



以有效治理保障人工智能健康有序发展

余晓晖

当前,人工智能作为引领新一轮科技革命和产业变革的战略性力量,正在深刻重构全球创新版图、发展范式与人类生产生活方式,同时其广泛应用也带来一系列风险挑战。习近平总书记高度重视人工智能治理,强调“要把握人工智能发展趋势和规律,加紧制定完善相关法律法规、政策制度、应用规范、伦理准则,构建技术监测、风险预警、应急响应体系”“形成具有广泛共识的人工智能治理框架和标准规范,不断提升人工智能技术的安全性、可靠性、可控性、公平性”等,系统回答了新时代为什么要加强人工智能治理、怎样加强人工智能治理等一系列具有全局性、前瞻性、战略性的重大理论和实践问题,为确保我国人工智能始终沿着有益、安全、公平方向发展提供了根本遵循。

新形势下加强人工智能治理具有重要意义

面对人工智能技术加速迭代、应用深度渗透、风险复杂交织的新形势,准确把握人工智能发展趋势和规律,加强人工智能治理,才能切实推动我国人工智能健康有序发展。

事关抓住新一轮科技革命和产业变革战略主动权。习近平总书记指出:“人工智能是新一轮科技革命和产业变革的重要驱动力量”。当前,人工智能技术创新进入密集活跃期,产业化进程加快,为赢得未来发展先机创造了有利条件。世界主要大国均在布局加快人工智能发展,催生智能体、无人驾驶、具身智能、智能穿戴等新领域新赛道,成为改变商业格局、重塑世界经济的新场域。加强人工智能治理,构建稳定、透明、可预期的治理环境,能够为企业、科研单位等提供明确规则指引,使其敢于投入、安心创新,进而最大程度激发创新潜能,提升国家创新体系整体效能,进一步夯实我国人工智能技术与产业发展的根基,并在这场科技竞赛中牢牢掌握战略主动。

事关推动我国经济社会高质量发展。习近平总书记强调:“我们要深入把握新一代人工智能发展的特点,加强人工智能和产业发展融合,为高质量发展提供新动能。”人工智能作为发展新质生产力的重要推动力,是推动产业智能化升级、培育经济新增长点的关键动能。

截至2025年底,我国人工智能企业超过6000家,形成了从基础底座、模型框架到行业应用的完整产业体系。一批国产大模型通过开源战略引领全球生态,推动人工智能从少数企业的尖端科技转变为千行百业可便捷调用的普惠性工具,为我国发挥超大规模市场与完整产业体系优势、实现换道超车奠定了独特生态基础。根据有关测试,我国的大语言模型基础能力与美国都居全球前列,模型复杂推理能力与美国的差距显著缩小,视频生成模型能力持续领先。在此背景下,探索科学有效的人工智能治理路径,能够更好防范风险,营造稳定的政策环境与发展预期,构建公平公正、健康有序、富有活力的产业生态,让人工智能真正成为驱动经济提质增效、社会全面进步的强大引擎。

事关维护国家和社会稳定。习近平总书记指出:“完善人工智能监管体制机制,牢牢掌握人工智能发展和治理主动权”。当前,人工智能安全风险日益凸显,呈现复杂性、系统性和泛在化特征。一方面,技术发展伴生安全风险,人工智能大模型的内在特性增加了安全可控的难度与复杂性。与传统网络安全相比,人工智能安全领域的攻防非对称性加剧,技术安全面临“易攻难守”新形势。另一方面,应用延展衍生安全新难题。人工智能大模型的广泛应用既带来应用形态迭代、开源生态滥用、软件供应链漏洞等外部性安全难题,也引发个人、群体、社会多层次安全风险。例如,人工智能拟人化互动服务诱导情感过度依赖现象高发,智能体任务失控及安全问题初显。因此,加强人工智能治理是贯彻落实总体国家安全观的内在要求,必须建立健全覆盖从技术研发、产品应用到社会影响全链条的风险防控体系,确保人工智能始终被用于维护国家利益和人民福祉。

事关构建人类命运共同体。习近平总书记指出:“人工智能应该是造福全人类的国际公共产品”“推动各方加强发展战略、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和标准规范”。当前,人工智能发展处于技术跃迁、应用爆发、治理探索的关键窗口期,地缘政治动荡和逆全球化思潮,严重影响全球协作和可持续发展。加快推动我国人工智能治理

理论创新、制度创新、实践创新,主动提出蕴含中国智慧、符合世界各国共同利益的治理方案,有助于构建一个更加公平、合理、包容、共享的全球人工智能治理体系,反对以意识形态划线或构筑“小院高墙”,防止全球人工智能治理的碎片化和阵营化,展现负责任大国担当,推动构建人类命运共同体。

加强人工智能治理需要把握若干基本原则

有效的治理必须以科学理念和清晰的原则作为指导。加强人工智能治理,要以习近平总书记关于人工智能治理的重要论述为指引,准确把握以下基本原则。

确立智能向善的价值取向。加强人工智能治理,应确保技术发展的最终目的是增进人类福祉、推动人的全面发展,实现智能向善。增进人类福祉,就是以人民群众获得感、幸福感、安全感为衡量标准,保障各个社会群体智能服务可及性,加速弥合“数字鸿沟”;推动人的全面发展,就是要保证人的主体性永远在场,让人工智能成为拓展人类认知边界和行动能力的“放大器”。只有将以人为本、智能向善作为人工智能治理的价值取向,才能确保技术进步不偏离人类价值坐标,使人工智能真正成为服务人民、造福社会的强大正能量。

贯彻统筹兼顾的系统思维。人工智能发展和安全具有高度内在一致性。从发展看,人工智能既是事关国家竞争力和国家安全的战略性技术,也是推动生产力跃迁的关键通用技术,不发展就难以在激烈的国际竞争中占领先机、赢得优势;从安全看,我国人工智能发展中面临的风险触点多、传导快,各类风险跨界性、关联性、穿透性较强,坚守安全底线才能在技术变革浪潮中行稳致远。处理好发展和安全动态依存的关系,必须牢牢掌握人工智能发展和治理主动权,以前瞻性统筹、整体性谋划、系统性布局,主动构建动态适配技术快速演进的治理框架,在可控边界内释放创新活力。

夯实良法善治的制度根基。加强人工智能依法治理是坚持在法治轨道上全面建设社会主义现代化国家的时代要求,能够更好为人工智能发展提供制度保障。从原则要求看,加强人工智能依法治理既要适应生产力进步所带来的新需求,也要回应技术产业发展伴生的新问题,进而实现以良法善治保障新业态新模式健康发展。从实践路径看,加强人工智能依法治理,需要把握人工智能发展趋势和规律,建设涵盖高中低位阶的法律法规、监管政策、技术标准等制度体系,明确数据、知识产权、开源模型等相关权责边界,统筹推进立法执法司法普法工作,真正做到科技发展到哪里、法治建设就跟进到哪里。

创新敏捷动态的治理模式。应将包容普惠导向贯穿人工智能治理全过程,推动算力、算法、数据等核心资源开放共享,降低技术应用门槛;支持开源社区建设和公共数据开放,促进知识扩散与技术扩散;保障各社会群体在智能时代的平等参与权与发展权,让人工智能发展成果更好惠及全体人民。针对人工智能应用的复杂性和不确定性,加快建立与技术演进相匹配的治理机制,探索“沙盒监管”、触发式监管等新型监管方式,降低制度性成本,为处于初创期、成长期的人工智能技术和应用建立必要的容错纠错机制,为市场提供稳定预期。

多维协同推进人工智能治理加速落地

理论的科学性最终要体现为实践指导性。贯彻落实习近平总书记关于人工智能治理的重要论述精神,应运用系统思维,构建覆盖人工智能研发、部署、应用与影响全链条的人工智能治理体系。

夯实数据治理基础。数据是人工智能发展的核心要素与战略资源,其质量、规模与治理水平直接决定了人工智能发展的深度与广度。加强数据治理,旨在突破数据要素化进程中的

关键瓶颈,构建从原始数据到智能价值的安全可靠通道。我国数据资源规模庞大、应用场景丰富,为人工智能发展提供了独特优势,但数据治理仍面临挑战,如数据基础质量与供给效率有待提升、数据安全威胁持续演进、生成式技术滥用导致数据生态污染等。为此,要创新数据要素市场化配置机制与生态,建立数据产权结构性分置制度;开展常态化的数据安全风险评估,对各类数据来源开展合规性审查,提升训练数据清洗效能;通过分场景、分层级、分应用进行数据管控,建立数据溯源体系,积极回应云端协同等新技术形态下用户数据保护难题,夯实人工智能发展的数据根基。

提升模型治理效能。人工智能大模型已成为智能系统的核心载体与能力中枢。模型自身的可靠性、安全性与价值观等,直接决定了立基于其上所有应用的品质与安全。因此,模型治理是人工智能治理体系中的关键环节。当前,模型安全面临新挑战,一是模型复杂度跃升,新型攻击手段难防御,补丁式防护思路不适用;二是模型风险识别难度高,一些风险指标量化难;三是随着智能体自主性的增强,其对物理世界和数字空间的运行秩序构成潜在威胁。我国已出台《生成式人工智能服务管理暂行办法》《人工智能生成合成内容标识办法》等文件,推动算法备案、大模型安全评估、内容标识等治理手段取得积极成效。下一步,应构建制度建设与技术创新并重、风险防控与应用赋能协同的模型治理体系。一方面,加强模型安全能力建设。可加强人工智能测试验证平台建设,支持第三方审计、评估、认证机构发展,提供模型测试验证、供需对接、内容标识检测等服务,落地模型对抗安全、后门安全、可解释性等检测能力,推进加固工具等技术开发共享。另一方面,可构建多主体协同联动的模型风险预警与应对机制。探索推动监管结果互认,统一监管要求与执行标准;推动企业定期报送风险阶段性评估报告,建立完善内部监测与预警机制;充分发挥行业组织的协调作用,加强对行业企业发展态势跟踪监测和合规评估。

优化应用治理生态。随着“人工智能+”行动纵深推进,人工智能正加速从

“能思考”向“能实干”转变,应用场景持续扩大。这在解放和发展生产力的同时,也导致技术应用呈现跨域融合与系统性外溢特征,风险从虚拟数字空间向物理实体空间、伦理价值范畴等不断扩展和深化。优化人工智能应用治理生态,应确立场景化、精细化的治理导向,摒弃“一刀切”的监管模式,针对不同应用领域、具体应用场景实施科学的分级分类管理;着力厘清人机协作新形态下的权责边界,确保在关键决策节点上人类始终掌握最终的价值判断权与控制权,切实防范技术失控可能带来的实质性伤害;致力于维护公平竞争的市场秩序,营造开放包容的创新环境,推动算力、模型、数据等核心要素开放共享,从而实现人工智能红利的广泛覆盖与社会公平。

细化伦理治理要求。人工智能的伦理治理,不仅是技术规则的制定,更是价值秩序的构建。从全球来看,关于算法公平、数据隐私、机器责任的讨论大都深植于各国的历史文化传统之中。中华优秀传统文化中蕴含的“民为邦本”思想、“厚德载物”精神、“天人合一”理念、“执两用中”方法等,为我们构建以人为本、智能向善的伦理框架提供了深厚文化根基。借中华优秀传统文化融入人工智能伦理治理,是形成中国特色治理范式、为全球贡献中国方案的关键路径。应着力构建人工智能伦理治理的文化内核,系统挖掘中华优秀传统文化资源,形成一套与社会主义核心价值观相契合的人工智能伦理规则体系,推动研究机构、企业等将文化伦理理念转化为具体的技术标准与评估工具,并以多种形式前置到产品开发流程中。

强化全球治理协同。人工智能治理关乎全人类命运,是世界各国面临的共同课题。近年来,受地缘政治、文化传统、体制差异等影响,全球人工智能治理导向发生深刻分化。特别是个别国家采取去监管化、去多边化、出口管制等策略,将人工智能视为大国竞争的战略工具而非共同治理对象,极大削弱了全球人工智能治理的协调性。为此,必须积极落实《全球人工智能治理倡议》,构建开放、公正、有效的治理机制,从对人类文明进步负责的高度,加快形成具有广泛包容性的人工智能治理框架。坚定维护联合国在国际治理中的作用,在充分尊重各国国家主权、法律体系及文化传统的基础上,形成具有广泛共识的国际治理框架和标准规范。坚决反对以意识形态划线或构建排他性集团,恶意阻挠他人工智能发展。倡导并践行真正的多边主义,通过多种方式支持发展中国家提升人工智能研发与应用能力,确保不同地区、不同群体公平地获取和受益于人工智能技术。

(作者为中国信息通信研究院院长)

莫把「痕迹」当「政绩」

李永领

福祉。

纠正“以痕论绩”的政绩观偏差,需要把握好三组关系。把握好手段与目的的关系。留痕是手段,落实是目的,手段可以服务于目的,但不能取代目的。因此,必要的留痕应当保留,但必须明确,留痕是为了更好地落实,而不能为了留痕而留痕。把握好过程与结果的关系。过程和结果相互依存,但过程的完备不等于结果的有效。有些工作过程轰轰烈烈、台账满满当当,结果却是“雨过地皮湿”,问题依然如故。真正的政绩,是让群众紧锁的眉头舒展了,发展的堵点痛点打通了,承诺的民生实事落地了。“金杯银杯不如百姓口碑”,真正的政绩不在一张张图片中、一页页记录里,而是在人民群众的心坎上。把握好形式与内容的关系。形式服务于内容,形式也反作用于内容。当形式过于繁复,就会挤压内容的空间,甚至吞噬内容。

莫把“痕迹”当“政绩”,更需要从制度层面发力。一方面,完善绩效考核评价体系,用好考核指挥棒,降低“留痕”材料权重,增加实效指标占比。比如,进一步明确必要的痕迹材料的范围,取消无实质意义的“打卡”拍照、台账要求;将群众满意度、问题解决率等作为考核核心指标;对不同领域、不同层级实施差异化考核,避免“一刀切”。另一方面,畅通群众监督渠道,让群众评价成为考核的重要依据。同时,加大对形式主义整治力度,对“造痕”迎检、弄虚作假的行为严肃问责。

唯成绩,在为民、在实干、在实效;痕迹,唯有服务于实、贴合于实,方有价值。党员干部树立和践行正确政绩观,必须坚决纠正“以痕论绩”的偏差,甩开“镜头”的包袱、走出“材料”的围城,真正沉到矛盾问题集中的地方去、群众最需要的地方去,真正以实干出政绩。

推动公共安全治理模式向事前预防转型

许欢

习近平总书记指出:“坚持从源头上防范化解重大安全风险,真正把问题解决在萌芽之时、成灾之前。”“十五五”规划纲要提出:“完善公共安全体系,推动公共安全治理模式向事前预防转型,切实维护人民群众生命财产安全。”推进公共安全治理模式向事前预防转型,不仅要有方法、环节之变,更要实现治理逻辑、能力结构的重塑,推动公共安全治理由“末端急救”向“源头防控、关口前移、系统减灾”转变。

我国发展处于战略机遇和风险挑战并存、不确定难预料因素增多的时期,公共安全形势总体稳定但局部严峻、隐患突出。极端天气频发,能源交通通信等关键基础设施联动紧密,城市生命线系统安全韧性亟待提升;数据与算法引发的新型风险与传统灾害交织叠加,呈现复合性、链式性、放大性特征。若公共安全治理仍以“事后处置”为主,易陷入“风险累积—事故爆发—治理追赶”的被动循环。推动公共安全治理模式向事前预防转型,是顺应风险演进规律、推动治理升级的必然要求,能让安全成为发展的“先手棋”。这一转型并非简单增加排查,而是将安全要求嵌入规划、建设、生产、运行、服务全链条,形成制度化、常态化、可度量的系统能力,其核心涵义有三:一是以本质安全托举高质量发展,将其融入发展全过程各环节,从源头上消除或减少危险因素;二是以“少出事、不出大事”的硬标准和“提前发现、有效避免”的硬能力,满足群众对高品质安全的期待;三是以数字化、智能化拓展预防空间,从“经验判断”转向“数据驱动、模型赋能”,实现风险可视、可测、可控。

健全风险治理制度体系,为关口前移夯实保障。制度设计的关键作用,在于通过前置约束、刚性规范和常态化机制,从源头上减少风险生成和隐患积累。要形成“规划有红线、建设有门槛、运行有标准、监管有抓手”的制度闭环,将风险评估前置国土空间规划、重大工程立项等环节。完善“风险清单—隐患清单—

责任清单—整改清单”四单管理,对高危行业、老旧小区、地下空间、燃气管网、老旧桥隧等实施分区分级、分类分色管控。完善突发事件相关法律法规及配套制度,加快应急与行业安全标准修订,以法治化、标准化和责任追溯固化事前预防成效。比如,北京市应急管理局印发《北京市突发事件应急预案管理办法》,构建覆盖市、区、乡镇(街道)三级的应急预案体系,强化应急管理全流程制度保障。此外,要严格落实党政同责、一岗双责、齐抓共管、失职追责,压实“三管三必须”,建立“谁审批谁负责、谁运营谁负责、谁监管谁负责”的责任链条,让每一项预防措施都有明确的责任主体、时限要求和监督路径。

健全监测预警体系,实现风险早识别早处置。织密监测预警网络,实质上是把风险感知触角延伸到全域全时,把应对环节前置到事故发生之前,为科学决策、精准防控、智能响应和高效联动提供支撑。可整合气象、水文、地质、海洋、消防、交通、能源、通信等数据,建设公共安全“一图统览、一网协同、一体联动”的综合监测预警平台,强化数据治理与模型供给,提升对极端场景的识别能力和对链式风险的推演能力。比如,湖南长沙的城市生命线系统汇聚融合全市时空底座、规划管控等数据资源,通过构建预警模型等方式,实现了对城市内涝、燃气等领域风险的高效管理和精准预报,有效提升风险防控能力。健全预警响应机制,完善预警“叫应”机制和避险转移预案。可探索将预警精准到网格、到社区、到家庭的新途径,提高预警信息传递效率,实现“人防+技防”协同发力。

提升基层应急管理能力和筑牢安全底座。基层是预防体系的承重层,基层能力强,预防就有着着力点;基层机制顺,预防就有稳定器。县乡村(社区)是风险早识别、早报告、早处置的前沿阵地。要健全“常态化巡查—直报直达—先期处置—快速联动”的工作机制,推动网格管理与专业支撑有机衔接。完善社区应急服务站、

微型消防站、避险场所等设施,推广“应急广播大喇叭”“叫应终端”等简便有效的工具,提升群众自救互救能力与组织动员效率。比如,浙江温州苍南县在2025年应对第4号台风“丹娜丝”的过程中,依托应急广播系统高频次发布预警信息,联动基层防汛责任人精准开展群众转移工作,实现地质灾害风险区人员安全避险。要充分发挥基层干部、网格员、群众、社会组织等多方主体作用,凝聚多方合力,全面提升群防群治水平。

以数字技术为预防赋能增效,提升预防的精度和效率。技术赋能的意义,不仅在于提高风险感知速度和处置效率,更在于推动公共安全治理由经验主导转向数据支撑、由事后响应转向事前推演、由分散处置转向协同治理,让“看得见、算得清、调得动、管得住”成为常态。要推动关键基础设施、重点企业和重要场景建设数字孪生与在线感知系统,开展设备健康诊断与预测性维护;建设跨地区、跨行业的“风险底图”,以场景化模型复盘历史事件、推演极端情形、检验预案可用性;完善安全数据要素的汇聚与合规流通,发挥保险与再保险在减灾防灾中的市场机制作用,探索灾害风险分担与激励约束并重的制度安排。比如,宁夏中卫市通过打造城市运行数据底座、建设全域覆盖的物联网感知体系等,对全域高风险区和安全敏感区进行安全监测,实现了对城市整体运行态势和风险的有效把控。

(作者为清华大学社会治理与发展研究院研究员)

