一张图片就能“活化”成视频

警惕AI深度合成击穿风险底线

一段视频、一段语音，未必是真人拍摄或录制，在你不知道的手机App后台、支付界面、门禁闸机，或许有人正在盗刷你的脸。随着人工智能(AI)深度合成技术日益精湛，合成的音频、视频等伪造内容越来越能以假乱真。毫无疑问，我们生活的现实世界正在面临技术滥用的风险与挑战。

资料图片

盗刷人脸篡改声音

近两年来，在浙江、安徽、江苏等地，多名盗取个人信息的犯罪嫌疑人被公安部门抓获。犯罪嫌疑人作案流程极为雷同：先是非法获取他人照片或有偿收购他人声音等“物料”，然后利用人工智能技术将照片“活化”、合成动态视频，之后或直接骗过社交平台、支付宝账户的人脸核验机制，进行非法获利；或骗过手机卡注册过程中的人工审核环节，继而利用他人名下的手机号进行电信网络诈骗、网络赌博等，使被收集信息的人遭受安全威胁和财产损失。

记者在清华大学人工智能研究院实验室的演示电脑前看到，一张刚从微信朋友圈中下载的陌生人的正脸静态照片导入电脑后，在技术人员的操作下，照片上的人物可瞬间“活”起来，根据指令做出相应的眨眼、张嘴、皱眉等精细动作和表情变化，并在短短十几秒内生成流畅视频。

“完成由静到动这一驱动操作的技术叫深度合成技术，是人工智能内容合成技术的一种。”清华大学人工智能研究院工程师萧子豪说，深度合成技术已经衍生出包括图像合成、视频合成、声音合成、文本生成等多种技术。

在技术加持下，盗刷人脸不再是难事。在手机卡注册、银行卡申请、支付软件登录等需要人脸动态识别的环节，这些伪造的合成视频可协助不法分子通过后台审核验证。

技术人员向记者演示了声音合成的操作。几段60秒的陌生人语音通过深度合成技术，即可生成“不用打卡，直接微信转账给我吧”“今天你不用去接孩子了。我就在学校附近，顺路去接孩子”等语音，效果如同真人发出的声音。这种声音合成令人“细思极恐”。

深度合成正瓦解“眼见为实”

在国内外内容平台、社交平台上，深度合成内容呈现“量质齐升”。其中合成的影视剧片段、话题人物的换脸视频等因具有较强娱乐性而被大量传播。

清华大学人工智能研究院、北京瑞莱智慧科技有限公司、清华大学智媒研究中心、国家工业信息安全发展研究中心、北京市大数据中心联合发布的《深度合成十大趋势报告(2022)》显示，2017年至2021年国内外主流音视频网站、社交媒体平台上，深度合成视频数量的年均增长率超过77.8%。2021年新发布的深度合成视频数量是2017年的11倍。与此同时，深度合成内容的曝光度、关注度、传播力也呈指数级增长，2021年新发布深度合成视频的点赞数已超3亿次。

“网上流传的视频、语音，未必是真人拍摄或录制。”浙江大学网络空间安全学院院长任奎说，是全脸合成、音频合成，还是真实拍摄录制，许多时候凭借人眼难以分辨。

清华大学计算机系教授、人工智能研究院基础理论研究中心主任朱军认为，深度合成技术正在改变信息传播内容信任链的底层逻辑和复杂程度，风险隐患在迅速加大。一方面，“眼见为实”的定义发生改变。尽管公众对照片等静态信息易被篡改已有认知，但对视频、声音等动态信息仍持有较高信任度，深度合成技术再次瓦解了“眼见为实”的信任逻辑。二是短视频的广泛传播，使深度合成技术的滥用产生了较大范围的影响力和破坏力。

清华大学苏世民书院院长、教授薛澜认为，当深度合成等人工智能技术走向“滥用”，就会带来一系列的伦理和治理问题：轻则侵犯个人财产安全、伤害个人尊严和隐私，重则威胁国家安全、影响社会稳定。

技术是一把双刃剑。用好这把双刃剑，既不能让技术成为脱缰的野马，也不能让技术创新原地踏步。

从善用技术的角度，中国工程院院士、信息技术专家邬贺铨提出，对于技术的新应用、新发展，不能“一刀切式”地禁止和干预，以免阻碍其创新。而应当从源头上解决技术衍生的安全问题，利用技术创新、技术对抗等方式，持续提升和迭代检测技术的能力。

朱军认为，当前针对深度合成应用的检测技术仍处于探索阶段，手段尚不成熟。建议充分发挥科研院所、科技企业等力量，尽快形成有效、高效的深度合成应用技术检测能力，以在舆论战、信息战中争取技术优势。

从风险治理的角度，国家工业信息安全发展研究中心副总工程师邱惠君指出，近年来的数字化转型倒逼多国人工智能安全风险治理落地。欧盟率先在人工智能领域开展了立法，基于风险分析的方法，重点明确针对高风险人工智能系统的监管框架。

“人工智能安全包括数据安全、框架安全、算法安全、模型安全、运营安全等组成部分。对此，我们应当构建‘规定+标准+法律’的一体化治理规则体系，出台风险治理的指南、标准、评估规范，在条件具备时完善立法。”邱惠君建议，重点围绕数据、算法、模型和运维的角度，一是构建数据采集质量规范；二是根据应用场景对人工智能进行系统风险分级分类；三是建立安全责任体系，明确设计开发单位、运维单位、数据提供方的各自责任。

中伦律师事务所合伙人陈际红表示，打击“变脸”诈骗犯罪，应从技术的合法使用边界、技术的安全评估程序、滥用技术的法律规制等方面予以规范，提高技术滥用的违法成本。

朱军提示，公众应当对深度合成新技术、新应用形成正确认知，对其不良应用提高防范意识，保护好个人声纹、照片等信息，不轻易提供人脸、指纹、虹膜等个人生物信息给他人。

真的假不了，假的真不了

尽管有刑侦专家指出，当前的AI造假技术尚能通过加强辨识来防范，但随着技术快速迭代，人类迟早有一天会无法通过自身鉴别真假。“眼见为实”的基本常识，已成明日黄花，我们究竟还能不能在高科技发展的未来，维持互相信任？难道我们向往科技昌明的明天，会是“高科技、低生活”的“赛博朋克”世界？

科技进步的脚步无法也不应被人阻滞。但科技带来的副作用，却需要我们运用智慧来规避。正如当下流行的网络语言“只有用魔法才能打败魔法”，研发更多监管性“白科技”，将科技创新约束在可控范围，势在必行。

人工智能其实已经奋斗在了反诈骗一线。现在，当你接到疑似电信诈骗电话时，这通电话很可能已经被公安部门甄别。电信诈骗总是有类似特征，通过大数据分析，很容易就能从正常通讯中发现异常。许多人接到疑似诈骗电话后，第一时间就会接到公安部门反诈骗中心的提示电话或信息。这种人工智能“道高一丈”，可以很好地从根上铲断诈骗黑手。当然，如果这一道防线被突破，在人们进行转账时，人工智能卫士仍会守住防线——陌生人之间没有征兆的大金额转账，往往意味着危险，银行、支付应用系统已经“标配”了这样的预警和干预机制，让人就算被骗也不会蒙受损失。

诈骗其实是一个“产业链”，在这个链条上有诈骗者、洗钱者，也有收卡者、贩卖个人隐私者等等犯罪团伙，只有彻底断链，诈骗才能被禁绝。下一步，警方也应利用人工智能工具打击这些犯罪。近年来，警方、银监会、电信运营商和不少互联网公司携起手来，打击电信、金融犯罪的“智慧大脑”已经初步形成，麟、鹰眼、神茶、神侦、神羊等十余款反诈骗产品已经帮助相关部门破获众多诈骗案件。

而面对迷惑性极强的AI换脸、变声，最好的甄别方法，依旧是AI自身。真的假不了，假的真不了，还是颠扑不破的真理——人的感官存在缺陷，而AI能够很好弥补人的不足。最近，诸如图神经网络、关联图谱等AI反欺诈算法，以及数据指纹、智能多因子认证技术等反诈骗技术也日渐成熟。利用算法和技术创新，计算机可以发挥机器学习的优势，快速发现造假者的马脚，划分真假的界限，为人们决策作出安全、准确的参考。

回溯历史，我们已经多次面临新技术带来的社会挑战。比如照相机的发明改变了传媒界，却也带来了图片造假技术，人们对此也曾忧心忡忡——但总体而言，当下人们依旧相信“有图有真相”，社会的透明度也确实在不断增加。我们有理由相信，未来的世界，也不会因为“造假技术”出现而变得真假莫测，相反，在技术助力下，我们会拥有更明亮的慧眼。

(来源：半月谈、深圳特区报)

社址：上海市小木桥路268弄1号 (200032) 电话总机：34160933 订阅热线：33675000 广告热线：64177374 交通安全周刊电话：28953353 零售价：1.50元 上报印刷