtering bubbles and information cocoon”, and “subcultural space, value penetration and digital fetishism”. The different stages of risk are inherited and intertwined in reality, constructing a complex ideological risk network. Based on a comprehensive grasp of the logic of the generation of ideological risks in the AI era and a systematic sorting out of risk phenomena, we can target the root causes of ideological risks in the AI era through the application of the logic of socialism, the technical prevention of ideological risks in the AI era through the advancement of systematic regulation and technological innovation, and the systematic prevention of ideological risks in the AI era through the creation of a virtuous ecology and other pluralistic governance. This is also the ideological attempt of Chinese modernization to create a digital civilization in the age of artificial intelligence.

Key words: artificial intelligence; algorithms; ideology; ideological risk; capital logic