让人工智能更安全地赋能产业发展——

打造一个负责任的AI

本报记者 林晓晖

乌镇网事

如果世界互联网大会乌镇峰会有个“热搜榜”,“人工智能”定是火了10年的最热词之一。

从每个论坛绕不开的顶流ChatGPT到场馆里巡检的智能机器狗,今年的乌镇峰会,AI还在升温。

过去一年,人工智能在技术、产品、应用等各个层面,都以“周”为迭代速度向前突进,以超出人类想象的速度持续进化。带着想象,也有不安。人们在乌镇再次讨论起AI,它和人类的关系究竟是什么?我们应该如何与之相处?在交流碰撞的火花中,我们或许能获得一些启发。

重塑

“人工智能已经不再是一个简单的流行词,而是正在重塑我们所生活的世界。”在人工智能赋能产业发展论坛中,阿拉伯信息通信技术组织秘书长本·阿莫说。

1950年图灵最早提出了人工智能的概念,他在论文中直截了当地提问,“机器是否可能具有人类智能?”开创了人工智能领域的先河。今天,生成式人工智能不断取得突破,展现出强大的生成创造能力,开始涌现出“智慧”,而这是一种重构性的变化。

人们显然发现人工智能更聪明了。就像聊天机器人ChatGPT,它能理解我们的问题和指令,能够结合上下文的语境给出回答,并且能够处理文字、图片等多种信息,现在它不仅能在搜集、处理资料的时候帮上忙,甚至能写出各类像模像样的文学作品。在这届乌镇峰会上,2023“直通乌镇”全球互联网大赛人工智能(大模型及数字人)专题赛决赛上,京东言犀多模态大模型、达观曹植大模型、星蝠科技大模型等10个项目同台竞技。极高量级的算力、庞大的数据资源让以大模型为代表的AI技术飞速发展,令人惊叹。

同时,这些人工智能已经注入经济社会的各个支脉。在“互联网之光”博览会的未来生活数字体验区,观众可感受智能驾驶的便利、建立自己的3D数字人分身、体验人工智能在智慧教育中的运用……生活的各个场景都与人工智能相关。

现阶段,人工智能技术正处在量变引发质变的关键节点,从流行到重塑,开始渗透千行百业。360集团创始人周鸿祎也谈到,GPT大模型的诞生代表着通用人工智能、强人工智能的到来,是真正的智能涌现。人工智能大模型已经从感知进化到了认知,能够理解文字、语言,作出分析、规划,传统算法将被代替。未来在自动驾驶、机器人控制等很多重要领域,大语言模型都会大显身手。

人类是AI的塑造者。站在新技术降临的黎明,我们真切感受到数字世界和现实世界的深刻重构,也开始思考技术的更多侧面。

追问

在人工智能的交互界面上输入问题或需求,几秒内,我们就能获得丰富的信息回应,许多人停留于赞美效率,享受这样“被解放”的体验。

但是,与会的互联网代表都在追问更多——你是否想过这些数据的来源?谁投喂了这些AI,生成的数据是否真的客观、中立且真实?

我们应该留意数据本身的质量。“人工智能以大量数据为基础,但数据的生命周期涉及数据采集、传输、计算、存储、利用、消亡等方方面面。如果数据安全、可信没有做好,那么基于数据的人工智能就会面临各种各样的问题。如果使用错误的数据去训练人工智能,一定会产生错误导向。”伏羲智库创始人、中科院计算所研究员、清华大学互联网治理研究中心主任李晓东反复提醒,如今人类每年产生大量的数据,各种各样的数据充斥在一起良莠不齐,所以人工智能本身的安全和治理非常重要。

11月9日,人工智能赋能产业发展论坛上,与会嘉宾讨论的不仅仅是追赶风口,赋能产业应用,聊得更多的是安全和治理。机遇和风险是技术之刃的两面,隐私风险、数据偏见、伦理失范……人工智能带来的挑战不可避免,那我们又能做些什么?

“我们要构建全生命周期的安全保障。”思科大中华区副总裁、安全事业部总经理卜宪录在讨论中强调。如何保证数据采集合法,如何保证用户隐私,如何保证数据可信,如何保证可信数据都能被合理利用……以上种种,都涉及整个数据生命周期的安全与可信。作为全球领先的网络解决方案供应商代表,卜宪录分享,思科已经围绕透明度、公平性、问责制、隐私、安全性和可靠性展开风险的排查和控制,作为责任主体的企业更应该参与制定相应的AI框架。

“为AI制定框架是兜住底线,我们也应该反观自身,是否有足够的能力和素养理解并规范它们。”中国教育和科研计算机网(CERNET)网络中心副主任李星认为,发明技术的智慧,把握技术发展的智慧我们应该同时具备。他多次提及人工智能时代的教育和素养问题,“现在很多人都是‘人工智能时代的移民’,我们应当培养更多的年轻人成为人工智能时代的‘原住民’,具备更高的知识素养,能够正确地看待它、理性地使用它。”

技术的发展基于伦理、安全,来到乌镇峰会人们对此不断追问。安恒信息首席技术官刘博深有同感:“在10年前,网络安全产业相对没有那么受大家关注,整个行业的体量还是比较小的。这10年是网络安全产业蓬勃发展的10年。”刘博觉得,乌镇峰会的举办,让人们关注到更多的网络安全问题,这对于他所在的行业而言同样意义巨大。

希冀

“打造一个负责任的AI”——人们对这一具备颠覆性力量的技术满怀希冀。

智能治理问题、伦理问题、安全问题……提及治理问题,与会代表们不约而同谈到了“全球视野”——并不是一方就能解决所有问题,让AI“负责”需要各个国家与不同利益主体之间,构筑通力合作的机制。

我们比任何时候都需要这种共识,这种共同认可的治理框架和准则。

同样在人工智能赋能产业发展论坛上,来自企业、研究机构、国际组织等各方的代表,共同发布了《发展负责任的生成式人工智能研究报告及共识文件》。这套56页的文件,是此次发布的标志性成果之一。

文件里分析了生成式人工智能技术发展态势,对经济社会、科学研究带来的机遇,以及在安全与伦理等方面的挑战,全面梳理了全球主要国家、地区、组织以及产业界面对生成式人工智能发展与治理的努力,汇总了产业界在金融、交通、教育等14个应用场景的相关案例,提出发展负责任的生成式人工智能十条共识。

“许多人工智能领域的应用与社会、文化息息相关。比如生成式人工智能生成的内容基于不同国家和文化,生成的内容服务于不同文化的人民。然而许多人工智能产品都面向全球提供服务,这个时候,必须遵守不同的监管框架。因此,不同区域监管框架之间的接口与互操作性对全球化至关重要。”中国科学院自动化研究所人工智能伦理与治理研究中心主任曾毅多次呼吁创建全球人工智能治理网络。如今,共识形成,各方携手,朝着推进构建生成式人工智能治理模式的方向,迈出了坚实的一步。

技术与人性,在这里同时闪耀。

乌镇talk

让人工智能助孩子成长

随着新技术新媒介的普及,人工智能成为青少年必须掌握的新工具之一。因此,建设一个儿童友好型人工智能社会,让人工智能成为推动青少年健康成长的新工具,就显得尤为重要。

前不久,国务院发布了《未成年人网络保护条例》。未来,我们要坚持保护和发展并重,在“儿童为中心”“保护儿童权利”和“多方共治”“技术向善”理念的指导下,坚持最有利于未成年人的原则,兼顾未成年人网络安全和数字发展,注重人工智能技术应用符合科技伦理,鼓励互联网企业积极参与行业共治,促使人工智能成为促进未成年人发展的新动能。

未来,我们还应该构建人工智能素养培育生态系统。加强义务教育阶段“信息科技”等课程内容、教学方式方法的创新,以中小学基础教育为主阵地,形成政府主导,学界和企业共同建设的人工智能公共服务平台,推进人工智能素养教育及创新实践活动,推动优质教育资源均衡共享,让人工智能更好为青少年的成长成才服务。

——中国工程院院士、新疆大学教授吾守尔·斯拉木在“未成年人网络保护论坛”上的发言

(本报记者 宋彬彬 整理/摄)

人工智能需要正确“打开”

在与ChatGPT对话时,人工智能常说“我认为(I think)”“我建议(I suppose)”。但此“我”非本“我”。能力再强,ChatGPT也只是个信息处理工具——缺乏理解力,或容易以人们难以预期的方式犯错,更不必说具备道德感、责任感。

因此,为降低潜在风险的伤害,首先,不应且不能混淆人工智能与人类的界限。其次,不应过多依赖,且不能抱有过高期待。不可否认,今年人工智能的突破,给各行各业的发展带去了无限可能。但眼下,类人工智能生成技术能替代的,只是工业化的、流水线上出来的东西,任由它在生活中无处不在是不合适的。“一半规划,一半野生”,全球协作、安全评估后适度使用,才能实现人工智能与人类的和谐共生,共同迈向一个更加智能和可持续的未来。

——联合国高级别人工智能咨询机构专家组成员、中国科学院自动化研究所研究员曾毅在“全球发展倡议数字合作论坛”的发言

(本报记者 谢丹颖 整理 本报记者 俞碧寅 摄)

以科技创新促安全发展

网络安全防护正面临三方面问题,一是下一代互联网规模扩大导致部分安全威胁“看不见”,二是互联网体系结构复杂多样导致安全威胁“理不清”,三是新兴应用场景导致安全威胁“防不住”。因此,我们建议,要以技术创新推动互联网安全治理。面对互联网技术发展带来的安全挑战,需要通过持续的技术创新来解决“看不见、理不清、防不住”问题,保护互联网安全。

安全是下一代互联网发展的基础和保障,互联网安全治理是一个复杂的系统性工程,需要各方共同努力、深化协同治理,综合运用政策、技术、标准等多种手段,构建统筹协同的网络安全防护体系,不断提升网络安全综合防护能力。同时协同产学研各方资源和能力,进行网络安全技术研发和创新攻关,以科技创新促安全发展,共同构建更加安全、健壮的下一代互联网。

——奇安信科技集团股份有限公司总裁吴云坤在“下一代互联网创新发展论坛”上的发言

(本报记者 宋彬彬 整理/摄)

好演员要提升数字技能

数字技术和艺术的关系,其实是相互依存、相互促进的。数字时代,影视制作的效率大大提高,尤其是拍摄过程中不再需要等雨来、等风来,可以通过数字摄影、特效创作、虚拟现实等技术实现,不再受限于时间与空间。只要创作者想得到,观众就能看得到。

数字技术给影视产业带来新的机遇和挑战,对演员的素质和演技要求也将会更高。在三维虚拟场景拍摄中,通过计算机的实时渲染,导演可以在监视器中看到合成后的画面,会有更多试错和打磨的机会,而演员需要在没有实物的空间中表演。在数字时代,想当一名好演员,不仅要有极强的艺术感知力,还要提升对技术的掌握水平。

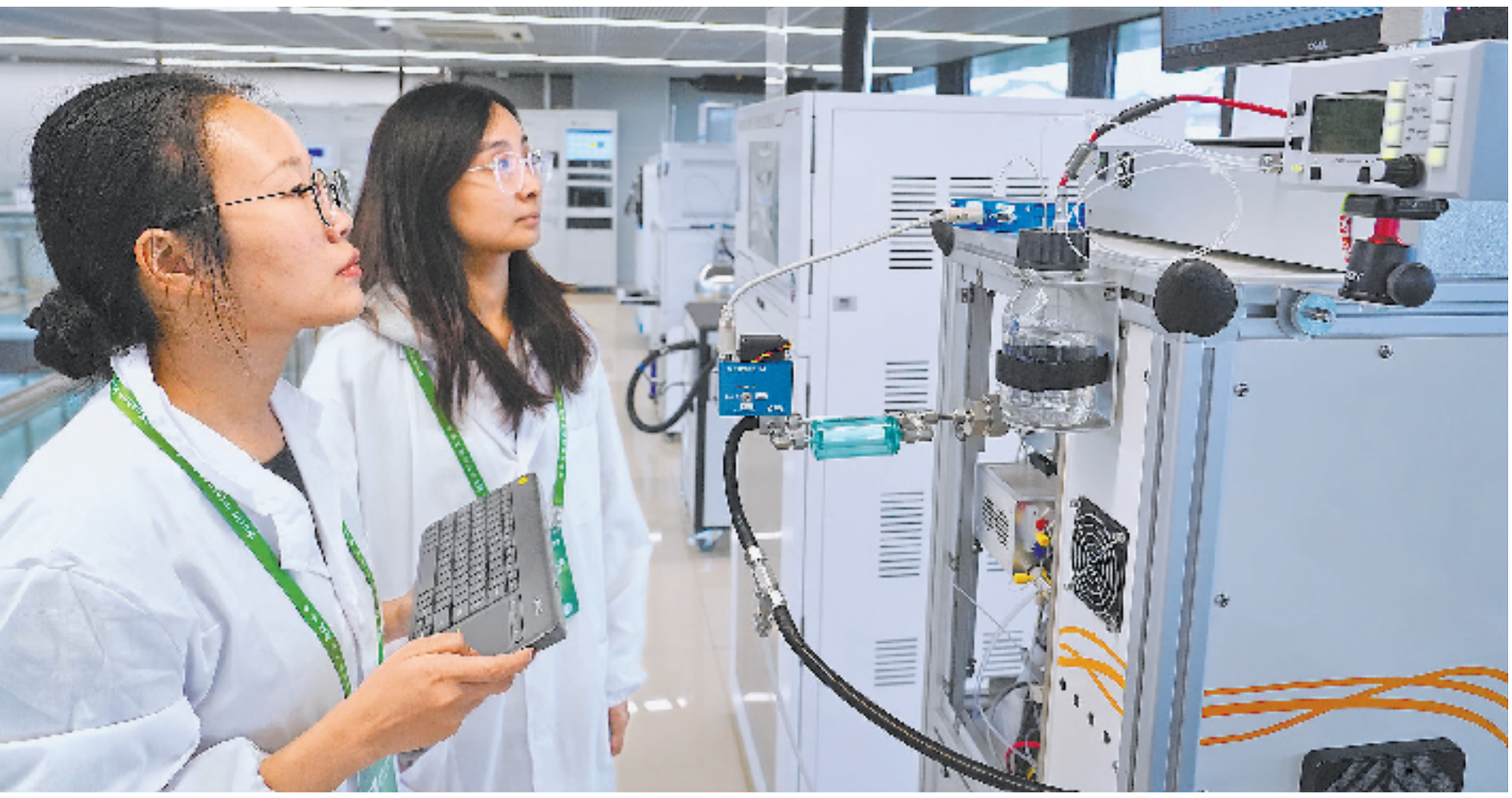
——演员雷佳音在“数字素养与技能提升论坛”上的发言

(本报记者 张源 整理 本报记者 王建龙 摄)

乌镇治气有利器

乌镇大气监测超级站近日启用,该站能进行大气复合污染过程、关键因子、形成机制等研究型监测,并及时、精准地发现问题,实现靶向管控。图为工作人员检查电喷雾化学电离质谱仪。

本报记者 钱璐斌 徐文迪 摄



论坛嘉宾共议风险应对与技术赋能前景——

人工智能潮涌,应兼顾创新与安全

乌镇时间

本报乌镇11月9日电(记者 杨一凡)以大模型为代表的人工智能技术快速进步,引领科技革命和产业变革浪潮。而安全与发展犹如鸟之两翼,缺一不可,确保网络安全成为全球的一项共同责任,如何兼顾技术创新和安全发展成为重要话题。

11月9日,2023年世界互联网大会乌镇峰会前沿数字技术创新与安全论坛举行,与会嘉宾就前沿数字技术发展和应用探索,共议安全风险应对与技术赋能前景。

大模型发展推动了人类数字化进程加速,业内对其风险的关注度也在提升。

中关村实验室主任助理、首席科学家卿昱通过对2019年到2023年顶级会议论文的热词分析,揭示了这种变化:这5年来,研究者从对隐私计算、恶意软件、数字取证、异常检测技术的关注,扩展到了人工智能、深度学习和隐私保护相关技术。她表示,各种新技术的出现同时也带来了新的安全威胁。

伦理风险、算法风险、数据风险、技术滥用,在清华大学人工智能研究院副院长、计算机系教授朱军看来,生成式人工智能的新型安全风险正在显现,这些新型风险还可能与传统网络风险形成叠加效应。

在论坛上,嘉宾们分享了诸多有效治理工具,以解决安全问题、规范技术发展。

奇安信集团总裁吴云坤介绍,为应对网络安全攻防难题,奇安信基于安全能力、知识和实践打造了工业级安全大模型,将其应用于网络安全攻防场景,增强安全能力,可以提升生产力和生产效率。朱军演示了人工智能安全评测平台、人脸人工智能安全防火墙、数据安全产品及人工智能内容检测平台等针对性解决人工智能发展问题的工具。

互联网始终担当连接世界的桥梁,兼顾创新与安全的同时,让数字技术为每个人带来机遇、赋能行业,打造智能互联新世界,成为与会嘉宾的共识。

在论坛上,记者看到国内外企业展示技术普惠与技术赋能的成果。

高通公司全球高级副总裁钱堃表示,5G和AI的协同发展,渗透到大众生活中的智能手机、移动PC和联网汽车等终端,在智慧城市、医疗系统、资源优化等领域也发挥着越来越重要的作用,助力数字世界的可持续发展。

文心大模型4.0全面开放两个多月,用户达到7000万,场景达4300个,联合研制10余个行业大模型。百度首席技术官王海峰表示,大语言模型的产业落地,正助力产业智能化升级。阿里云“通义千问”大模型已经向全社会开放,并且启动了“千问伙伴”计划,阿里云智能集团首席风险官兼首席财务官郑俊芳表示将与行业伙伴共同创新生态。沟通、普惠、提效,腾讯集团副总裁蒋杰用3个关键词介绍了已经落地应用的AI工业质检、大模型辅助设计、服务听障群体、支持生态保育诸多项目。

网络空间已成为人类发展的新疆域,也是网络安全前沿技术发展和运用的“秀场”。人工智能潮涌,围绕网络空间安全新形势和需求,推动安全发展,走向负责任的人工智能新时代,赋能行业普惠美好生活——这正是论坛展现的前景。