“数字时代安全科技价值”专题研讨会举行

AI技术为安全治理提供“新工具箱”

“数字化进入新阶段,我们面临着全新的安全挑战。”近日,中国社会科学院大学数字中国研究院举办“数字时代安全科技价值”专题研讨会,中国社会科学院大学数字中国研究院执行院长吕鹏在会上说,“安全风险呈现出快迭代、高智能、全覆盖的新特点。尤其是有了生成式人工智能以后,关注人工智能的安全风险更加紧迫。”

研讨会上发布了《数字时代安全科技价值报告》。该报告指出,未来,安全科技将成为公共品,与人工智能(以下简称AI)并列成为两项通用技术。具体来看,AI作为核心关键技术,将成为未来生产力的“发动机”。安全科技将作为“方向盘”,把新兴科技控制在向善的道路上。新的安全技术发展得越好,个人与社会为数字化转型所付出的代价也将越小。

安全科技将成为公共品

“当人们提到安全科技时,想到的往往是防病毒软件和防火墙软件。但全球安全科技的版图和技术工具已远远超过这个范围。”吕鹏介绍,“在网络安全、系统安全之外,还有数据安全、终端安全、AI安全、云安全等技术门类,也包括区块链、隐私计算、量子计算等前沿技术。”

浙江工业大学网络空间安全研究院院长宣琦指出,本质上,安全科技是一种伴生技术。它永远在面向新科技、新发展。比如,伴随AI发展提出智能安全,针对生物科技提出生物安全。新技术的发展有时非常快速,所以安全技术的发展和创新也同样在高速进行。

中国社会科学院科学技术和社会研究中心研究员段伟文认为,未来,安全科技必将成为公共品。整体来看,安全科技具有“压舱石”与“助燃剂”的双重价值:守住技术的安全底线,防御外部风险隐患,让技术“难作恶”;提高技术的安全上限,降低技术运行成本,让新技术得以规模化落地,让产业在安全的基础上“跑起来”。

吕鹏举例说明安全科技在产业发展中的应用。“北京中铁建工物资有限公司和蚂蚁蚁盾共建产业风控平台,用数据智能防范上下游的协作风险,产生了较好的效果。”吕鹏说,“工作效率提高了50%以上,产业授信额度的评定科学化、可量化程度也得到大幅提高。风险预测、预警、事后风险处置等都更好更精准。”

AI安全风险主要分三类

2023年,AI大模型安全风险凸显。AI技术在带来强有力的新工具的同时,也带来数据隐私、技术滥用、失控等安全问题。“加强对AI这一新兴技术的潜在风险研判和防范,确保AI安全、可靠、可控,已成为产业发展的核心要素。”吕鹏说。

在段伟文看来,AI安全风险目前来看主要可以分为三类:内生风险、衍生风险、外生风险。

在内生风险方面,AI存在技术本身的脆弱性、对数据的依赖性等自身缺陷带来的安全问题。比如,如果给数据库不断投喂带有特定价值观的数据,会对AI系统形成严重干扰,产生“数据偏见”“观点霸权”等问题。衍生风险是指AI系统因其自身脆弱性被利用或不恰当使用,可能引发其他领域的安全问题。例如生成虚假新闻、利用深度合成伪造进行诈骗等,这涉及人身安全、隐私保护等一系列社会治理挑战。外生风险也就是面向AI系统的外部网络攻击。

《数字时代安全科技价值报告》认为,当前,安全风险变得更加复杂隐蔽、强对抗、更具破坏力,将AI驱动的业务风控系统建设得更强、更智能,更好地应对大规模网络攻击与入侵,成为行业健康发展的必需。过去几年,通过应用AI来提高安全技术的效率和成功率,已经成为技术领先企业的常态。业界开始推出“大模型质检”类安全产品,成为推进大模型安全健康发展的方式之一。

“快”“慢”结合维护大模型安全

吕鹏指出,AI技术发展给安全治理增加挑战的同时,也形成AI安全治理的“新工具箱”。

谈到“用AI对抗AI”的具体产业实践,他介绍,一方面,可以使用智能对抗技术向大模型“投射问题”,观察模型生成的回答,以此实现对AI生成图片、视频等多模内容进行“真伪”辨别和安全性检测。另一方面,通过智能化风控技术,可以帮助大模型拦截外界的恶意提问,确保外部恶意诱导无法传入大模型。同时,对生成的回答内容能够进行风险过滤,保障大模型上线后从用户输入到生成输出实现整体安全防御。

整体来看,通过从已有数据中学习,AI可以更快地识别攻击的模式和趋势,从而预测未来攻击,并配置自动响应威胁功能,在更快的时间内对抗网络威胁。

吕鹏表示,维护大模型安全既要“快”也要“慢”。大模型安全防御方面要“快”,要能快速检测、查杀病毒,确保服务无毒害。大模型安全可信方面要“慢”,要能长远地、体系化地保证整个系统环境的可控、可信。 (崔爽)

巡检充电桩 保障绿色出行

2月19日,在滁州市全椒县G40沪陕高速全椒服务区内,国网全椒县供电公司工作人员在对辖区内充电桩等供电设施开展全面巡视检查维护。 宋卫星 杨东升 摄

芜湖智能网联无人驾驶驶入“快车道”

今年以来,随着芜湖市鸠江区启动“无人驾驶先行区”建设,一幕幕无人驾驶场景走进现实,智能网联无人驾驶在芜湖驶入“快车道”。

今年春节前夕,芜湖市发布2024年第一批(总第三批)智能网联汽车公开测试道路,此次公开的测试道路双向里程超过200公里,实现了县域范围内首次开放,并进一步向芜湖中心城区延伸,覆盖江南江北,横跨多个区域。至此,该市三批次智能网联汽车公开测试道路双向总里程达352.2公里。

早在2022年8月,芜湖市就正式出台“智能网联汽车道路测试与示范应用管理办法(试行)”,规范智能网联汽车道路测试工作,推动全市路权开放,并在鼓励开展示范应用、支持模拟商业化运营、推动测试牌照互认及建立通用检测项目制度等方面推出了一系列创新举措,助力该市打造智能网联汽车国家先进制造业集群。

近年来,芜湖市将新能源和智能网联汽车列为“首位产业”精心培育。作为该市汽车零部件制造强区,鸠江区完善政策支持体系,优化产业政策、创新创业支持政策、激励政策,出台建设“无人驾驶先行区”实施方案,聚焦无人驾驶新领域新赛道,瞄准无人驾驶细分领域,重点耕耘、先行先试,无人驾驶产业生态初具规模,智能网联应用场景呈现多元化发展态势。

在城市环卫方面,酷哇科技研发投放41台无人驾驶环卫车,在芜湖市率先实现智慧化环卫作业,减少人力投入25.9%,降低运营成本5%以上;在智慧物流方面,京东物流(绿色城配)推出“无接触式配送”,推动芜湖市无人配送车投放工作进入国内第一梯队,在全省率先进入“无人配送”时代;无人运输方面,灵动科技推出无人叉车,实现仓储物流、汽车制造等行业全业务场景搬运需求。海星智驾在芜湖港朱家桥建设的集装箱无人智能堆场项目正式启用,北京踏歌智行及艾尔动力矿山自动驾驶前装基地和新能源矿卡改装基地项目落户。在智慧乘用车方面,奇瑞集团旗下大卓智能、奇瑞商用车获颁芜湖市首批智能网联汽车公开道路测试牌照,测试里程达5000公里。

作为国家邮政局授予的全国唯一快递科技创新试验基地,短短数年时间,芜湖市南陵县已发展成为全国规模最大、创新最强、增速最快、链条最全的快递物流装备产业集聚基地,正致力于加快低速无人驾驶商业化落地步伐。该县规划了16个无人化示范应用场景,并配套出台了一系列扶持政策,支持相关产业发展,让更多无人化新技术、新产品在南陵率先应用、推广和迭代,打造“全域无人化应用示范城市”。

目前,芜湖市正加快“人—车—路—云”高度协同的智能基础设施建设,推进智能网联汽车创新发展,拓展无人驾驶商业化运营应用“全场景落地”,已创建无人驾驶技术研发创新平台15个,企业拥有无人驾驶相关专利233项。 (沈宫石 阮孟玥)

铁路机车打温监控系统代替人工实现远程巡检

“原来担心机车部件在寒冷的天气中被冻坏,要不停巡视去打温。自从监控系统上线,需要打温时系统会有报警提示,大大降低了我们的作业强度。”中国铁路哈尔滨局集团公司三棵树机务段司机米刘兵说。近日,笔者获悉,针对高寒地区研发的动态物联网机车打温监控系统近日首次应用于中国铁路。该系统将人工打温变成智能监控提示打温,大大降低了人工作业强度,提高了打温效率。

中国铁路哈尔滨局集团公司三棵树机务段党委书记王宇介绍,机车在寒冷气温下停留时易发生冻结,因此工作人员要执行间隔时长不等的打温作业。打温作业可以让机车原地不动时发动柴油机,为机车提供足够热量。机车处于打温状态时,需要工作人员24小时“陪护”。黑龙江省冬季寒冷漫长,打温作业一般持续半年之久。在齐齐哈尔、加格达奇、塔河、海拉尔等纬度更高地区,打温时间则会更长。

针对以上问题,中国铁路哈尔滨局集团公司研发出了动态物联网机车打温监控系统。系统由车载数据采集、地面数据收发、监控分析三部分构成,可实现机车股道、水温、油温、柴油机转速、蓄电池电压、冷却水温度等多项重要部件数据自动实时采集。

“以往判断是否需要打温作业,需要工作人员登上机车查看操纵台的数据。现在有了这套系统,加装在机车上的传感器会代替工作人员精准‘把脉’,打温司机可实现远程巡检。”王宇告诉笔者。目前该系统已在23台内燃机车启用,并将逐步推广。

该系统还兼具火灾、转速异常、温度异常等14种报警及机车故障提示功能,并可实现精准估算打温油耗,杜绝“应间歇不间歇”“应停机不停机”过度打温行为,最大程度杜绝浪费资源和环境污染。

(李丽云 朱虹 李敏)