构建人工智能安全治理体系刻不容缓

中国电子信息产业发展研究院
党委副书记 胡国栋

- 加强人工智能安全治理是深入贯彻落实总体国家安全观的需要。
- 加强人工智能安全治理是防范人工智能技术现实风险的需要。
- 加强人工智能安全治理是应对各国抢抓安全治理主导权的需要。

党的二十届三中全会通过《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》，提出“建立人工智能安全监管制度”“完善生成式人工智能发展和管理机制”。当前，新一轮科技革命和产业变革深入发展，人工智能赋能经济社会千行百业，特别是生成式人工智能深刻影响人们的工作生活，也带来一系列新挑战。在推进人工智能创新应用的同时，要深入贯彻落实总体国家安全观，形成人工智能安全治理的强大合力，推动人工智能安全发展。

加强人工智能安全治理意义重大

加强人工智能安全治理是深入贯彻落实总体国家安全观的需要。新时代以来，以习近平同志为核心的党中央高度关注人工智能安全发展，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”。加强人工智能安全治理，关系加快培育和发展以人工智能为重要引擎的新质生产力，关系国家重大战略实施和重点领域安全能力建设，关系推进国家治理体系和治理能力现代化，关系推动人工智能造福全人类，必须统筹发展和安全，加强顶层设计、适当前瞻部署，坚持系统观念、协同推进，有力应对人工智能快速发展所带来的挑战。

加强人工智能安全治理是防范人工智能技术现实风险的需要。全球性安全组织开放式Web应用程序安全项目指出，人工智能大模型存在十大安全风险，包括提示词注入、训练数据投毒、模型拒绝服务、供应链漏洞、敏感信息泄露、不安全的插件设计、过度代理、过度依赖、模型盗窃等。我国奇安信集团发布的《2024人工智能安全报告》指出，2023年基于人工智能的深度伪造欺诈增加30倍，基于人工智能的钓鱼邮件增加10倍。

加强人工智能安全治理是应对各国抢抓安全治理主导权的需要。根据有关机构Insight Partners的分析，美国在公共安全领域的人工智能支出预计将从2022年的93亿美元增加到2030年的710亿美元，增长超过6倍。同时，美国宣布成立人工智能安全研究所联盟，汇集200多家科技企业、高等院校、金融组织和政府机构等单位，包括谷歌、英伟达、麻省理工学院、兰德公司等，旨在支持安全可信的人工智能的开发部署，争夺国际人工智能安全治理主导权。

人工智能安全治理的主要关切

人工智能安全治理攸关人类命运，引起国际社会各界广泛关注和讨论，目前各方主要关切以下领域。

一是尊重人工智能主权和文化多样性。联合国教科文组织强调在人工智能系统的整个生命周期内，确保尊重、保护和促进多样性，要符合国际人权法、标准和原则，以及人口、文化、性别和社会多样性。有关方面提出“主权人工智能”概念，认为人工智能强大的学习和推演能力使其成为世界文明的重要生产工具，可用来编纂整个国家的文化、社会智慧、常识、历史等各类数据，每个国家都应拥有自主人工智能基础设施，在保护自己文化的同时利用其经济潜力。

二是以人为本、智能向善。2023年12月，联合国人工智能咨询机构发布《以人为本的人工智能治理》报告，强调最大限度地发挥其对人类的益处。基辛格在《人工智能军控之路》一文中呼吁，从国家层面建立“全球人工智能秩序”，以防止人工智能最危险和潜在的灾难性后果。OpenAI的CEO山姆·奥尔特曼呼吁：“建立一个像国际原子能机构的机构来监督人工智能的发展。”

三是包容性、平等性。《布莱切利宣言》指出，发展人工智能要促进包容性经济增长、可持续发展和创新，强调“包容性人工智能和弥合数字鸿沟的重要性”。牛津大学对全球181个国家在公共服务中使用人工智能情况评估显示，得分最低的地区包括南半球大部分地区，例如撒哈拉以南非洲、一些中亚和南亚国家，人工智能准备程度的差异将加剧全球不平等。

四是安全、可靠、可信。保证人工智能系统生命周期中的数据集、流程和决策的可追溯性十分必要，参与者应采取与人工智能发展阶段相适应的系统风险管理方法，解决隐私、数字安全、偏见等相关风险。人工智能系统应保持透明和负责任的对外资料披露，促进公众对系统输入和输出内容的理解。

五是多边主义、国际合作。世界经济论坛《多方合作才能实现负责任的人工智能治理》报告指出，对先进的人工智能进行集体监督不仅有利，而且势在必行。全球局势正处于一个关键时刻，既要迅速发展人工智能技术，又迫切需要治理引导这一飞速发展的技术，在二者之间找到平衡，多方参与的人工智能治理是降低其技术风险并最大化利用其可能性的最合适的方式。

六是分类监管、可问责性。欧盟《人工智能法案》提出，根据人工智能对社会造成危害的能力，遵循“基于风险”的方法来分级监管人工智能：风险越高，规则越严格。2023年经合组织发布《推进AI问责制：围绕可信赖AI生命周期的治理和管理风险》报告，强调人工智能问责性原则，建议人工智能治理能够提供针对风险管理流程的审查，包括对流程和结果的监测、审查、记录、沟通和协商，同时提供多种机制，将人工智能风险管理流程嵌入更广泛的组织治理。

积极构建我国人工智能安全治理体系

人工智能安全治理体系要面向未来“人机物”协同社会、智能化产业体系、数实一体世界、人类命运共同体，坚持“多层、多维、多方”推进，从硬件层、模型层、工具层、应用层等多个层次，主权安全、产业链安全、意识形态安全和能源资源安全等多个维度，建立自主可控、智能向善、综合防护、架构健全的安全治理体系。

首先，构建“多层次”安全体系。

一是硬件层安全方面，确保人工智能产业的硬件底座安全。在芯片、服务器和智能终端层面，增强芯片硬件设计安全性，采用先进检测技术识别硬件木马或后门。在智算中心和算力网络层面，实施物理层安全技术加密，强化身份认证和密钥管理，加强网络安全管理。二是大模型安全方面，增强人工智能应用的安全性、可靠性和合规性。有效管理人工智能研发、部署和应用，构建源头语料审查、生成内容审核、伦理规范对齐等机制，开展通用、专用领域模型的评测审计，识别潜在风险，推进最佳实践。三是工具层安全方面，维护我国人工智能开发生态安全。推动自主操作系统研发，减少对外技术依赖。积极推广基于自主技术的并行计算框架与工具链，确保人工智能开发软件安全。四是应用层安全方面，确保人工智能安全嵌入现有工控系统、生产装备、社会运行和基础设施。在汽车、机器人、石化等工业领域，加强人工智能系统安全测试，确保智能系统和设备安全可靠运行。在民生、公共服务领域，严格用户数据管理和安全审计，确保社会治理有序高效。在交通、电力、金融等关键基础设施领域，充分利用人工智能技术建立监测预警系统，建立部署人工智能系统后的网络隔离、数据加密、入侵检测和人类及时接管等安全机制。

其次，推进“多维度”安全治理。

一是主权安全方面，从技术与内生安全、数据与隐私安全、系统与应用安全、检测与监管安全等多方面综合施策，提高国家人工智能系统的抗攻击能力。二是产业链安全方面，从关键技术保障、前沿科技布局、产业链自主可控、产业生态协同、标准制定实施等方面入手，提升我国在全球人工智能产业中的话语权和影响力。三是意识形态安全方面，加强对数据语料、算法规则、价值导向的管理，确保生成式人工智能产出内容符合国家要求、彰显中国特色。四是能源资源安全方面，加强人工智能产业用能用水趋势预判，完善人工智能能效水效评价及标准体系，加强人工智能产业节能降碳节水技术攻关，优化智算中心空间布局。

最后，形成“多方面”工作合力。

一是促进政府、企业联动。在加强政府对人工智能发展监管的同时，发挥内部大市场和新技术高效应用推广的优势，鼓励企业通过自愿行为准则进行自我监管，探索适应不同行业应用需求的监管规则。二是促进科技、产业联动。建立产学研合作平台，支持高校和研究机构的安全技术创新，联合制定安全技术标准和规范，推动产业界人工智能应用安全能力建设。三是促进“平急两用”联动。人工智能大模型是新型基础设施的关键底座，要确保在平时和应急情况下都能发挥效能，有效提高使用效率和公共安全保障能力。四是促进国内、国外联动。主动参与全球人工智能安全治理体系分工合作，深度融入全球创新网络，积极参与国际标准制定和国际组织构建，提出人工智能安全治理中国方案，引领全球人工智能安全治理向更加科学合理、公平正义方向发展。

2024金砖国家新工业革命伙伴关系论坛在厦门召开

（上接第1版）大力发展新一代信息技术产业，加快制造业数字化转型，构建数字产业生态合作网络，开展人工智能技术交流、产业对接、项目合作和能力建设。四是倡导普惠包容，促进共同发展。加强分工协作，共同维护全球产业链供应链韧性和稳定。中方将继续通过金砖创新基地提供更多公共产品，推动设立金砖国家工业能力中心中国分中心，向更多发展中国家开放合作资源。

周祖翼指出，福建是中国改革开放的重要省份。近年来，我们牢记习近平总书记嘱托，不断拉紧与金砖国家的合作纽带，推动金砖国家新工业革命伙伴关系创新基地建设走深走实，取得丰硕成果。本届论坛是金砖国家历史性扩员后的首届论坛。福建将以此为契机，深入贯彻党的二十届三中全会精神，秉承开放包容、合作共赢的金砖精神，加强同金砖各国的交流合作，努力为深化金砖国家新工业革命伙伴关系作出新的贡献。我们将加强与金砖国家在产业链供应链方面的合作，积极引进金砖国家相关产业项目，支持福建企业到金砖国家拓展业务，构筑开放、韧性、高效、稳定的产业链供应链体系；加强与金砖国家在发展新质生产力方面的合作，合力推进共性关键技术联合攻关，加快行业数字化转型、智能化改造，共同打造现代化产业体系；加强与金砖国家在人才培养方面的合作，创新金砖人才培养模式，架起金砖人文交流的“连心桥”，助力金砖合作不断迈上新台阶。真诚期待各方积极参与金砖创新基地建设，带动更多资源来福建投资兴业，深化共商共建，实现共赢共享。

本次论坛以“携手构建高质量伙伴关系，开启新型工业化合作新征程”为主题，其间举办部长圆桌论坛和产业高峰对话，分享各国工业化、数字化发展实践，并就深化伙伴关系合作进行深入交流。国内领军企业、智库、金融机构负责人参与产业高峰对话，共同探讨深化金砖产业合作机遇和路径。围绕人工智能、智能制造、低空产业、能源电子、工业设计、产融合作、产业科技创新、历史经典产业、数字产业生态、工业绿色低碳等领域举行10场专题研讨。

论坛上，中国、俄罗斯、巴西、埃塞俄比亚、伊朗、玻利维亚、古巴、圭亚那、科威特、摩洛哥、尼日利亚、塔吉克斯坦、乌干达、委内瑞拉、津巴布韦等15国主管部门联合发布《新型工业化国际合作倡议》。金砖创新基地等13家中方单位正式加入联合国工业发展组织全球工业与制造业人工智能联盟。全球发展倡议新工业革命伙伴关系网络发布合作成果，该合作网络已包含19个国家的111家多元主体。论坛首次发布《金砖国家产业合作案例集》，举办了金砖国家工业创新大赛决赛、金砖国家人工智能技术与治理卓越人才研修班、金砖国家新工业革命展等系列活动。

论坛由工业和信息化部、福建省人民政府共同主办，中共中央对外联络部、外交部、国家发展和改革委员会、科学技术部等部门相关司局负责人，以及来自31个国家的政府主管部门、驻华使节、知名企业、行业协会、智库、金融机构代表，联合国工业发展组织、数字合作组织等国际组织代表，共计500余人参会。

山西：加快培育新质生产力 奋力推动制造业振兴升级

（上接第1版）山西工信系统将深入学习贯彻党的二十大和二十届三中全会精神，以及习近平总书记对山西工作的重要讲话重要指示精神，以进一步全面深化改革为关键抓手，固根基、扬优势、补短板、强弱项，扎实推进新型工业化和制造业振兴升级，为推进中国式现代化山西实践作出新的更大贡献。

一是要夯实工业基础支撑。围绕“促增长”“稳投资”，着力从点上强企（企业）、连点成线（行业）、扩线成面（区域），“三位一体”稳定全省工业经济运行，不断夯实高质量发展基础。全面提升产业服务水平，探索组建制造业振兴专项基金，深化干部包联和入企服务工作。抢抓国家“两新”“两重”等各项政策机遇，加快推进企业设备更新和技术改造，持续提升产业发展能级和水平。

二是要推进产业结构优化升级。坚持传统优势产业、新兴产业、未来产业“三业”并举。传统优势产业要加快在振兴上破题，落实国家标准提升要求，支持企业用数智技术、绿色技术改造提升传统优势产业，提高先进产能占比，优化产品结构，提升产品附加值，实现焕新发展。新兴产业要加快在规模上破题，围绕高端装备制造、电子信息等新兴产业，加强政策供给，承接产业转移，持续壮大产业规模。未来产业要坚持精准选择、点上突破，发挥既有装置作用，沿途下蛋、培育转化，力争抢占制高点。

三是要推进科技创新和产业创新深度融合。深入推进“四链”融合，主动融入国家科技创新体系，推动制造业创新中心、企业技术中心等创新平台建设，进一步强化高质量科技创新供给，打造以企业为主体、产学研深度融合的技术创新体系。探索建立技术创新成果转化落地机制，在工信领域开展“揭榜挂帅”工作，支持实施若干具有引领性、协同性的创新项目，推进创新成果转移转化。强化“晋创谷”与新型工业化战略协同、政策衔接，放大“晋创谷”效应。

四是要促进数实融合数智赋能。加大制造业智改数转网联力度，布局建设5G基站、数据中心、算力中心等数字基础设施，拓展“5G+工业互联网”创新应用，筑牢数字基础设施底座。深入推进制造业数字化转型，实施中小企业数字化改造，推动工业互联网在煤炭、装备制造等重点行业广泛应用。深入开展智能制造诊断评估工作，开展“人工智能+”行动，支持企业开展智能化改造。

五是要提升产业链供应链韧性和安全水平。健全强化重点产业链发展机制，深化产业链“链长制”，推动山西新能源、新材料等领域产业链融入国家产业链体系。支持中小企业专精特新发展，与行业龙头企业、“链主”企业共同参与产业链上下游协同攻关，保障产业链供应链稳定畅通。完善承接产业转移协作机制，深化与京津冀、长三角、粤港澳大湾区等地区合作，大力推广“政府+链主+园区”招商模式，高水平有序承接产业梯度转移。

金砖国家开启新型工业化合作新征程

（上接第1版）“一是继续发扬伙伴精神，坚持共商、共建、共享的基本原则；二是充分利用金砖国家多元、互补的特点，充分发挥各自的优势，形成互补效应；三是结合金砖国家的地缘与分布特点，将国家发展转化为区位优势，通过国家带动洲际整体发展。”邹刺勇告诉记者，金砖这一平台立足务实的产业对接合作，如果使用得当，将对金砖国家与全球其他国家的合作，特别是南南框架下的合作具有引领性意义。

优势互补发展趋势初步形成

“新五国”的加入，为金砖国家形成优势互补格局提供了新的抓手。金砖国家智库合作中方理事会秘书长、中共中央对外联络部研究室主任金鑫对此做出了详细解读：“新加入的国家中有几个是能源生产国，而中国、印度等国家是能源消费国；俄罗斯是粮食生产大国，而部分非洲国家则仍然面临缺少粮食的困境……当前，无论是原材料、能源、粮食，还是大宗商品，都能形成优势互补的局面。”

除本次历史性扩容外，金砖国家新工业革命伙伴的队伍仍在不断壮大——全球发展倡议新工业革命伙伴关系网络已包含19个国家的111家多元主体；论坛期间，金砖创新基地等13家中方单位正式加入联合国工业发展组织全球工业与制造业人工智能联盟……源源不断的新鲜血液，无疑能够对金砖国家发挥互补优势、激发发展潜能起到积极作用。

谈到未来金砖国家如何深挖互补优势潜力，金鑫给出了“三条公式”：“首先要达到‘5+5=1’，即新老成员国要发出一致的声音，形成一致的行动；同时要实现‘5+5>10’，加强金砖国家间的互补合作，形成合力，反哺整体发展；此外，还要积极进行‘5+5+N’模式的合作，新老金砖国家要与其他新兴市场国家、其他南方成员国家进行合作，共同推进工业化进程。”

在本次论坛上，这一发展趋势已经初现雏形。记者了解到，本次论坛邀请了40个国家，以及新开发银行、联合国工业发展组织、数字合作组织等国际组织的代表参加，充分扩大交流覆盖面；此外，俄罗斯外交部发言人扎哈罗娃也于近日表示，已收到数十份加入金砖国家合作机制的申请。金砖国家进行大面积交流合作、充分发挥互补优势、实现共赢发展，已成大势所趋。

共同奏响新型工业化和声

“纵观全球工业发展，呈现出绿色化、数智化两大趋势，这也是金砖国家推进新型工业化的两个主要方向。”邹刺勇表示。

金砖论坛期间，中国、俄罗斯、巴西、埃塞俄比亚、伊朗、玻利维亚、古巴、圭亚那、科威特、摩洛哥、尼日利亚、塔吉克斯坦、乌干达、委内瑞拉、津巴布韦等15国主管部门联合发布《新型工业化国际合作倡议》，从注重创新驱动、努力实现绿色低碳可持续发展、探索符合各自特色的新型工业化路径等七方面提出合作建议，旨在共同推进新型工业化。

“面对新一轮科技革命和产业变革机遇，我们需要全球协作、携手并进，积极推动和加强有效的伙伴关系。”圭亚那国家数据管理局局长克里斯托弗·迪恩在此次金砖论坛“部长圆桌论坛”上表示。摩洛哥工业与贸易大臣利雅德·梅祖尔说，摩洛哥愿意与金砖国家开展更多双向交流，借鉴今天金砖国家的成功经验，在新型工业化道路上不断迈进，携手持续推进工业化道路创新。

金砖创新基地产业创新联盟理事长陆益民建议，要把握时代发展机遇，扩大“金砖朋友圈”，努力打造促进双向投资的国际公共平台，围绕智能制造、绿色工业、新能源、中小企业数字化转型等热点领域，助力金砖国家产业合作。

记者了解到，不少中国企业正发挥自身优势，携手“金砖+”伙伴，共同推进工业化进程。例如，宁德时代首席科学家吴凯透露，宁德时代正在与道达尔集团合作，在非洲建设其最大储能电站项目。据悉，宁德时代连续7年位居全球动力电池使用量第一，已成为推动全球新能源汽车产业发展的关键力量。“宁德时代将以此为契机，进一步加大对非洲市场的开拓，用高安全长寿命电池助力金砖国家储能设施高水平建设。”吴凯表示。