近年来，AI深度伪造技术被犯罪分子用于电信诈骗，引发社会广泛关注。

AI深度伪造是一种利用人工智能技术合成、修改或替换图像、视频和音频内容的技术，可以将一个人的面部表情、言辞和动作应用到其他人的图像或视频上，使生成的内容看似原始内容。

因为易得性强、成本低、仿真度高，AI深度伪造技术巨大的应用潜力为艺术、社交、医疗等领域的发展带来了新的可能性。与此同时，由于该技术通过一张照片、一段语音就能生成一段伪造视频，一些不法分子利用AI深度伪造新工具实施电信诈骗、散布虚假视频、激化社会矛盾，给安全领域带来了诸多风险。

如何规制AI深度伪造技术被滥用？如何升级反制技术破除监管难点？这些问题的有效破解，不仅关乎社会治安和国家安全，也影响着新一代人工智能技术的发展走向。

图据《瞭望》新闻周刊

谨防AI沦为电诈神器

AI深伪诈骗危害性更大

近年来，一些网络犯罪分子使用“深度伪造”的文本、图像、音频或视频，进行欺诈活动。记者从公安机关采访获悉，依靠深度伪造技术工具，AI客服可以同时给上万人打电话，从事电信诈骗的危害性更强、数额更大。

——假冒熟人进行诈骗。2023年4月20日发生在内蒙古包头市的一起金额高达430万元的诈骗案件，竟是利用AI换脸技术得手的。当天，郭先生接到来自好友的求助电话，对方称自己在外地竞标，需要430万元保证金。巨大的金额也让郭先生产生了怀疑，于是拨打视频通话确认对方身份，近乎一模一样的面容与声音让郭先生消除了疑虑。短短十分钟，430万元便已被转入骗子账户。好在事后经再次打电话确认，郭先生识破骗局及时报警，300多万元受骗金额被冻结。

——假冒知名人士误导公众。在一些案件中，犯罪分子伪造视频或录音，使生成的内容看起来像是知名人士正在说他们从未说过的话、做从未做过的事，不仅给个人带来名誉损害，也对公众形成误导。

——假冒官方网站或账户发布不实信息。有的犯罪分子利用AI技术伪造各大知名企业、平台或社交媒体官方账户进行诈骗。互联网上的“V”字认证往往是官方认证的标识，可当官方认证也能被AI深度伪造技术造假，一模一样的头像主页和认证标识，会让公众无法分辨“真假美猴王”。近期，职场社交平台领英发现，其平台上有1000多个用AI生成的虚假“V”字认证账户，发送大量推销信息及钓鱼邮件，甚至形成了相关产业链。

——假冒他人身份实施诈骗。利用虚拟或合成身份，犯罪分子可以盗用或注册他人账号，实现骗取养老金、骗取人寿保险的犯罪目的，潜在风险极大。业内人士提醒，保险行业很可能成为遭遇AI深度伪造欺诈风险最高的行业。

升级反制技术“道高一丈”

面对不断升级的AI深度伪造欺诈手段，需要尽快升级技术手段，“道高一丈”实现反制破局。揭秘AI换脸、语音变声等深度伪造手段的网络安全科普，以量子加密技术保障金融、电力等基础设施安全，用大数据反诈系统守护公民人身财产安全……我国正多措并举，升级反制技术，加强网络安全防护。

随着湖北省黄石市公安局联合科大讯飞股份有限公司、电信运营商联合研发的反诈智能语音机器人“小飞”在黄石“上岗”，一些“不听劝”的受害者回过神来，避免了经济损失。“以往开展劝阻，一名民警每天拨打几十个电话，累得嗓子冒烟。换成‘小飞’之后，一天可以拨打几百乃至上千个电话，效率大大提升。”黄石市公安局科信支队大数据中心负责人李雪松说。据了解，以前，民警拨打电诈劝阻电话，一般通过公安局座机或者自己的手机，容易被当成普通来电甚至是推销电话。而“小飞”系统可以筛查出电信网络诈骗高危级潜在受害人，自动通过反电诈专用号码96110联系对方；联系劝阻形成的大数据又进一步训练“小飞”升级反电诈劝阻办法，提升劝阻精准度和成功率。

以技治网，更多与“小飞”一样的技术创新，正在破解技术滥用带来的负面问题。

华为云诈骗载体智能检测技术通过即时内容获取来捕捉APP涉诈内容，快速识别“黑灰产”；小米移动研发的“灵犬”骚扰诈骗防治系统，针对开卡入网的前、中、后环节设立防骚扰、诈骗等异常行为的触发机制；中国电信开发的“翼网”平台综合采用短信电话预警、运营商“断号”、预警劝阻等措施提升电信网络反诈能力……

深圳计算机学会秘书长、北京大学深圳研究生院深圳市内容中心网络与区块链重点实验室主任雷凯表示，可以采取“全周期沙盒”管理方式在源头上保护重要内容不被非法使用。“将那些经过严格审核、真实可信的图像、音频和视频纳入白名单，把可能被用于生成虚假内容的图像、音频和视频以及AI工具都放在‘沙盒’中，进行全周期监控和管理，通过细粒度监管，对白名单内容的发布加强内容鉴权和追溯，确保其真实性和完整性，防止被恶意篡改。”雷凯说。

依法治理亟需细化配套措施

电信网络诈骗的潜在受害人面广量大，通过法律加强对“事前（内容源头监管）—事中（诈骗快速查处）—事后（追偿、救济）”的全方位规制，源头规范深度伪造技术的使用和发展，具有较强的必要性和紧迫性。

目前，我国现行法律法规，已对深度伪造作出一定规制：2022年12月施行的反电信网络诈骗法加强预防性法律制度构建，加大对违法犯罪人员的处罚。2023年1月施行的《互联网信息服务深度合成管理规定》明确规定，“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”；“提供人脸、人声等生物识别信息编辑功能的，应当提示深度合成服务使用者依法告知被编辑的个人，并取得其单独同意”；“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。

受访专家认为，从现有司法实践来看，还需在多个方面进一步完善法律规范配套措施。比如，深度伪造技术风险最有可能涉及肖像权和名誉权侵权，但根据民法典侵权责任编的规定，在被侵权人对损害后果难以准确证明的情况下，损害赔偿数额难以确定，精神损害赔偿不易实现，被侵权人难以得到充分救济。为进一步防止不法分子利用AI深度伪造等技术实施犯罪活动，仍需细化配套措施，让法律条款的落实更加简便易行。

一方面，通过数字版权管理等方式，加强对生成内容甄别、溯源、追责。深圳光子晶体科技有限公司创始人、首席科学家王勇竞博士建议，加强深度伪造内容制作者和网络服务提供者的责任，在事前规范落实制作者的披露义务，明确标识其制作的内容为AI合成，同时强化网络服务提供者的监督管理责任。侵权行为发生后，被侵权人可以通过人格权禁令主张对深度伪造记录予以封存，简化损害赔偿的证明责任，明确损害赔偿的计算模式，以更好保护被侵权人合法权益。

另一方面，将AI深度伪造技术的潜在风险纳入应急响应机制。国际测试委员会创始人、中国科学院计算所研究员詹剑锋教授建议，将AI深度伪造纳入舆情监测机制，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，提高公众的判断能力、鉴别能力。同时，建立一定规模的志愿者群体标识数据内容，为人工智能监管平台提供样本和数据，供算法学习训练，从而提高监管效率，更好规范人工智能技术有序发展，形成良性发展的生态闭环。 （据《瞭望》新闻周刊 记者 白瑜 实习生 罗梓凝）