“AI换脸诈骗潮”来了?

公安部门:全国爆发消息不实,但涉诈风险正在积聚

zhi liao 知了

一张静态半身照,经过特殊软件处理,就能变成视频,而这样一段视频,可能被不法分子用来诈骗。那么,“AI换脸”是如何实现的?目前全国真的爆发“AI诈骗潮”了吗?对普通公众而言,如何识别“AI假脸”?

记者 于梅君

1 花钱就可“买脸” “AI换脸”暗藏黑产

“AI换脸”技术,一直是网络安全监管的重点对象,许多换脸软件,不得不通过短视频平台、社交账号等私密渠道进行资源共享。

某社交平台一位商家介绍,可以提供国内明星换脸服务,软件价格2000元一套,可送一个模型,“模型可以定制,另外加500元一个。”此外,该商家提供了多组套餐供选择:包括单图换脸、实时换脸、全自动换脸等,每套价格多集中在2000元至3000元。商家表示,可采用多个人脸混合的方式,将使用者本人的脸和明星的脸进行混合,以降低侵权风险。

在某销售网站上,一名软件开发者则展示了如何将知名女星的眼睛和嘴巴拼接融合到一起,生成全新虚拟数字人的过程。该款软件提供了从初级到超级再到全套的不同版本。以初级套餐来说,付费499元可提供软件界面、视频指导、入门级模型、基础指导等。付费2888元,则可得到全套等级的视频换脸模型。

软件不仅可以实现视频换脸,还支持直播换脸。“如果你有看中的脸,可以找到某个人的正面、侧面等各种角度视频拿来训练。”该软件开发者同样表示,“如果你想要明星脸,又怕侵权,不想那么像,就可以找几个人的脸放在一起训练,看上去又像又不像,达到明星脸的效果。”

此外,网络上不乏AI换脸教程。在国内某知名App上输入“换脸”,弹出的高频检索记录显示,有“换脸软件”“换脸App免费”“换脸视频怎么做”“换脸算法”等。

2 视频“大变活人” 是如何实现的

“AI换脸”也叫“深度伪造技术”,堪称现代网络“易容术”,是比PS强大很多的动态换脸技术。这一方式被一些不法分子用于盗刷银行卡、非法办理手机卡等诈骗场景。除了欺骗机器外,还可以通过伪造电话语音、视频通话等欺骗真人,引诱转账、实施敲诈等。那么,“AI换脸”到底是如何实现的?

“深度伪造技术主要分两部分:自动编码器和生成对抗网络。”天津大学智能与计算学部教授翁仲铭介绍,其核心原理,是利用“生成对抗网络”等算法,将目标对象的面部,“嫁接”到被模仿对象上,一般可分四类:重现、替换、编辑和合成。

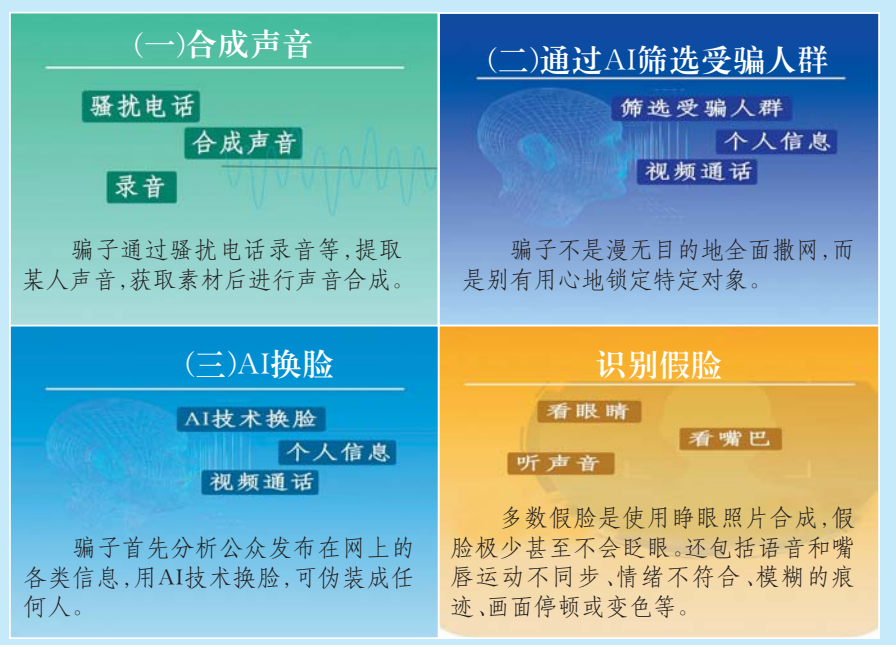
随着深度学习技术的发展,自动编码器、生成对抗网络等,逐渐被应用到“深度伪造”中。那么,什么是自动编码器?翁仲铭介绍,自动编码器是一种神经网络技术,就是把一个人的照片特征抓取出来,然后用数字代表。

有编码器就需要解码器,解码器会把一串串数字再还原成照片。不同解码器,可以还原不同照片,比如朱茵的解码器可以还原朱茵的照片,而还原施瓦辛格照片,则需要施瓦辛格的解码器。

“深度伪造技术,就是在设计、训练精准的编码器和解码器。”翁仲铭解释,因为编码器是抓取照片的特征,所以基本上只需要一套就可以了。但解码器就需要训练很久,以换成史泰龙的脸为例,这个过程需要输入史泰龙600—3000张照片,并经过48—72小时来训练深度模型。

“自动编码器做出的照片是否自然真实,还需要判别把关,这就需要‘生成对抗网络’技术。”翁仲铭解释,这包括两个机器学习模型,分别为“生成网络”和“辨别网络”,这种网络包含两个相互博弈的神经网络,像“猫捉老鼠”一样对抗:一个尽力去忽悠另一个,让它相信自己生成的是真样本,而另一个则尽力去区分真假样本。

生成网络扮演“造假者”,在模型训练后,产生伪造影片;而辨别网络则扮演“检测者”,不断地检视假影片,直至它再辨别不了结果是假的。数据越多,效果越理想,假影片越真实。



3 AI进入快速迭代期,涉诈犯罪风险正在积聚

通过AI换脸和拟声技术,10分钟骗430万元;AI虚拟人在聊天中筛选出受害者,人工接力实施诈骗……近期,多起宣称利用AI技术实施诈骗的案件引发关注。

那么,“AI诈骗潮”是否真的到来了?近日,公安部门确认,“AI诈骗全国爆发”的消息不实,目前此类诈骗发案占比很低。但公安机关已注意到此犯罪新手法,将会同有关部门,开展技术反制和宣传防范。

专家表示,AI在技术上确实能做到换脸、拟音,但被用来进行“广撒网”式诈骗需要具备很多条件。

这类诈骗如果得手必须做到:收集到被换脸对象的个人身份信息、大量人脸图片、语音素材,通过AI生成以假乱真的音视频;窃取被换脸对象的微信号;充分掌握诈骗对象个人身份信息,熟悉其与被换脸对象的社会关系,综合作案成本很高。

公安机关研判,目前AI涉诈案件仍属零星发案状态,成熟的类型化诈骗犯罪,往往具有在全国多地集中爆发的特点,但当前没有成规模的AI诈骗类案件发生。

所以,近期网上“AI换脸、换声诈骗在全国爆发”传言不实,目前全国此类案件发生不到10起,但该动向值得高度关注。网上一键换脸功能的App、小程序有技术滥用风险,需要加强技术防范反制等工作。

国家开发投资集团特级专家赵建强表示,虽然AI诈骗类案件尚未大规模爆发,但AI技术正加速向网络诈骗、虚假信息、色情等领域渗透。“如假冒明星换脸直播、一键脱衣、造谣、制作色情视频等。虽然AI诈骗案件未成气候,但这一趋势值得关注,必须提早防范。”一位反诈民警说。

4 用“白AI”对抗“黑AI”,及时画红线、踩刹车

中国移动信息安全中心品质管理处副处长周晶表示,近年来,各界正在研判AI技术给社会带来的风险和潜在威胁,设法将AI技术发展纳入一定规则中,做到安全可控。

业内人士建议,要加强AI反制技术研究,“以AI制AI”。一些科技公司正加强对图像、声音伪造技术的反制研究,在公安、金融的视频认证场景已有应用。有一线民警建议,要加强AI安全技术应用研发,将AI技术应用于犯罪识别、预警、对抗中,实现以“白”AI对抗“黑”AI。

其次,加强源头治理和行业引导,及时更新、完善相关法律、标准、规则,为AI技术发展保驾护航。“数据是AI犯罪的源头,保护好公民的个人隐私数据安全,就能在最大程度上降低AI违法犯罪的能力。”周晶说。

中国互联网协会监管支撑部主任郝智超建议,AI技术发展还要有相关法律法规来画红线、踩刹车。对AI技术的研发、传播、使用做到有规可循,并根据技术发展实际情况,及时完善对技术服务商行为的规范引导。

此外,还要有针对性地加强反诈宣传。郝智超表示,未来AI可根据大数据创造出无比接近真实的“真实”。“要通过不断的教育,改变大众观念,让人知道眼见不一定为实,有图不一定有真相,提升对网络信息的辨识力。”

知多一点

火眼金睛 揪出“AI假脸”

如今,AI生成的人脸已十分逼真,有没有办法“揪出”它们?

北京邮电大学网络空间安全学院教授周琳娜表示,“AI换脸”检测有多种手段,对普通人而言,可以依据“AI假脸”的破绽进行简单识别。

首先,注意观察“AI假脸”的纹理特征。伪造后的视频人物,眼睛或牙齿轮廓细节容易不一致。

现实中,人眼一般会看向同一个方向,眼球一般颜色一样,但AI生成的人眼,大部分看的方向不一致,且眼球差异较大,两只耳朵的大小甚至高度也不一致。

AI擅长生成概括性画面,但目前仍难以处理像牙齿这样的半规则细节问题。

有时它会生成错位的牙齿,拉伸或收缩每颗牙齿,如果画面中人物的牙口看着很怪异,那么这是假脸的可能性就很高。

再就是看头发。AI生成的直发,看起来像画上去的一样,没有那种略微自然弯曲的感觉。有些不是头发的地方,常被它变成头发纹理。所以,如果图像中的人脸很精致,头发却跟丐帮弟子一样凌乱不堪,那多半这是个假脸。

注意观察“AI假脸”的生物特征。一个健康成年人,一般间隔2-10秒眨一次眼,每次眨眼用时0.1-0.4秒,而在伪造视频中,人的眨眼频率,可能不符合上述规律,甚至不会眨眼,因为它们都是使用睁眼照片进行训练的。

伪造视频往往忽略了“自发的、无意识的生理活动,例如呼吸、脉搏和眼球运动”。

还有,就是注意观察“AI假脸”的嘴部特征。嘴部运动频繁且快速,AI软件无法真实准确地渲染其连续动作,常造成语音和嘴唇运动不同步或情绪不符。

此外,如果人脸背景中的文字扭曲变形,或者背景痕迹模糊、画面停顿或变色,基本就可判定这是AI生成的结果。

专家建议,对普通人来说,图像在最初拍摄时就要标记数据,将拍摄时所在地的坐标、时间、海拔高度、附近的Wi-Fi列表数据等信息,都上传到数据库,在后续传播过程中,可随时通过比对原始数据来确认图像的可信度。