人民日报海外版

2024年6月10日 星期一

责编：卢泽华 邮箱：wesley1111@163.com

“融”观中国

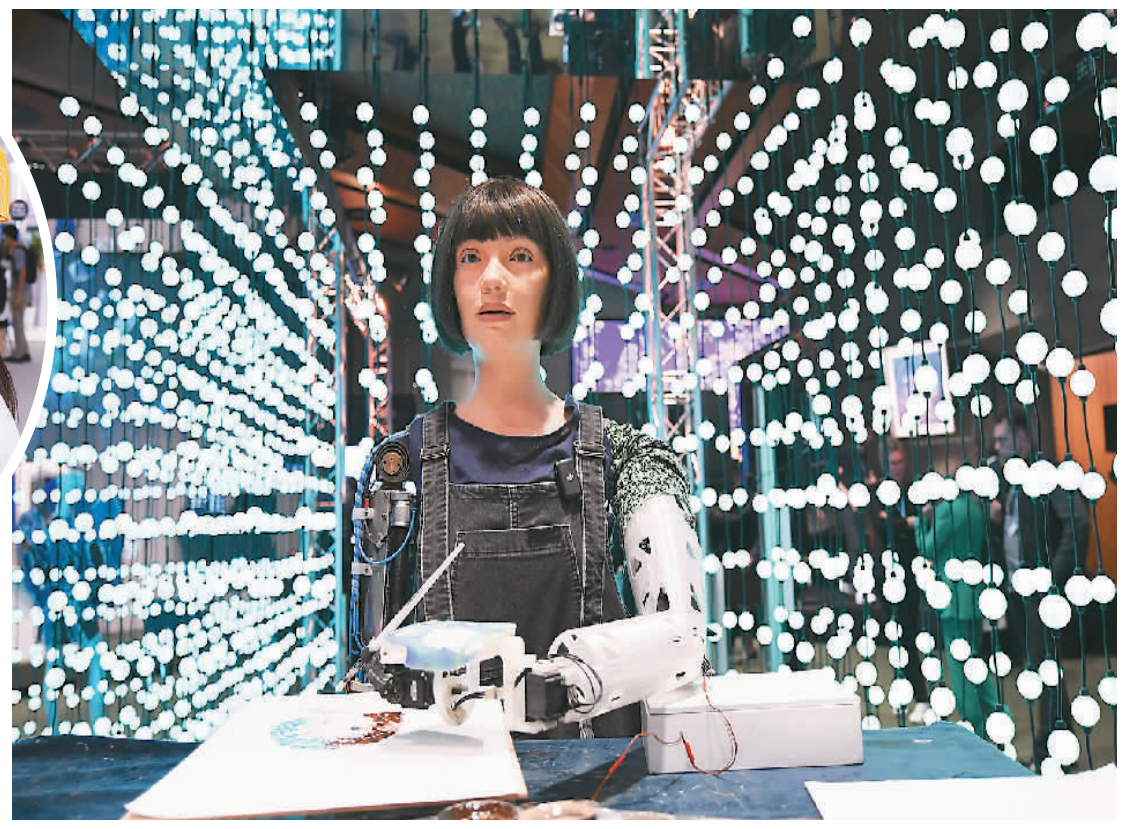
AI发展快，系紧“安全带”

——“人工智能与信息保护”系列报道之二

韩维正 毛翊兴

AI“换脸”实施诈骗、AI伪造图片传播谣言、AI“复活”逝者引发争议……随着人工智能技术的不断发展，AI深度伪造的例子屡见不鲜。不少网民惊呼，以前总说“有图有真相”，如今我们连图片和视频也不敢轻易相信了。

AI深度伪造是怎么回事？该如何防范相关风险？用户、平台、监管、司法等社会各界该如何形成合力，建立生成式AI信息安全规范？我们对此进行了采访。



▲ 在北京举行的全球数字经济大会上，参观者体验智能安全帽。陈晓根摄（人民视觉）

▶ 在瑞士日内瓦举行的“2024人工智能造福人类全球峰会”聚焦人工智能监管与治理等议题，现场展览的绘画机器人吸引众多参会者目光。新华社记者 连 漪摄

“眼见不一定为实”

关于AI深度伪造的热点事件，正引发全世界广泛关注。

今年1月，美国一位流行歌手被人用AI恶意生成虚假照片，在社交媒体迅速传播，给歌手本人造成困扰。2月，香港一家公司遭遇“AI变脸”诈骗，损失高达2亿元港币。据悉，这家公司一名员工在视频会议中被首席财务官要求转账。然而，会议中的这位“领导”和其他员工，实际都是深度伪造的AI影像。诈骗者通过公开渠道获取的资料，合成了首席财务官的形象和声音，并制作出多人参与视频会议的虚假场景。

基于深度合成技术引发的侵权案例，常见的手法是冒充熟人实施电信诈骗。不久前，江苏句容的杨女士，在收到自己“女儿”多条要求缴纳报名费的语音后，向骗子账户转账3.5万元。相关办案民警反复提醒：“遇到转账一定要慎之又慎，眼见不一定为实。”

“深度伪造技术利用AI深度学习功能，实现图像、声音、视频的篡改、伪造和自动生成，产生以假乱真的效果。”上海人工智能研究院院长宋海涛给公众应对深度伪造支了三招：一是掌握甄别AI“换脸”的简单技巧，比如要求对方在视频对话时在脸前挥挥手，看是否出现图像扰动等；二是学会使用检测深度伪造的工具和软件；三是保持合理怀疑。“保持谨慎和警惕，是公众应对AI造假的第一道防线。”宋海涛说。

“用技术治理技术”

如何把生成式AI的强大能力用于建立信息安全规范，将制约行业发展的“绊脚石”变为“压舱石”？业界一直在探索“用技术治理技术”的方案。

瑞莱智慧是清华大学人工智能研究院孵化的企业，专攻人工智能安全领域。瑞莱智慧总裁田天介绍，其公司研发的生成式人工智能内容检测平台，支持多种合成类型的图片、视频、音频、文本的真伪检测，应用场景包括打击网络诈骗和声誉侵害行为、检测网络内容合规性、检测音视频物证真实性等。

“利用AI技术治理AI犯罪，本身也是一个不断博弈的过程。”田天介绍，“红队测试”是目前生成式AI治理的重要手段，旨在通过模拟攻击者行为，对目标系统进行全面网络攻击，针对性地发现、修补潜在系统漏洞，使模型在面向公众开放前，充分接受安全技术检验。视频合成AI平台Sora，在上市前就曾邀请数名从事信息安全漏洞研究的专家充当红队进行对抗测试，找出相当数量的安全漏洞。

此外，业界也在推动落实AI生成内容标识制度。中国政法大学数据法治研究院教授张凌寒表示，全国信息安全标准化技术委员会在《网络安全标准实践指南——生成式人工智能服务内容标识方法》中给出了内容标识方法：通过在交互界面中添加半透明文字的方式显示水印标识，或通过人类无法直接感知但可通过技术手段从内容中提取的隐式水印标识，提示内容由人工智能生成。“标识制度可以提升AI信息内容治理能力，减少虚假信息生成，防止虚假信息污染下一代训练数据，营造良好的网络信息生态环境。”张凌寒说。

国际测试委员会创始人、中国科学院计算所研究员詹剑锋建议，应将AI深度伪造纳入监测机制，遇到负面影响较大的造假行为，第一时间快速反应。通过建立针对深度伪造有害内容的群众举报机制，提高公众的判断力、鉴别力。

“制度引导技术向善”

生成式AI技术是一把双刃剑，如何在释放创新活力的同时，有效防范信息安全风险？

不少专家表示，尽快建立健全相关治理规范至关重要。“这一领域的治理不能完全寄希望于企业自治，更需要法律硬性监管。”对外经济贸易大学数字经济与法律创新研究中心主任张欣表示，“通过立法为AI开发划定基本底线，明确合规义务，可以从源头防范风险，避免‘先污染后治理’的被动局面。”

去年国家网信办等部门发布的《生成式人工智能服务管理暂行办法》标志着生成式AI有了专门性行政法规。2023年1月施行的《互联网信息服务深度合成管理规定》明确提出，“任何组织和个人不得利用深度合成服务制作、复制、发布、传播法律、行政法规禁止的信息”“可能导致公众混淆或者误认的，应当在生成或者编辑的信息内容的合理位置、区域进行显著标识”，等等。

在法律层面，中国已经出台了《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律，但针对AI领域的司法治理仍有进一步细化的空间。“现有法律法规对数据权属、保护范围等问题暂无具体规定。在人工智能生成结果的保护方面，《民法典》的原则性条款在具体适用方面还有一定难度。”北京互联网法院综合审判三庭负责人颜君建议，要进一步加快人工智能方面的法律供给，推动国家层面生成式人工智能规范的确立。对已有案例出现的疑难法律问题，要通过学术讨论和司法实践，逐步形成共识，总结确立法律适用规则，推动《人工智能法草案》的立法进程。

在伦理层面，AI“复活”等行为引起部分用户的反感和恐惧。对此，科技部等10部门印发《科技伦理审查办法（试行）》，列出了需要开展伦理审查复核的科技活动清单，其中就包括具有社会动员能力和社会意识引导能力的算法模型、应用程序及系统研发等。

“我们要正确看待新业态带来的可能性，但也不能让其偏离合法性、合理性轨道。无论是技术开发者、使用者还是监管者，都有必要用类似的伦理视角来审视技术发展。在追求科技进步的同时，确保AI技术循着以人为本和技术向善的理念发展。”张凌寒说。

新媒视点

避免对AI的过度依赖

卢泽华

身边有个无所不知、无所不能的助手，是千百年来人类的梦想。《山海经》中就记载了一个名叫“白泽”的神兽，“能言语、通万物之情”，不仅有求必应，还能占卜吉凶。

如今，人工智能（AI）的发展，正让这个“神兽”走进千家万户。琳琅满目的智能设备不仅能轻松做到“能言语”，“通万物之情”也已基本实现。借助海量数据库，各种AI设备有如百科全书，再生僻的问题都难不倒它。“有问题找AI”已渐渐成为人们的习惯。

AI解决问题的能力有目共睹：教师可以用AI批改作业，分析学生特点，优化教学方案；律师可以用AI审核案卷、分析案例、草拟合同，还能制定辩护策略；会计可以用AI记录数据、设计账目、生成报表，过去耗时费力的财务核算，AI可以一键完成。最近，一位从事留学中介服务的朋友向笔者抱怨，五月份只接到了两单活，因为生成式AI的兴起，学生对写作入学申请的人工辅导服务需求量大减……

值得注意的是，随着AI应用的深化，“有问题找AI”的现象也引起了人们的警觉。教师过度依赖AI批改作业，对学生情况缺乏直观感受，只能相信AI数据分析。可是，学生的成长受到环境、性格、学习等因素综合影响，AI分析会不会因为偏差而“误人子弟”？律师界担忧，青年律师刚入行，就频繁使用AI撰写文书，从业底子怎么会厚实？会计行业也在呼吁，不要过度依赖AI，因为财务工作不是简单的数据分析，还需要大量的专业判断，“没有价值观的智能人”很可能引发财务监管和伦理上的风险……

如果说，各行各业的担忧还不够直观，一个小故事也许更能引起公众的重视。如今，一些中小学生开始利用AI帮助自己写作文和读后感。面对家长对孩子“学习不要偷懒”的告诫，孩子振振有词：“将来都是机器人帮我们干活，学习早就没有意义了！”

这样的回答，能不引起我们的深思吗？

在决策式人工智能时代，AI技术越来越“拟人化”。但绝大多数科学家认为，应该将AI控制在“常规替代性”领域，即替代人类从事那些可重复性高的常规任务，而社会洞察、情感表达、艺术审美等高阶任务，仍要人类来完成。

“有问题找AI”的思维惰性和依赖，很可能在不知不觉中弱化人们的管理意识、防护意识和学习意识。世界最大一家AI巨头的首席执行官表达过这样的忧虑：“AI技术无疑是人类迄今为止发展出的最伟大技术，它将重塑社会——但同时，AI也会带来风险，可我们的社会却没有太多时间去思考如何监管、如何处理这类事情。”

业界早对这种风险进行过归纳，主要有两个方面：一是安全性风险。我们只注重为AI喂料并获得计算结果，却疏于监管其过程中发生了什么，由此滋生出各类信息安全隐患。二是退化性风险。对AI的依赖，逐渐消磨了人们的学习和认知能力，由此带来人类被“取代”的隐患。这两种风险都在警示我们，要避免对AI的过度依赖。

毕竟，有问题就去找AI，一旦AI出了问题，又该去找谁呢？

▲ 在福建福州举办的网络安全博览会上，市民体验数字货币购物功能。王旺旺摄（人民视觉）

◀ 安徽省含山县姚庙中心学校为学生们开展网络安全等宣讲活动。欧宗涛摄（人民视觉）

前沿动态▶▶▶

人工智能让车辆识别行人速度提高百倍

据新华社日内瓦电 瑞士苏黎世大学近日发布公报称，该校研究人员将仿生摄像头与人工智能技术相结合开发出一套车载系统，能以比现有车载摄像头快100倍的速度识别行人和障碍物，可大大提高行车安全性。相关成果已发表在英国《自然》杂志上。

这套最新开发的系统与传统相机不同，它不是通过定期拍照捕捉画面，而是以模仿人眼感知图像的方式，在每次检测到快速运动时记录信息。

人工智能将为劳动力市场带来重大变化

据新华社柏林电（记者**褚怡**）麦肯锡全球研究院日前发布一份名为《工作的新未来：在欧洲及其他地区部署人工智能和提升技能的竞争》的报告，认为包括德国在内的多个国家劳动力市场将因人工智能出现重大变化。

报告说，随着人工智能技术的快速推广，劳动力市场将迎来重大变革。预计到2030年，生成式人工智能将帮助美国和欧洲近1/3的工作时间实现自动化。