一张图片就能“活化”成视频?

警惕AI深度合成击穿风险底线

《半月谈》记者 张漫子 张超

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。

盗刷人脸、篡改声音,那都不叫事儿

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。

一张陌生人的图片,如何“活化”成为视频?

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。

引导技术向善,完善AI风险治理体系

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。

推进乡村善治,这个小山村有“法宝”

新华社 商意盈 魏一骏 许舜达

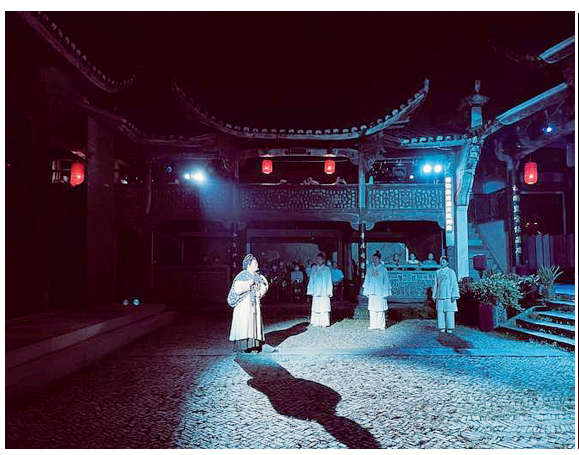
暮色四合,位于浙江西南部山区的江山市大陈村,逐渐褪去白天的暑热。村头,始建于清初的汪氏宗祠热闹起来——灯光准备就绪、演员纷纷到位、观众渐次落座,一场演出正要拉开序幕。

“派由大坂,族衍须江;文光射斗,六州保障!”……“黄连清心经,如为人而友善,决明和肝气,如尊老与敬贤,小麦养心力,如处事之有信……”以宗祠作为演出的天然布景,演出阵容大部分由村民组成,这场名为《大陈见面》的沉浸式村歌剧让台下观众看得入神,不时发出赞叹。

“这部剧创作题材来源于汪氏先人的故事,主题是挖掘孝道故事,将优秀传统文化与当下主旋律相结合,培育文明乡风、良好家风和淳朴民风。”谈起文化治村的力量,大陈村党总支书记汪衍君打开了话匣子。

第六批中国历史文化名村、浙江省3A级景区村庄……如今,在这个粉墙黛瓦、古建错落、卵石铺陈的古村,一系列荣誉纷至沓来。但2005年汪衍君刚回村上任时,大陈村不仅环境脏乱,村集体还负债了68万元。怎么不辜负大家期望,让“后进村”改变面貌?他常思索着。

“一把扫帚扫出一片新天地。”从村容村貌入手,汪衍君发动党员干部带头,自己挨家挨户上门做工作。“我要让大家知道,每个村民都是村子‘大家庭’的一分子,改变环境就要从改变每个人自身做起。”慢慢地,村子里形成了“脸面、灶面、桌面、地面”“四面干净”文化,村民们在浙江省“千村示范、万村整治”工程推动下,共同打造出一个“面子美、里子更美”的新家园。



《大陈见面》村歌剧以宗祠作为演出的天然布景

除了领衔的专业演员,演出阵容大部分由村民组成

搭上了美丽乡村建设规划的快车,汪衍君开始思考如何结合本村特色,让村民精神生活更加丰富。

2007年,借着宗祠续写族谱等契机,大陈村创作了村歌《妈妈的那碗大陈面》。悠长深情的旋律,唱到了全国村歌总决赛的舞台。

“用村歌治理村庄,能管用吗?”面对类似的质疑,汪衍君认为,文化的力量往往见于无形处,而且是自发且持久的。

“歌曲歌词承载的内容有限,为了深入挖掘当地故事,我们开始考虑由‘歌’往‘剧’的方向发展。”2020年初,尽管面临突如其来的疫情等不利影响,浙江小百花越剧团策划人周国清和团队开始了《大陈见面》剧本的创作。

“该剧结合孝道文化和面条产业,呈现农文旅融合的方式,同时选在宗祠空间内演出,容易引起演员和观众的共情,能更好发挥优秀传统文化的教化作用。”周国清说。

剧本创作完成后,听说《大陈见面》村歌剧开始遴选演员,在村里经营农家乐的邱红霞和许多村民第一时间报了名。尽管只是众多配角中的一员,每次演出收入也不多,但对邱红霞来说,参演的意义超出了经济回报本身。

有专家表示,大陈村的村歌、村歌剧等文化活动是农村精神文明建设的有效平台和载体,也是乡村治理体系的重要组成部分。

“只有来自村民自觉的才是扎下根、有生命的文化。”汪衍君说,下一步,村里打算用好村歌剧等形式,用文化的力量持续探索乡村振兴的路径。