技术与安全“双轮”驱动AI产业发展

热点透视 rediantoushi

新技术应用往往先于规范，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，才能营造良好的创新生态。着眼未来，在重视防范人工智能风险的同时，也应同步建立容错、纠错机制，努力实现规范与发展的动态平衡。

近期，人工智能领域大模型声势高涨，短短数月间，数十家国内公司先后宣布进军大模型赛道。7月6日在上海举办的2023世界人工智能大会成为大模型集中秀场，30余款来自不同企业的大模型产品和技术相继亮相。

科技部新一代人工智能发展研究中心近期发布的《中国人工智能大模型地图研究报告》显示，以“大数据+大算力+强算法”相结合的人工智能大模型，在中国正迅猛发展，中国研发的大模型数量已居全球第二，仅次于美国，目前中国发布的10亿参数规模以上的大模型已达79个。

以大模型为代表的人工智能技术，其巨大潜力正在加速释放，想要抓住它所带来的巨大机遇，就需要警惕它的潜在安全风险和隐患，这是当前人工智能产业界所面临的双向任务。

人工智能风险具“自有”特点

据中国信息通信研究院测算，2022年中国人工智能核心产业规模已达5080亿元人民币。

人工智能已成为全球数字技术创新最活跃的领域之一，为人们的生产生活带来了巨大的变革和便利，但也带来了诸多风险与挑战，如何构建安全可信的人工智能是当前各界关注的焦点。

阿拉伯信息通信技术组织秘书长穆罕默德·本·阿莫在日前举办的世界互联网大会数字文明“尼山对话”主论坛上表示，建立一个安全可信的人工智能系统，要首先考虑数据隐私与安全、透明度、责任与问责、稳健性与弹性等因素；此外，还需要建立一个道德框架，通过以人为本的设计，优先考虑人类福祉。

全球移动通信系统协会首席执行官洪曜庄表示，只有在道德准则的约束下，人工智能才能真正改善世界。我们必须共同努力构建一个可信赖的环境，建立以人为本的方法体系，确保人工智能对于每个人都可靠、负责和公平，最重要的是，能够普惠所有人。

加强人工智能发展中潜在风险的研判和防范，维护人民利益和国家安全，确保人工智能安全、可靠、可控，是中国推进人工智能治理的重要方向。

早在2017年，国务院印发《新一代人工智能发展规划》，明确提出到2025年，初步建立人工智能法律法规、伦理规范和政策体系，形成人工智能安全评估和管控能力；2030年，建成更加完善的人工智能法律法规、伦理规范和政策体系。

尽管相关科研机构和科技企业在人工智能系统设计之初就考虑到要保障人工智能安全、可控，但技术应用的趋利避害却往往难以一蹴而就。

中国科技大学公共事务学院网络空间安全学院教授左晓栋表示，人工智能的风险和传统技术的风险相比，有一些“自有”的特点。比如，人工智能是高度自主智能的，极大依赖数据基础，而且还存在“算法黑箱”、算法不可解释等问题，使得人工智能系统存在大量未知因素，风险预测难度较大。

为应对可见的挑战和不可知的风险，我国应加快建立人工智能领域相关法律法规、伦理规范和政策体系，形成人工智能安全评估和管控能力，这是我国各界对人工智能发展的共识。

大模型“放大”AI安全问题

百度创始人、董事长兼首席执行官李彦宏说：“过去一年，人工智能在技术、产品、应用等各个层面，以‘周’为迭代速度向前突进。大模型成功压缩了人类对于世界的认知，让我们看到了实现通用人工智能的路径。”

以今年3月百度率先发布的大语言模型文心一言为标志，我国大模型创业潮奔涌，伴随而来的，是社会各界越来越多对于大模型安全的疑惑。但在百度看来，安全问题并不是大模型出现才带来的新问题。

“大模型之前的人工智能时代，我们已经发现人工智能本身具有所谓的内在安全问题。人工智能算法可能会被对象样本攻击，正常样本加入少量对抗就会误导识别结果。不管是数字世界还是物理世界，很多场景都存在这种情况。”清华大学计算机系长聘教授、清华大学人工智能研究院副院长朱军指出。

在朱军看来，特别是ChatGPT出现以后，生成式人工智能的安全问题越来越严重，而算法本身是否存在政治偏见和数字鸿沟，数据采集过程中会不会侵犯知识产权等问题，也是大模型时代需要重点关注的问题，可从以下几个层面尝试解决。

首先，从人工智能基础层面，针对深度学习、深度神经网络，学术界一直在探索第三代人工智能新范式，希望能够发展更加安全可靠的人工智能框架。第三代人工智能新范式的优势就是在于安全、可信、可靠和可拓展。

视觉中国供图

其次，提升安全评测能力，主要关注对抗攻击评测、角色扮演与诱导欺骗评测、混淆指令欺骗评测、标识性能评测、数据安全评测、伦理安全评测等。

还有，构建人工智能安全治理有效工具，如可以构建人工智能安全平台，通过平台化的方式对人工智能的算法和服务进行评测。

业界不断尝试新的“解题思路”

正如北京智源人工智能研究院院长、北京大学多媒体信息处理全国重点实验室主任黄铁军所言，人工智能越来越强大，风险与日俱增，但对于如何构建一个安全可信的人工智能，我们仍知之甚少。

面对这样的现实，业界一直在不断尝试新的“解题思路”。

6月28日，火山引擎发布大模型服务平台“火山方舟”，面向企业提供模型精调、评测、推理等全方位的平台服务。

“企业使用大模型，首先要解决安全与信任问题。”火山引擎总裁谭待表示，“火山方舟”实现了大模型安全互信计算，为企业客户护佑数据资产安全。基于“火山方舟”独特的多模型架构，企业可同步试用多个大模型，选用更适合自身业务需要的模型组合。

与小模型的“自产自用”不同，大模型的生产门槛很高，数据安全成为大模型时代的新命题。谭待认为，企业使用大模型，最担心的是数据泄露，但如果将大模型私有化部署，企业将承担更高的成本，模型生产方也会担心知识资产安全。“火山方舟”的首要任务就是做好安全保障，使大模型使用者、提供者和云平台各方可以互相信任。

据火山引擎智能算法负责人吴迪介绍，“火山方舟”已上线了基于安全沙箱的大模型安全互信计算方案，利用计算隔离、存储隔离、网络隔离、流量审计等方式，实现了模型的机密性、完整性和可用性保证，适用于对训练和推理延时要求较低的客户。

黄铁军表示，所有的探索才刚刚开始，我们面临着全新的挑战，原有的经验和方法可能都无法解决新问题。

“新技术应用往往先于规范，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，才能营造良好的创新生态。着眼未来，在重视防范人工智能风险的同时，也应同步建立容错、纠错机制，努力实现规范与发展的动态平衡。”李彦宏说。

但是，无论从技术趋势，还是产业应用来看，大模型都绝不是昙花一现的风口，而是影响人类发展的重大技术变革，是拉动全球经济增长的重要引擎，是绝对不能错过的重大战略机遇。李彦宏说：“坚持技术发展和安全可控的双轮驱动，才能行稳致远。如果我们安全、负责任地驾驭人工智能发展之路，大模型就会重塑数字世界，人工智能就可以为中国经济乃至全球经济创造无与伦比的繁荣，提高全人类福祉。”

刘艳

青岛海关：为高水平开放创新提供服务支撑

科学观察 kexueguancha

上合示范区多式联运中心、上合示范区原产地证书审签中心等“上合元素”集聚发力，为这座“上合之城”的蝶变注入新动能。

7月5日，谈起中国—上海合作组织地方经贸合作示范区（以下简称上合示范区）5年来高质量发展的成就时，青岛海关所属胶州海关副关长徐月静在接受科技日报记者采访时表示，这其中得益于青岛海关聚焦国际物流、现代贸易等重点建设任务，持续搭建开放平台、创新监管模式、拓宽物流通道、优化服务举措，先后确定“8+12”项重点支持项目，推出“532”工程，支持上合示范区高质量发展。

数字最有说服力。2018至2022年，青岛市对上合组织其他成员国累计进出口2114.7亿元，年均增长13.7%。今年前5个月，青岛市对上合组织其他成员国进出口336亿元，增长68.5%。

中欧班列“趟趟爆满”

7月2日，在青岛海关所属胶州海关办结过境运输手续后，一列中欧班列过境货物专列从上合示范区多式联运中心发出。本次班列主要载有来自日韩的农用机械、汽车及配件，共110个标准集装箱，货值约530万元人民币，去往吉尔吉斯斯坦。

集装箱进出不停，班列穿梭往返，炎炎夏日里，上合示范区多式联运中心一番火热发运场景，“趟趟爆满”是这里中欧班列的真实写照。从东南亚、日韩入境的集装箱，经过中心转运至中亚和欧洲，这是一条最便捷的物流通道。本地企业创新开展冷链、跨境电商、对外承包工程等特色货物专列，使得班列运行线路持续拓展，发运货物种类不断丰富。

图为6月30日，上合示范区多式联运中心，办结海关监管通关手续的集装箱被吊装入中欧班列等待发运。　陈星华摄

青岛海关打造连接青岛港和上合示范区多式联运中心的“上合海铁通道”物流通关环境，不断优化过境和出口货物监管手续，服务中欧班列高质量运行。

截至目前，上合示范区已常态化开行21条国际班列线路，通达上合组织和“一带一路”沿线23个国家54个城市。五年间，青岛海关共监管上合示范区到发中欧班列2765列，年均增长约35%。

开放创新活力迸发

作为一家上合示范区内向日本出口床单、被罩、枕套等床上用品的企业，青岛吉诚纺织有限公司通过自主研发，创新产品涉及种类多，对应众多不同税号。“哪些能减免关税，我们也搞不懂，海关专家帮我们开展针对性分析指导，精准指导我们享惠。”该公司业务经理王婷婷告诉科技日报记者，从2022年2月企业开始申请RCEP证书，至今已签发了500多份。受利好驱动，该企业出口订单开始逐渐增加，继去年实现对日本出口额增长近10%，今年相关订单已排到年底。

这是去年《区域全面经济伙伴关系协定》（RCEP）生效以来所产生的积极效应。徐月静介绍，青岛海关率先在上合示范区建立RCEP监管服务创新试验基地（上合），并联合上合示范区管委会设立上合示范区原产地证书审签中心，成为国内首个以服务上合组织成员国经贸合作为特色的原产地证书审签中心。

RCEP是近两年来我国稳步扩大制度型开放的“热词”。当“上合”遇上“RCEP”，青岛海关用一次次制度创新推动更高水平开放和更高质量发展。记者了解到，胶州海关在上合示范区原产地证书审签中心开展专班服务、专业指导、专窗办理“三专”工作，针对重点享惠商品量身定制帮扶措施，通过贸易数据比对分类开展靶向推介，指导企业享惠RCEP协定。

最新统计数据显示，上合示范区原产地证书审签中心成立以来，已累计为1000余家出口企业签发原产地证书7万余份，货值超10亿美元。

为破解国际经贸合作的痛点、堵点、难点，2022年11月，中国—上海合作组织地方经贸合作综合服务平台上线发布，实现了“贸易+通关+物流+金融”全周期、全要素、全链条一站式综合服务，成为中国国际贸易单一窗口三大专区之一。

青岛海关深度参与该平台建设，结合辖区上合示范区企业实际需求，开发了一系列本地特色功能，为企业开展外贸业务、拓展国际市场提供便利。截至目前，该平台已累计注册企业近5000家，认证供应商731家、认证采购商191家。在今年6月份举办的上海合作组织产业链供应链论坛暨2023上合国际投资贸易博览会上，该平台的2.0版本发布上线，创新增加了多个服务系统，将继续有效降低企业综合通关物流成本。

徐月静表示，胶州海关积极推进“信用上合”建设，通过与地方政府联合选企、联合培育、联合激励，共建上合示范区高级认证企业产业园。目前，上合示范区内共有13家高级认证企业，包括5家上市公司、2家专精特新企业和2家行业国际龙头企业。

王健高

让新型研发机构发挥更大作用

创新杂谈 chuangxinzatan

许强

在不久前举办的中关村论坛上，一大批赋予科技人员更大自主权、更灵活管理模式、更精准资源支持的新型研发机构受到广泛关注。作为基础研究和应用研究的融汇点，新型研发机构不仅是“从0到1”的原始创新策源地，而且担负着“从1到10”的产业技术研发任务，是推动我国实现高水平科技自立自强、建设世界科技强国的有力支撑，被寄予提升国家创新体系整体效能的重任。

党的十八大以来，北京、上海、广州、深圳等地探索建设了多种形式的新型研发机构，发力打造国家战略科技力量，取得了丰硕的科创成果。短短几年，这些机构在《自然》《科学》《细胞》等国际顶尖科学刊物发表高水平科技论文超过百篇，推出了全球首款96核区块链专用加速芯片、长寿命超导量子比特芯片、量子直接通信样机等一批世界级原创成果。分析这些机构成功的原因，以下几点经验值得借鉴。

给予科研单位更多自主权是新型研发机构健康发展的机制保障。北京的新型研发机构大多推行理事会领导下的院（所）长负责制，赋予实验室主任在选人用人、技术路线等方面全方位的自主权，在确定的重点方向、重点领域范围内可以自主确定研究课题。这样的体制机制给予科研单位更多自主权，赋予科学家更大技术路线决定权和经费使用权，让科研单位和科研人员从烦琐、不必要的体制机制束缚中解放出来。

持续经费支持是新型研发机构专注创新的物质基础。北京新型研发机构多以5年为一周期享有财政科技经费稳定支持，编制年度预算，实行负面清单管理，研发人员根据已确定的研究课题自主安排科研经费使用。除特殊规定外，北京市财政资金支持产生的科技成果及知识产权由新型研发机构依法取得，自主决定转化及推广应用。比如，北京脑科学与类脑研究中心实行科研经费包干制，科研人员无须填写基金申请，即可获得长达6年的稳定支持。

以为人才松绑减负为主要目标的新机制让新型研发机构动力强劲、成果迸发。在人才评价上，很多新型研发机构坚持“破四唯”和“立新标”并举，加快建立以创新价值、能力、贡献为导向的科技人才评价体系。不以论文为导向，由理事会下设的评估委员会进行评估，围绕科研投入、创新产出质量、成果转化、原创价值、实际贡献、人才集聚和培养等，作出符合机构设立目标和符合科研规律的国际同行评价。

总体来看，北京在新型研发机构建设及创新方面已初见成效。与此同时，也要看到，新型研发机构还是个新事物，仍需不断探索和完善。比如，进一步明确新型研发机构的定位，聚焦国家战略需求，形成超前谋划引领前沿研究方向的重要力量；不断优化和提升治理能力，建立健全“全职院长+项目经理人+学术院长”的结构，完善双聘人员兼职取酬政策；探索多种途径的资金支持方式，为科研项目长远发展造血；通过不断夯实新型研发机构的制度保障，为创新松绑、为创业加油、为创造助力，促进科技创新和经济社会发展深度融合，共同托举实现高水平科技自立自强。

±500千伏延庆换流站有了国产『心肌』

科学导报讯　7月6日，随着双极OLT试验（空载加压试验）顺利完成，±500千伏延庆换流站完成本年度年检，进入送电阶段。值得一提的是，今年年检中，国网北京电力首次将国家电网有限公司研发的具有自主知识产权的4500伏/3000安IGBT（绝缘栅双极型晶体管）在延庆换流站挂网应用。

国网北京检修公司柔直调相机运检中心副主任周凯告诉笔者，IGBT是一种复合全控型电压驱动式功率半导体器件，如果说换流阀是延庆换流站的“心脏”，IGBT就是心脏里的“心肌”。相比传统大直流工程换流阀采用的半控器件晶闸管，IGBT既能控制导通，又能控制关断，是柔性直流换流阀的核心器件，就像一个智能开关，能够自动、快速地实现能量高效转换。以往使用IGBT的换流站里，90%的IGBT依赖进口，特别是3300伏以上高电压等级的IGBT。具有自主知识产权的4500伏/3000安IGBT的挂网应用，将推进自主研发的功率半导体器件在智能电网领域的应用。据了解，今年±500千伏延庆换流站检修工作时间为9天，共有来自全国24家参检单位的565名电力检修人员分赴23个作业面，投入使用46辆大型机具、320台套检修试验仪器，集中开展例行修试项目16594项、特殊性检修25项、重点检查验证30项等。

周凯介绍，今年年检中，在换流阀、直流断路器、直流控制保护、直流测量装置等核心设备作业面上，国网北京电力首次安排延庆换流站内自有运检人员担任工作负责人及主要工作班成员，完成例行检修试验、可靠性提升、状态感知能力提升，确保设备以最佳状态迎接迎峰度夏大负荷考验。

张北柔性直流电网由中都、康巴诺尔、阜康和延庆4座换流站和666公里的输电线路组成。作为张北柔性直流电网试验示范工程重要受端换流站，±500千伏延庆换流站年检将进一步提升站内设备安全稳定运行状态，畅通迎峰度夏绿电进京通道，为首都电网提供安全稳定的绿色电能，为迎峰度夏电力保供注入“强心剂”。

郭菲