

构建可持续发展的人工智能生态： 问题、启示与路径

邬晓燕 李欣颖

[摘要] 快速发展的人工智能通过重构社会生产要素与资源配置模式,正在重塑人类文明的发展图景。人工智能具有技术增效、社会治理赋能、人类认知边界重构等变革动能,但其衍生的生态代价、技术风险、主体性价值消解等负面效应亦须审慎应对。构建一种可持续发展的人工智能生态成为当务之急,可借鉴生态学理论对人工智能治理的启示,即应用系统观念规划人工智能生态、倡导共生发展的人工智能治理策略、确立可持续发展的人工智能生态价值目标,进而构建起贯穿绿色理念、实现信息共享、形成协同支撑、推动负责任创新的人工智能生态。

[关键词] 人工智能;可持续发展;人工智能生态

[中图分类号] D63

[文献标识码] A

[文章编号] 1674-7453(2025)07-0017-09

近年来,人工智能技术席卷全球,已成为推动各行业进步的关键力量,深刻影响人类社会的发展格局。然而,问题与风险也不可忽视。习近平总书记强调,“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,确保人工智能安全、可靠、可控”,“要整合多学科力量,加强人工智能相关法律、伦理、社会问题研究,建立健全保障人工智能健康发展的法律法规、制度体系、伦理道德”。^[1]生态学中的系统观念与共生规律,能够为可持续发展的人工智能治理提供理论思路,进而构建起贯穿绿色理念、实现信息共享、形成协

同支撑、推动负责任创新的人工智能生态。

一、人工智能的广泛应用及其问题

人工智能技术在近年来取得了显著进展,赋能人类生活的各个领域。自动驾驶技术推动交通管理体系智能化升级,数字化教学系统推动教育模式创新,医疗领域通过人工智能辅助提升诊断效率,金融行业借助智能算法优化风险评估。人工智能不仅大幅提高了生产效率,还催生了数据服务平台、自动化商业系统等新型业态。然而,人

[基金项目] 国家社会科学基金一般项目“绿色技术创新与生态文明转型的协同演化研究”(21BZX051)。

[作者简介] 邬晓燕,北京交通大学马克思主义学院教授、博士生导师;李欣颖,北京交通大学马克思主义学院硕士研究生。

工智能在被广泛应用的同时也带来了风险与挑战,如算法决策过程缺乏透明性,用户隐私数据存在安全隐患,部分行业面临就业结构调整的压力等。解决这些挑战需要协同推进技术创新与规范管理,通过制定法规明确技术应用边界,建立数据使用的伦理准则,促进技术研发与产业需求的精准对接。只有当技术发展和公共治理形成有效适配,人工智能才能持续转化为推动社会进步的核心动力。

(一)技术增效与生态代价

人工智能在生态治理中展现出技术增效的潜力,但随着人工智能的发展,算力和预训练对能源的消耗逐渐增加,暴露出技术进步与可持续发展目标之间的矛盾。在生态治理领域,以智能算法为核心的人工智能技术全面革新生态治理范式。气候预测系统通过解析海量复杂数据,能够将地球气候系统的动态变化转化为可量化的演变趋势预测;能源管理领域的智能算法则通过实时优化电网运行,显著提升了风能、太阳能等清洁能源的利用效率。人工智能技术正在突破传统生态治理的局限,利用算法模型模拟生态系统自我调节机制的强大潜力,为维护生态平衡提供了新的技术路径。然而,人工智能技术在提质增效的同时,也面临热力学定律的终极拷问。人工智能发展的环境代价,凸显了在技术进步与生态保护之间寻求平衡的紧迫性。

第一,数据分析、云计算会消耗大量能源。数据中心是人工智能蓬勃发展及运行的关键支撑,但它却以消耗海量的能源为代价,能耗问题日趋严峻。根据国际能源署估算,到2030年,为互联网和越来越多的人工智能供电的计算机仓库可能需要高达400太瓦时电力,这超过了英国2022年全年发电量。^[2]随着人工智能技术不断迭代、数据量呈指数级增长,能耗只会居高不下。根据斯坦福大学发布的《2023年人工智能指数报告》,GPT-3模型训练耗费的电力可供一个美国家庭使用数

百年,CO₂排放量也相当于一个家庭近百年的排放量。同时,基于大模型算力需求的高能耗运行,其热量释放也呈现倍增态势。^[3]这一方面折射出人工智能产业高速发展背后隐藏的资源危机,可能制约人工智能产业的长远健康发展;另一方面大量的电力消耗往往意味着更多的碳排放和生态环境破坏,与全球倡导的绿色可持续发展背道而驰。

第二,技术迭代产生的电子废弃物会加剧环境污染。随着人工智能的普及,其更新迭代所产生的电子废弃物呈爆炸式增长,给生态环境带来沉重负担。生成式人工智能的底层是一个材料密集型行业,数据中心和服务器群中包括多种高性能计算硬件,比如图形处理器(GPU)、中央处理器(CPU)、内存模块、存储设备等,且服务器寿命普遍在3年左右。当其寿命结束后,就会成为电子废弃物。如果不采取任何循环经济措施,至2030年,生成式人工智能产生的电子废弃物将高达250万吨。从2023年到2030年,在激进、中等和保守情景中,产生的电子废弃物总量分别是500万吨、300万吨和180万吨。^[4]由此可见,人工智能依赖的硬件系统本质上是“人造代谢体”,其生产、消耗与废弃的速率远超自然生态系统的再生与净化能力。目前亟须建立人工智能生态经济循环网络,避免资源流动的单向枯竭与毒性物质的空间累积。

(二)社会治理赋能与技术风险衍生

人工智能依托数据模型重构社会运行的信息处理机制,优化公共资源调配,可以实现治理效能的系统性跃升,但随着人类对技术依赖性的增强,新型技术风险逐渐显露。人工智能驱动的数字化治理模式,本质上实现了由经验判断向数据建模的根本性转型。基于机器学习的算法决策系统,通过实时解析多源数据流,建立起预测、感知、反馈的闭环治理体系,助力公共资源配置突破传统科层制的信息衰减困境,形成精准识别、动态调整

的弹性治理模式,推动社会治理从“被动应对”向“主动预防”转型,从职能驱动的科层制管理向流程驱动的矩阵式管理、数据驱动的网络型管理、智能驱动的价值生态共生管理等管理方式转变,最终实现由数据治理向数智治理转变。然而,技术赋能的同时也催生难以规避的系统性风险。

第一,“算法黑箱”与不可解释性会导致伦理风险。“算法黑箱”由于符号系统的不可通约性,导致算法内部具体的操作链条无法被翻译为人类可理解的逻辑命题。人们很难明晰意向性编码或解码过程,也难以理解、预测和验证算法的运行逻辑及结果,所以也无法明确人工智能算法的运算过程是否违反社会的伦理规范,从而导致算法伦理溯源深具挑战性。^[5]人工智能目前还未完全达到人的智能水平,却被赋予了为人决策的权限。人工智能的决策机制面临着“算法黑箱”和“不可解释”的双重困境。一方面,算法内部逻辑的生成过程呈现出高度的不可解释性。算法是从信息体向智能体发展的机制基础,正是由此逐渐获得智能算法的自主性。^[6]另一方面,算法开发者虽然能够编写程序设定算法架构,但对其自主学习产生的结果却难以解释,这不仅造成“创造者不可控”的问题,更因“算法黑箱”给社会治理带来挑战。

第二,算法的不透明性和不可解释性,会进一步导致算法偏见。算法训练可能将数据中携带的历史偏见转化为系统性歧视,威胁算法公正。如性别、种族等敏感属性的隐性关联被编码为决策规则,因推理路径不可视、歧视行为具有隐蔽性而难以识别。算法偏见的来源大致可分为两类:一是算法模型自身存在与偏见相关的固有缺陷;二是开发者个人认知与价值观念的局限性通过开发设计过程融入模型。^[7]这些偏见被算法模型放大,可能导致大模型输出歧视性内容,^[8]其产生的后果远超技术范畴,深刻影响着社会公平。

(三)人类认知边界重构与被动性遮蔽

人工智能在拓展人类认知边界、推动决策能

力跃升的同时,其技术主导性扩张正逐步消解着人类主体性的价值根基。其一,人工智能系统凭借计算与模式识别能力,可实时分析海量动态参数,突破了人脑只能处理有限的决策变量困境。其二,人工智能通过传感器网络与数据分析,正在突破人类感官的物理局限,将不可见、不可感的微观与宏观世界转化为可理解的认知图景。其三,人工智能通过模拟推演,帮助人类建立动态的系统认知框架,促使人类思维从“线性逻辑”向“系统思维”不断进化。然而,人工智能技术的快速迭代也正将人类推入前所未有的认知困境。一方面,社交机器人可能在网络空间中争夺话语权和议程设置权,操纵公众的价值观和思维认知。^[9]另一方面,由个性化推荐导致的信息茧房,可将人类思维禁锢于算法编织的认知牢笼之中。

第一,人工智能生成的虚假信息泛滥,正引发严重的认知污染问题。在2025年两会期间,人工智能技术滥用与虚假信息治理问题成为热议焦点。^[10]随着人工智能技术的发展,生成文字、图片、视频的“门槛”逐渐降低,越来越多的人已经能够熟练掌握生成式人工智能技术。生成式人工智能技术看似是创造性生产,实际上其运行逻辑是对现有数据的重新排列组合,本身存在生成虚假信息风险。^[11]同时,人工智能技术会加速信息传播,使得虚假信息在短时间内指数级扩散,由于人们在短时间内无法确定信息的真伪,这些虚假信息可能会侵蚀公众的价值观和思想,造成“认知污染”,同时可能威胁政治安全和网络安全。第二,信息茧房的禁锢效应严重阻碍人类认知的拓展与深化。信息茧房现象标志着人类正经历认知领域的“静默革命”,其实质是算法权力通过神经认知机制和符号权力运作,完成对思维结构的渐进式改造。这一过程包含五个递进层次的技术控制。第一层,注意力被算法劫持,个性化推荐系统重构注意力分配模式。第二层,算法通过控制词语联想网络,将自由的思想转化为预设路径的选

择,技术对认知多样性产生系统性扼杀。第三层,算法通过构建“超真实”,使用户逐渐失去区分事实与虚构的能力。第四层,社交媒体平台通过情绪化内容推荐,制造“认知部落化”现象,^[12]即算法通过多巴胺刺激强化人产生道德愤怒,使群体认同从理性共识转向情绪化站队。第五层,在认知驯化的终极阶段,人类思维彻底沦为算法系统的外延器官。当系统能直接读取神经信号并反馈定制化信息时,笛卡尔“我思故我在”的哲学根基将被颠覆。这种异化不仅是工具理性的胜利,更标志着人类在技术系统中正经历存在论层面的自我消解,人类将不再是使用工具的主体,而是技术系统自我复制的载体。

二、人工智能治理的生态学启示

生态是一个包含生产者、消费者和分解者的有机整体,具有整体性、动态平衡性、多样性和开放性。自然生态系统与人工智能生态系统有着深刻的内在相似性。与传统独立的人工智能系统不同,人工智能生态强调各要素间的紧密联系与协同共进。

(一)生态与人工智能生态的概念界定

生态,从广义而言,是指生物与环境之间相互作用、相互依存所形成的有机整体。其一,生态系统具有整体性,其中任何一个组成部分的变化都可能引发连锁反应,影响整个系统的功能和稳定。当某一区域的森林植被遭到破坏时,不仅会导致依赖森林生存的动物失去栖息地,还可能引发水土流失、气候调节功能减弱等一系列问题,进而影响整个生态系统的平衡。其二,生态系统处于不断的动态变化之中,在自然条件下,生态系统通过自我调节机制维持相对稳定的状态。当生态系统受到外界破坏时,其自身会在一定范围内进行自我修复和调整,恢复平衡。而当破坏程度超过了生态系统的承受能力,平衡就会被打破,导致生态

系统的退化。其三,生态系统具有多样性,具体表现为生物种类的多样性、生态环境的多样性等多个方面。其四,生态系统具有开放性,生态系统与外界环境之间存在着物质、能量和信息的交换。通过这种交换,生态系统不断从外界获取必要的资源,维持自身的生存和发展。同时,生态系统也会向外界输出物质和能量,对周围环境产生影响。

人工智能生态是指由算法、数据、算力为核心要素,以研发人员、科研机构、企业、用户等为主体,在特定的政策、法规、伦理规范与社会文化环境下,相互作用、协同演化所形成的复杂系统。其中,数据作为训练与验证原料,在驱动算法参数优化迭代的同时,反哺算力系统效能的提升需求;算法作为建模工具,通过对数据的特征提取与模式识别,实现智能功能;算力则是物理硬件基础,为算法的运行提供计算资源支撑。三者通过动态反馈形成技术进化循环,数据量的指数增长则持续推动算力架构创新,算法革新提升数据处理的维度与精度,算力升级使更大规模算法模型成为可能。人工智能生态呈现出显著的开放性、动态性与适应性的特征。其一,开放性使其能够广泛吸纳外界的新技术、新理念、新数据,如在智能医疗领域,人工智能的应用便于随时融入最新医学研究成果、临床数据,优化诊断算法。其二,动态性意味着人工智能系统内部的要素持续更新、结构不断优化调整,以契合外界瞬息万变的需求,如市场信息随市场需求的更迭实时优化。其三,适应性则让人工智能在面对复杂多变的环境时,能够自主调整策略,保持稳定运行,如自动驾驶系统能够在不同路况、天气条件下灵活决策,确保行车安全。这些特性相互交织,促使人工智能生态在科技浪潮中持续进化、深度融入,并重塑社会各个领域。

(二)人工智能治理的生态学意涵

以 DeepSeek 为代表的生成式人工智能的崛起,标志着中国在全球人工智能竞争格局中实现

了历史性突破。党的十八大以来,以习近平同志为核心的党中央深入推动实施创新驱动发展战略,我国人工智能技术创新取得重要进展,特别是近年来人工智能大模型研发应用步伐加快,目前已进入技术攻坚、应用深化、生态重构的新阶段。^[13]但人工智能的技术迭代和风险涌现速度远超传统技术,^[14]深度伪造、算法偏见、伦理失控等新型风险逐渐显现,现行的“碎片化、被动响应式”治理模式面临严峻挑战。生态学视角为构建人工智能生态提供了崭新的认知框架。生态学中的系统观念与共生规律,为解决当前人工智能治理中“头痛医头”的碎片化困境指明了方向,即将人工智能的发展置于技术、社会、自然的多维生态网中进行审视,治理策略从“控制论”转向“共生论”。

第一,应用系统观念规划人工智能生态。其一,将系统观念应用于人工智能治理。生态学强调整体性,有学者将生态学定义为“有机体与整体环境间关系的科学”,^[15]将生态系统视为有机整体,各部分相互关联、协同进化。规划人工智能生态要求确立系统观念,统筹算法、数据、算力等技术要素,构建协同创新体系;整合科研机构、企业、高校力量,避免各自为政,形成产学研用联动链条。其二,积极协调人工智能内部各要素,实现协同发展。自然生态系统常因内部要素失调导致整体失衡,如水污染致水生生物受损,从而影响整个食物链。人工智能生态中,各要素的发展能力制约着整体效能,比如,算力不足使算法运行迟缓,劣质数据导致模型偏差。所以,平衡发展各要素至关重要,必须加大算力基础设施投入,提升数据治理水平,鼓励算法创新突破,优化各要素配置,确保各环节紧密衔接、协同发力,最终实现整体效益最大化。

第二,倡导共生发展的人工智能治理策略。其一,促进不同人工智能领域的共生发展。德贝里在1879年首次提出“共生”的概念,将其定义为:“不同种属在不同获益关系中生活在一起。”^[16]

生态系统多样,物种共生共荣,比如在热带雨林中,菌类分解枯枝,藤蔓植物攀附支撑,共同构建了复杂的生态。共生理论后来拓展至社会科学领域。如今,人工智能领域也应倡导共生发展。如医疗影像诊断联合计算机视觉与医学知识,辅助精准识别病灶,人工智能共生性的治理有利于各领域优势互补,拓展智能边界,激发创新活力,丰富生态系统功能。其二,建立人工智能与其他产业的共生关系。在自然生态中,生物与非生物环境紧密依存,植物根系固土、微生物参与土壤形成。因此,人工智能与传统产业也应建立共生纽带,可赋能制造业实现智能制造,优化生产流程、预测设备故障;可助力农业精准种植,依土壤、气候、作物数据智能灌溉施肥;可融入服务业提升体验,智能物流优化配送、实现教育个性化教学;等等。

第三,确立可持续发展的人工智能生态价值目标。其一,确保人工智能生态的发展稳定可控。自然生态系统具有自我调节的能力,可以维持自身稳定,人工智能的发展亦需要稳健可控的生态机制。因此,搭建能够及时响应的防护网络至关重要,通过实时监测对可能出现的算力瓶颈、隐私泄露或伦理偏移等问题提前预警。其二,推动人工智能生态的绿色发展。将低碳理念贯穿技术全生命周期,在硬件层面建立算力分层架构,将高精度计算与轻量化任务分级处理,降低整体能耗。在算法设计中引入简约原则,数据治理则可借鉴生态系统的自净功能,自动清理冗余信息,提升存储效能。这种立体化的绿色实践,不仅能够降低行业碳排放,更能推动技术创新与生态保护的深度融合、和谐共生,顺应可持续发展的时代诉求。

三、构建可持续发展人工智能生态的多维路径

人工智能治理需要正确把握人工智能与生态

环境、社会发展、人类自身的关系,确保人工智能技术的发展朝着以人为本、技术向善的方向迈进。党的二十届三中全会明确提出,“完善生成式人工智能发展和管理机制”,“建立人工智能安全监管制度”等一系列关键性改革举措。^[17]因此,人工智能治理亟待搭建一套科学严谨、系统完备且行之高效的人工智能生态体系,助推人工智能可持续发展。

(一)打造“绿色贯穿”的人工智能

如果不处理好技术与生态保护之间的深层次矛盾,人工智能发展将会面临资源短缺和环境破坏的双重困境。因此,打造绿色贯穿的人工智能,不仅能够确保人工智能技术发展的可持续性,而且有益于为其他行业提供可持续发展范式。

其一,人工智能的可持续发展需要构建数字时代的生态治理体系。人工智能的可持续发展必须突破传统技术开发范式,在能源消耗、硬件生产、算法设计等全生命周期构建绿色生态。在数据中心节能优化层面,技术创新要与管理策略协同发力。“数据作为新型生产要素,对传统生产方式变革具有重大影响。”^[18]一方面,可研发高效能制冷技术,如用液冷技术取代传统风冷,冷却液可精准带走服务器热量,以降低能耗;智能电源管理系统通过动态调控电压电流,可减少电力浪费。另一方面,优化数据中心布局,可依气候条件选址,利用自然冷源,延长自然冷却时长。此外,可在数据传输高效低耗领域,应用前沿技术。

其二,在算法层面,建立“预防性创新”机制,通过前瞻性政策倒逼绿色技术突破。一方面,将碳强度作为算力资源分配的核心指标,促使企业优先开发生态友好型算法。这种主动干预并非限制创新,而是引导创新方向与人类整体利益保持一致。另一方面,探索新型人工智能算法,为资源节约型算法的发展开辟新方向,提升资源利用效率。

其三,推动人工智能赋能现代农业、工业现代化,助力经济社会全面绿色转型。譬如,推进

人工智能深度赋能精准农业,大力发展绿色农业。一方面,可借助先进的传感器技术和卫星遥感、无人机航拍等手段,构建全方位、多层次的农业数据采集体系。另一方面,可将智能监测与数据分析的结果,作为农业生产决策的有力支撑,以更好地驱动农业设备进行精准作业。更重要的是,人工智能可通过重构农业生产要素配置,在资源利用效率与生态承载力之间建立动态平衡,推动传统农业向“物质循环—能量流动—信息反馈”的生态闭环转型。这种系统性变革不仅可突破粗放式发展的路径依赖,更可在生产力提升与生态约束的辩证统一中,开辟出绿色农业的可持续发展新范式。

(二)打造“信息共享”的人工智能

在全球化背景下,各国在人工智能领域的相互依存度日益提升,在第三届“一带一路”国际合作高峰论坛期间,我国发布了《全球人工智能治理倡议》,^[19]倡导各国在人工智能治理进程中大力加强信息的交流共享与技术的协同合作,以利于整合全球资源,减少因误解或误判而引发的潜在风险。信息共享的人工智能旨在推动打造在全球范围内具有广泛共识性的人工智能治理框架与标准规范。

其一,传统产权制度将数据视为可独占的“物”,但在人工智能生态中,数据具有非竞争性、可复制性和网络外部性,单一主体的绝对控制权反而会抑制其社会价值。因此,人工智能生态要求突破“非此即彼”的确权逻辑,建立“权利束”分离机制:将所有权分解为“隐私控制权”“开发使用权”和“收益分配权”,分别赋予不同主体。欧盟《数据治理法案》提出的“数据中介服务”模式,正是通过第三方平台实现权利分离的尝试。这种制度设计的关键在于构建“数据中介”(Data Intermediary),由政府、行业协会、技术专家组成监督委员会,确保数据利用不偏离公共利益轨道,同时通过智能合约自动执行权益分配,避免传统确权制度下的高额交易成本。

其二,人工智能的信息共享无法依靠单一主体完成,需构建政府、市场、社会、公众多元共治的弹性框架。在算法设计中引入公众参与和公众反馈机制,是应对技术伦理挑战、构建社会信任的重要路径。公众参与的核心在于打破算法开发的“技术黑箱”,让普通用户、利益相关方和伦理专家共同介入。公众反馈机制能确保及时反馈算法在真实世界中的运行偏差。比如,当社交平台用户点击“不推荐此内容”时,当求职者申诉简历筛选系统的误判时,这些看似微小的反馈实则是优化算法的关键依据。这种反馈机制不仅让技术具备修正的能力,更重要的是赋予公众对抗算法霸权的权利,重建权力平衡。

其三,在国际组织与政府层面,应加强人工智能国际合作框架搭建。在国际治理的理论架构中,统一且科学的治理框架和标准规范是确保技术有序发展、保障公共利益的核心要素。在“一带一路”倡议下,沿线国家数字经济合作升温,比如,中国与东盟共建人工智能联合实验室,为区域经贸发展注入新动能。通过这一举措,旨在持续提升人工智能技术的安全性、可靠性、可控性和公平性,为全球人工智能的可持续发展营造良好环境,开启人类科技与社会协同进步的新篇章。

(三)打造“协同支撑”的人工智能

人工智能的复杂性决定了仅靠单一主体无法独立完成技术突破与应用转化。当前创新体系的碎片化,本质上是知识生产模式与产业需求结构的错配。2017年7月,国务院印发的《新一代人工智能发展规划》指出,要“构建开放协同的人工智能科技创新体系”,“新一代人工智能关键共性技术的研发部署要以算法为核心,以数据和硬件为基础,以提升感知识别、知识计算、认知推理、运动执行、人机交互能力为重点,形成开放兼容、稳定成熟的技术体系”,^[20]旨在建立人工智能的协同机制,从而为人工智能的发展提供有力支撑。

其一,构建“政产学研用”五位一体的人工智

能主体协同。人工智能的发展需要突破传统线性模式,因此要构建政府、企业、高校、科研机构及用户深度融合的协同网络,形成“基础研究—技术开发—产业应用—反哺研究”的闭环链条,实现知识生产、技术转化与价值创造的多维互动,最终构建起主体能力互补、资源共享的共生型人工智能生态。

其二,推进“架构—标准—能力”三位一体的人工智能技术协同。解决人工智能技术碎片化难题,需从系统层面推进技术要素的有机融合。通过为复杂场景应用提供弹性化、可持续的技术支撑体系,实现多元化技术要素的整合。

其三,建立“利益—风险—要素”联动的人工智能机制协同。人工智能协同生态的可持续运转依赖于制度层面的深层次变革。需构建基于区块链的贡献计量系统,实现数据、算法等生产要素的量化确权与价值分配;设立创新风险缓冲基金,通过社会化共担机制化解技术不确定性;完善算力、人才等要素的市场化流通体系,打破资源垄断壁垒。通过制度创新重构生产关系,将零和博弈转化为共生关系,使个体理性与集体利益动态统一,为人工智能生态发展注入持久动力。

(四)打造“负责任创新”的人工智能

人工智能技术飞速发展,同时也面临着“技术不可控”的困境。英国经济学家理查德·欧文(Richard Owen)最早提出“负责任创新”一词,指出“负责任创新注重技术创新活动与社会责任的有机结合,构建了由预测性、包容性、反身性以及反馈性组成四维度体系”。^[21]面对伦理困境、安全隐患、社会结构重塑等巨大挑战,打造负责任创新的人工智能旨在为人工智能发展锚定正确方向,确保其发展轨迹契合人类的根本利益与长远福祉,避免因技术的无序狂奔而陷入不可控的泥沼。

其一,“负责任创新”的人工智能要求突破“技术中立”的传统范式,将伦理考量内化为技术发展的内生维度。通过价值对齐技术(Value

Alignment),将人类伦理规范编码为算法的决策权重。同时,算法价值对齐不应是对社会复杂性问题进行价值偏差的简单调校,而应置于防止人类社会整体撕裂和极端思潮、激进主义的政治叙事和人类精神维度内,推动算法价值对齐和人类命运共同体同向同行。^[22]应从数据全生命周期保障和数据安全防护体系构建两方面着手强化数据管理责任。在数据全生命周期保障上,数据采集阶段要多源校验确保准确性;在数据安全防护体系构建方面,利用加密算法,保障数据传输、存储安全,防止数据被窃取或篡改。

其二,“负责任创新”的人工智能应做到算法透明可释。通过建立分层解释与社会监督机制,破解算法黑箱的信任危机,构建分层次的透明性体系。提升算法可靠程度、增强算法可解释性,可从算法设计优化与算法评估监测两方面发力。在算法设计优化上,运用决策树、贝叶斯网络等模型可视化展示决策逻辑;在算法评估监测方面,构建长期动态的评估体系,引入多方监督,包括专家、用户和监管机构,全方位保障算法可靠且可解释。

其三,“负责任创新”的人工智能应健全监督评估体系与完善监督机制,以解决技术失控风险与社会信任危机。监督评估体系应覆盖技术研

发、部署应用及社会影响三大维度:在研发阶段,制定伦理合规性指标;在应用阶段,建立“社会影响追踪”机制;在迭代阶段,引入第三方审计机构,对算法更新进行“伦理差分分析”,识别版本迭代中的价值偏移,同时要求企业以可视化报告形式向用户披露具体采集的信息类型、数据使用路径及算法应用场景,确保个人信息处理流程透明可溯。“第三方审计的相关技术治理措施有助于提高模型安全性和透明度,公众可以提高对利益相关者的信任,确保模型的负责任使用,并符合法律、行业标准和伦理要求。”^[23]

总之,“负责任创新”的人工智能需要超越单向度的技术效率追求,在算法设计中植入价值对齐机制,实现技术逻辑与社会价值的深度校准。将人类社会的公平、隐私、安全等核心伦理诉求转化为可操作的代码规则,确保人工智能系统的行为符合人类的价值观和伦理原则。这意味着算法不仅要打破“黑箱”效应,建立可解释、可追溯的信任纽带,还需构建多方参与的动态治理框架,使技术标准随社会发展持续优化。同时,只有将“负责任”意识贯穿人工智能的创新链条,才能真正使人工智能得到可持续发展,为构建智慧可信的数字文明培育可持续驱动力。

【参考文献】

- [1]习近平主持中共中央政治局第九次集体学习并讲话[EB/OL]. 中国政府网,https://www.gov.cn/xinwen/2018-10/31/content_5336251.htm.
- [2]王林.全球老旧电站加速“变身”数据中心[N]. 中国能源报,2024-10-14.
- [3]王祺,李冬露,张云,等.2023年中国AIGC产业全景报告:日就月将,学有缉熙于光明[C].2023艾瑞咨询8月研究报告会论文集.
- [4]赵广立,赵宇彤.预警!生成式AI将制造巨量电子垃圾[N]. 中国科学报,2024-11-04.
- [5]黄静秋,邓伯军.人工智能算法的伦理规制研究[J]. 北京科技大学学报(社会科学版),2025(2).
- [6]王天恩.人工智能算法的进化及其伦理效应[J]. 山西师大学报(社会科学版),2024(2).
- [7]毕昆,张华麟.算法、算据与算力:人工智能风险治理范式的技术支撑[J/OL]. 江苏社会科学,1-8[2025-06-06].
- [8]吴永兴,菅志宇,苏昌,等.生成式人工智能大模型的安全风险及治理路径[J]. 中国现代教育装备,2025(9).
- [9]韩蒙,潘智丹,李荣昌.《人工智能安全治理框架》加快人工智能风险治理[J]. 中国信息安全,2024(10).
- [10]文丽娟.代表委员建言加强AI虚假信息治理[N]. 法治日报,2025-03-06.

- [11]邓宏光,王雪璠.生成式人工智能虚假信息治理的风险与应对[J].理论月刊,2024(9).
- [12]Crockett, M.J. Moral Outrage in the Digital Age[J].Nature Human Behaviour,2017(11).
- [13]唐杰.大力推动我国人工智能大模型发展[N].人民日报,2025-05-08.
- [14]李诗悦.生成式人工智能嵌入公共服务风险的形成机理与防范对策[J/OL].电子政务,1-12[2025-06-06].
- [15]张金屯,李素清.应用生态学[M].科学出版社,2003:1.
- [16]DOUGLAS A E. The Symbiotic Habit [M]. New Jersey: Princeton University Press, 2010:5.
- [17]中共中央关于进一步全面深化改革 推进中国式现代化的决定[EB/OL]. 中国政府网,https://www.gov.cn/zhengce/202407/content_6963770.htm.
- [18]习近平谈治国理政(第四卷)[M]. 外文出版社,2022:206.
- [19]国家互联网信息办公室.全球人工智能治理倡议[EB/OL]. 中国网信网,https://www.cac.gov.cn/2023-10/18/c_1699291032884978.htm.
- [20]国务院关于印发新一代人工智能发展规划的通知[EB/OL]. 中国政府网,https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.
- [21]Stillgoe J, Owen R, Macnaghten P. Developing a Framework for Responsible Innovation[J].Research Policy, 2013(9).
- [22]曹克亮.算法时代的政党政治:重要影响、现实后果与应对逻辑[J].统一战线学研究,2025(2).
- [23]周辉,郭烘佑.大语言模型安全的技术治理:对抗测试与评估审计[J].西安交通大学学报(社会科学版),2025(2).

责任编辑:刘翠霞

Constructing a Sustainable Artificial Intelligence Ecosystem: Challenges, Insights, and Pathways

Wu Xiaoyan & LiXinying

[Abstract] The rapidly developing artificial intelligence is reshaping the development landscape of human civilization by reconstructing social production factors and resource allocation patterns. Artificial intelligence has the driving force of technological efficiency, social governance empowerment, and reconstruction of human cognitive boundaries, but its negative effects such as ecological costs, technological risks, and the dissolution of subjective values also need to be carefully addressed. Building a sustainable artificial intelligence ecosystem has become an urgent task, which can draw inspiration from ecological theory for artificial intelligence governance. This includes applying a system concept to plan the artificial intelligence ecosystem, advocating for symbiotic development of artificial intelligence governance strategies, establishing sustainable development goals for the artificial intelligence ecosystem, and then constructing a sustainable artificial intelligence ecosystem that is green, information sharing, collaborative support, and responsible innovation.

[Key Words] Artificial Intelligence (AI); Sustainable Development; AI Ecosystem