

以丰富国际公共产品供给引领人工智能全球治理

□ 薛澜 陈颖

习近平主席在2026世界人工智能大会暨人工智能全球治理高级别会议开幕式上发表主旨讲话,强调:“作为负责任大国,中国在人工智能领域始终致力于做国际公共产品的提供者。”习近平主席的重要讲话,着眼人类社会共同未来和人工智能发展大势,深刻阐明中国参与人工智能全球治理的价值立场和责任担当。人工智能作为引领新一轮科技革命与产业变革的战略性新兴产业,既蕴含巨大发展机遇,也带来治理挑战。我们要深入学习贯彻习近平主席重要讲话精神,以人工智能大国的胸怀与担当,不断丰富人工智能国际公共产品供给,为推动构建公正合理的全球人工智能治理体系贡献中国智慧、中国方案、中国力量。

统筹发展和安全,以治理型公共产品贡献中国方案

习近平主席强调:“人工智能应当成为人类的可信工具。要高度重视人工智能引发的各类内生和衍生风险,推动构建法律法规、技术监测、风险预警、应急响应体系,筑牢安全底线,防范滥用恶用,确保人工智能始终处于人类控制之下。”

人工智能在为经济社会发展注入强劲动能的同时,伴随着恶意使用、技术失灵、系统性风险等多重挑战。统筹发展和安全、平衡创新与治理,在推动人工智能广泛应用的同时,有效防范化解各类风险,成为国际社会共同面对的重大课题,也是人工智能全球治理的重要方向。特别是随着人工智能向通用化、自主化方向加速演进,其感知、推理、决策和行动能力不断增强,风险可能由单一环节向复杂系统传导,一旦系统目标偏离人类意图且无法被人类及时干预和有效控制,就可能引发难以预测的严重后果。当前,在特定实验环境中已发现人工智能系统存在策略性欺骗、规避干预和自我保护等倾向,存在隐瞒信息、误导操作者等策略性行为。尽管相关现象目前主要出现在模拟测试中,但其揭示的潜在风险不容忽视。

治理与创新相互促进、相辅相成。科学有效的治理能够明确规则边界、完善责任机制、强化风险防范,为技术创新营造稳定、透明、可预期的制度环境。人工智能技术迭代快、

应用场景广、风险形态不断变化,客观上要求及时更新治理思路、重塑治理关系,推动治理体系与技术发展同向而行、协同演进。要加强前瞻研判和政策储备,针对新技术、新业态、新风险及时出台相应规则,并根据技术演进、产业发展和实践反馈进行动态评估和持续完善,推动治理由相对静态的规则管理转向适应技术变化的动态调适。摒弃监管者与被监管者之间简单对立的博弈思维,健全政府、企业、科研机构、行业组织和社会公众之间的沟通协作机制,推动风险信息共享、治理责任共担,形成多元参与、协同共治的治理格局。根据应用场景和风险等级实行分类分级治理,在工具选择上突出柔性原则,综合运用监管沙盒、试点示范、动态评估等治理工具,实现规范与激励的有机统一。

早在2014年6月,习近平总书记在中国科学院第十七次院士大会、中国工程院第十二次院士大会上发表重要讲话,对人工智能等前沿技术发展作出前瞻性谋划。10余年来,中国人工智能发展始终沿着习近平总书记擘画的蓝图稳步前行,实现了从“跟跑”“并跑”到多个领域“领跑”的历史性跨越。同时,中国积极探索符合人工智能发展规律的敏捷治理路径。中国坚持包容审慎监管,在守住安全底线的前提下,为新技术、新业态成长预留必要空间。同时,面对隐私泄露、数据过度收集等新情况新问题,监管部门及时回应、动态调整,不断完善治理规则和监管措施。在此过程中,中国逐步形成了技术与治理完善协同推进的治理模式:技术创新不断向前,治理能力同步提升;风险问题及时识别,制度安排动态优化。这种治理模式既避免因监管过度抑制创新活力,也防止因治理缺位积累安全风险,实现了鼓励创新与规范发展的有机统一,为全球人工智能安全治理贡献了中国方案。

坚持开源开放,以技术型公共产品弥合数智鸿沟

习近平主席强调:“要抓住难得的历史性机遇,鼓励开源开放、合作共享,全面促进人工智能科技创新、产业发展、场景应用,协同推进传统产业改造升级、新兴产业培育壮大、

未来产业前瞻布局,让人工智能赋能千行百业。”

开源开放日益成为推动技术创新、成果扩散和产业应用的重要途径。前沿大模型研发具有技术密集、资金密集、人才密集和资源密集等特点,现阶段真正掌握大模型核心技术体系、具备从底层算法创新到模型规模化训练部署的国家屈指可数,全球约40%的人工智能核心技术研发集中在中美两国。当前,全球数智鸿沟不断扩大,少数发达国家收割技术红利、广大发展中国家沦为原始数据供应地的极端分化风险上升。国际货币基金组织制定的“人工智能准备指数”将全球数智鸿沟清晰量化,截至2024年4月,发达国家该指数为0.68,而新兴市场国家和低收入国家分别为0.46和0.32。作为人工智能领域“跟跑者”的许多发展中国家缺少有竞争力的科技企业,也缺乏足够的技术和人才储备,甚至在相关基础设施上都严重不足。国际电信联盟数据显示,至2024年底,全球仍有近26亿人未接入互联网,占全球人口的1/3,最不发达国家使用互联网的人数仅占总人口数的35%,内陆发展中国家仅有39%的人触网。

对于尚不具备前沿大模型自主研发能力的国家特别是广大发展中国家而言,开源大模型能够显著降低技术获取和应用的门槛,使这些国家得以在现有先进技术基础上开展本地化适配、场景化开发和产业化应用。越来越多的国家将开源大模型作为发展本国人工智能产业的重要技术底座,并以此加快应用能力提升和产业生态构建。促进人工智能研究成果与技术经验的开放共享,有利于扩大人工智能前沿技术的可及性,增强各国自主发展能力,为弥合全球数智鸿沟、推动人工智能普惠发展提供有力支撑。

中国始终坚持以开源开放促进人工智能普惠发展,以技术供给为发展中国家降低先进技术获取门槛、开展场景化应用和培育本土产业生态提供了重要支撑,有力推动了人工智能在全球范围的普及应用。当前,中国人工智能开源大模型全球累计下载量已突破100亿次。新一代开源模型Kimi K3正式发布,参数规模达

2.8万亿,为目前全球参数最大的开源模型,标志着我国人工智能模型发展取得新的成就。

践行多边主义,以制度型公共产品完善全球治理

习近平主席强调:“要践行真正的多边主义,切实发挥联合国的重要作用,加强人工智能发展战略、治理规则、技术标准的对接协调,早日形成具有广泛共识的全球治理框架,让这一前沿技术更好造福人类社会。”

人工智能带来的风险具有跨国性、复杂性和系统性,技术漏洞可能引发跨区域、跨领域的连锁反应。尤其是当人工智能应用基于同类基础模型和架构构建时,一旦出现共同漏洞则可能造成集体失效。这是国际社会面临的共同课题,没有任何国家能够独善其身、独自应对。确保人工智能安全、可靠、可控,离不开国际社会的协调合作。各国应把人工智能安全、可靠、可控作为重要的国际公共产品,在技术研发、安全评估、风险监测和应急处置等方面深化协调合作,推动形成信息共享、责任共担、风险共治的治理机制,共同筑牢人工智能安全、可靠、可控的全球防线。

当前,人工智能技术加速突破同地缘政治风险上升相互交织,全球治理面临大国博弈加剧、科技竞争升温、规则协调不畅等复杂局面。个别国家将人工智能治理异化为维护自身技术优势和竞争地位的工具,破坏长期形成的国际科研合作网络和产业协作体系,不仅阻碍人工智能技术进步和成果普惠,也无助于解决国际社会共同面临的风险挑战,反而损害各国共同利益。与此同时,全球人工智能治理体系尚处于加快形成阶段,不同国家和地区在发展阶段、治理理念、价值取向、制度安排和政策节奏等方面存在差异,各类治理机制相互交叠,规则标准衔接不足,国际协调难度不断加大。破解人工智能全球治理困境,必须坚持开放共赢、包容并蓄、和衷共济,在联合国框架下,推动形成人工智能全球治理共识,构建践行多边主义的国际合作平台。

(下转第10版)