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。

技术与安全“双轮”驱动AI产业发展

◎本报记者 刘 艳 杨 雪

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。

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。”

大模型让轨道交通更“聪明”

◎本报记者 叶 青

基于AI的建筑信息模型（BIM）智能审查平台、时速80公里A型中国标准地铁列车、轨道交通一体化智能电力监控系统……在广州大湾区轨道交通论坛暨展览会上，智能交通“黑科技”集中亮相。论坛上首次发布的佳都知行交通大模型展示出了“高智慧”，无论是服务场景的实时问答，应急指挥场景的统筹处置，还是地铁运维场景下复杂的故障处置任务，佳都知行交通大模型都能够应对自如。

当前AI大模型所展现出的能力与智慧，已远超过去大众对于AI的认知，逐渐从“弱人工智能”走向“强人工智能”。随着计算和存储技术的不断发展，类似ChatGPT这样的大模型，规模将会不断扩大，应用场景不断扩展，平台化的通用模型与行业专用模型相结合，将成为人工智能应用的重要方向。

“通用大模型的诞生只是一个起点，其最终还是要落脚到特定应用场景、解决行业具体问题上。”佳都科技集团（以下简称佳都科技）董事长刘伟表示，以行业大模型为关键驱动，赋能行业生产效率和服务质量提升，将会推动经济社会发展和产业的深刻变革。而实际应用场景的数据和反馈优势，又将进一步加速行业大模型技术迭代，提高国产大模型的竞争力。

据介绍，佳都知行交通大模型采用了混合云多模技术，整体上形成大模型基础设施、大模型中间件、大模型行业应用三层架构，从而进行“预训练+精调+基于行业数据反馈”的强化学习。

此外，佳都知行交通大模型已经具备面向轨道交通的智能客服、智能运维以及应急指挥三个场景的落地应用能力，其智慧化程度高于目前行业基本应用现状，能够促进轨道交通的数字化转型和智能化提升。

“佳都知行交通大模型可化身‘数字运维专家’，与维修人员进行单轮或多轮对话，协助其排查故障原因、分析解决方法并提供维修辅助决策，提升维修效率及维修质量。同时，通过大语言模型支撑辅助能力，可为城轨提供应急事前、事中、事后处置服务。”佳都科技智能轨道交通行业部副总裁严波表示，未来，佳都科技将以佳都知行交通大模型为基础，打造以乘客为中心、以信息为渠道，主动式、全链条乘客出行服务体系。

电力“AI军团”备战亚运会

◎本报记者 江 耘 实习生 卢馨怡 通讯员 钱 英

“石桥变10千伏朝湖州9419线跳闸。”7月6日，国网杭州供电公司供电服务指挥中心配网调控大厅内，供服虚拟数字员工智能监屏员发出响铃并告警。与此同时，数字员工“抢修专家”迅速锁定停电区域并通知抢修人员快速前往处理……这是国网杭州供电公司开发的虚拟数字员工办公平台首次协同作战演练。

去年3月，国网杭州供电公司启动虚拟数字员工建设方案，虚拟数字员工办公平台搭建包括虚拟调度、智慧监屏、智能抢修、虚拟办公、虚拟调度、智慧客服6大功能模块，分别对应6位数字员工，将虚拟机器人作业渗透到供电服务指挥体系各环节。

2018年，国网杭州供电公司将人工智能技术引入电网指挥领域，以电网知识图谱、多轮人机对话引擎和业务决策引擎为核心，创新提出了“电力大脑”概念，并成功打造了第一个数字员工——虚拟调度员“帕奇”。此后，他们以虚拟调度员“帕奇”为基础，又开发了智能监屏员、虚拟调试员、办公助手、智慧客服、抢修专家等一系列人工智能“新成员”。

“虚拟数字员工团队让我们的工作效率得到有效提升。”国网杭州供电公司供电指挥中心相关负责人姜建表示，人工智能技术让虚拟员工默契配合，协同实现供电服务指挥中心6大类27小类业务应用。

据介绍，智能监屏员可自动过滤筛选电网运行告警信息，联动虚拟调度员“帕奇”和抢修专家，自动推送异常信息，触发处置流程，故障研判和故障信息报送时间由以往的5分钟缩短至1分钟。虚拟办公助手可实现自动派单、工单辅助审核、客户简报自动生成等，平均每日完成7份报表。

杭州亚运会期间，这支电力“AI军团”将优化协同作战机制，通过数字化手段实现配电网设备实时监控、抢修现场快速指挥，保障电力安全稳定供应和高效优质服务。

加速实现智能计算“中国定义”

◎本报记者 江 耘
实习生 卢馨怡 通讯员 肖 乐

6月30日，全国智能计算标准化工作组（以下简称工作组）成立。中国工程院院士孙凝晖任主任委员，秘书处由之江实验室承担，浙江省市场监督管理局负责日常管理，国家标准化管理委员会负责业务指导。

同日，工作组编制的全国首个《智能计算标准化白皮书》发布，通过对国内外智能计算标准化工作现状的全面梳理和深入分析，确立了智能计算的标准体系框架。

“全国智能计算标准化工作组的正式成立，填补了智能计算领域标准化工作的空白，为实现智能计算‘中国定义’迈出了坚实的一步。”之江实验室主任朱世强表示。

我国智算标准工作刚刚起步

算力资源是数字经济发展的重要底座。随着数字经济蓬勃发展，数字化新事物、新业态、新模式推动应用场景向多元化发展，算力规模也在不断扩大，算力需求持续攀升。工信部日前发布的数据显示，2022年，全国在用数据中心机架总规模超过650万标准机架；近5年，算力总规模年均增速超过25%。

智能计算将人工智能技术与算力相结合，能够更好地满足实际应用场景的复杂计算需求。如何推动我国智能计算产业发展，已成为政产学研各界高度关注的问题。其中，标准制定工作尤为重要。

国内智能计算标准制定工作尚处于起步阶段，标准化工作对智能计算及相关产业的发展具有基础性、支撑性、引领性作用，是推动智能计算产业创新发展的关键抓手。

当前，我国智能计算领域相关产品和服务不断丰富，已出现标准化程度不足的问题。智能计算领域企业和平台众多，虽然有些企业和平台已经具备了一定的标准化基础，但是这些分散的标准化工作并不足以完全支撑整个智能计算产业的持续发展。

此外，智能计算属于新兴领域，产业发展方兴未艾。在国际标准领域，已知的三大国际标准化组织和国际先进的标准化机构中还没有智能计算领域的专业技术委员会和工作组开展标准制订工作，国内智能计算尚未形成完善、统一的标准体系，因此迫切需要把握机遇，加快对智能计算产业发展进行研判，系统梳理、加快建立智能计算各领域的标准体系，明确标准之间的依存性与制约关系，建立统一完善的标准体系，以标准化的手段促进我国智能计算技术、产业蓬勃发展。

国内智能计算标准制定工作尚处于起步阶段，标准化工作对智能计算及相关产业的发展具有基础性、支撑性、引领性作用，是推动智能计算产业创新发展的关键抓手。

3年时间打造一个标准体系

“在智能计算这个新兴领域，我们既面临着国内标准体系尚未成型的短板，又面临着国际同行的激烈竞争。”朱世强说，在智能计算领域开展标准体系建设，是智能计算技术和产业发展的迫切需求。

之江实验室以智能计算为主攻方向，始终坚持“科研与标准双轮并进”，积极打造智能计算领域国家级重大标准化平台。

“作为工作组的秘书处单位，之江实验室具备有力支撑标准化工作的技术实力。我们希望通过3年左右的时间，构建结构合理、层次分明、科学适用、符合智能计算产业发展需要的智能计算标准体系，将技术积累转化为标准优势。”之江实验室智能科技标准化研究中心主任、工作组秘书长潘洋说。

据悉，工作组将全面梳理智能计算产业标准化需求，制订并持续优化智能计算标准体系。围绕“基础通用”“计算技术”“计算架构”“应用”“计算安全”5个方面开展标准制修订工作，联动产业平台和创新载体，结合智能计算技术发展趋势和行业应用需求，在存算一体、图计算、类脑计算、光电计算、超算互联网、科学计算等关键技术领域加快标准研制工作，持续开展智能计算标准需求征集，激发智能计算行业创新活力。

图说智能

流水线上的智能“新伙伴”

近年来，中国一汽从创新端入手，以科技赋能，着力提升生产效率，一批高智能化、高自动化的新型工厂顺利投产。在智能化的生产车间内，无数智能机器人有条不紊地按照行程轨迹进行工作，冲压、焊装、涂装、总装……各生产工序中都有智能机器人的身影。图为位于吉林长春的一汽解放J7智能工厂内，智能机器人进行车辆玻璃装配及涂胶工作。

新华社记者 许畅摄