【法学研究】

“深度伪造”技术的刑法规制体系构建*

李　腾

摘　要:作为人工智能应用的重点领域,“深度伪造”技术借助于“深度学习”技术和“生成式对抗网络”模型,能对影音图像进行高精度合成。刑法规制“深度伪造”技术的必要性在于该技术易被滥用;正当性源于对该技术的自主性、便捷性、逼真性等特征及其隐患的审视;可行性基于域外有对该技术进行规制的立法经验。我国刑法规制“深度伪造”技术应秉持以下立场:以客观、中立的态度审视该技术的风险与收益,合理控制刑法介入的深度,以发挥刑法的预防机能为目标。具体可采取三条规制路径:通过强化个人生物识别信息的刑法保护,从前端防范“深度伪造”技术被滥用;通过增设“身份冒用罪”,强化个人身份的刑法保护,规范“深度伪造”技术的合理使用;通过“通知—删除”“发现—标识”“发现—删除”义务的履行,强化网络平台的刑事责任,预防“深度伪造”的危害后果传播。

关键词:“深度伪造”;社会危害;刑法介入;应然立场;规制路径

中图分类号:D924.1　　**文献标识码**:A　　**文章编号**:1003-0751(2020)10-0053-10

一、“深度伪造”技术带来实践挑战

2019 年 8 月底,一款名为 ZAO 的智能换脸软件爆红网络,该软件依托“深度伪造”(deep fake)技术,用户只需在该软件的应用程序中注册并上传自己的头像照片,即可将视频中人物的头像更换为自己的。该款智能应用程序的推广,使得“换脸”对公众而言不再是难以企及的高端科技,任何人都可以享受“换脸”带来的娱乐,甚至体验一次成为经典影视剧主角的快感。但是,仅仅数天后,该款应用程序设计团队便因用户隐私协议不规范、存在数据泄露风险等网络安全问题而被工信部约谈,随即该款应用程序被微信屏蔽访问。①智能换脸应用程序的“夭折”隐含全社会对“深度伪造”技术被滥用的担忧。

在美国,“深度伪造”技术走红同样引发诸多道德和法律问题。早在 2017 年 11 月,美国社交新闻网站红迪网(Reddit)的一位用户便通过“深度伪造”技术,将电影《神奇女侠》中女主角加尔·加多(Gal Gadot)的脸嫁接到色情电影中,并将该视频上传至网络,引发了极大轰动。2018 年 1 月,“深度伪造”技术的应用程序正式上线,换脸的对象从公众人物迅速扩展至亲友、同学、同事之间,加剧了合成视频在网络的传播,红迪网不得不关闭“深度伪造”技术讨论区以减少该技术被滥用的可能性。同年,加拿大蒙特利尔大学三名博士联合创办的一家名为“琴鸟”(Lyrebird)的公司更是开发出了“语音合成”软件,只要对目标人物的声音进行 1 分钟以上的录音,再将录音“喂”给该软件进行学习,就能得到一个特别的密钥,利用这个密钥就可以用目标人物的声音伪造出任何想说的话。②由此,经过“深度伪造”技术处理的视频、声音成为具有高度欺骗性的影音资料,使人们不仅陷入“耳听为虚”的窘境,更陷入“眼见不为实”的认知混乱,而换脸技术与语音合成技术的叠加适用更增加了合成视频的辨别难度,也加剧了各界对“深度伪造”技术被滥用的担忧。

作为人工智能技术发展的新兴领域,“深度伪造”技术确实能增加社交的互动性、娱乐性,这种功效与国家大力扶持人工智能产业发展的初衷是一致

收稿日期:2020-07-09
*基金项目:国家社会科学基金一般项目“民营企业的刑事法律风险及刑法保护研究”(19BFX072)。
作者简介:李腾,男,中国政法大学刑事司法学院博士后研究人员(北京　100088),荷兰鹿特丹大学访问学者。

的。但是,“深度伪造”技术被滥用会引发诸多风险,甚至造成严重的危害后果。面对“深度伪造”技术的“两面性”,刑法应当秉持何种立场来看待技术革新?在国家尚未对人工智能发展建立完备的法治体系的背景下,该如何平衡对自由价值的崇尚与对秩序价值的追求?面对“深度伪造”技术被滥用的风险,刑法应以何种路径对该技术进行规制?回应上述议题不仅能使我们正视“深度伪造”技术发展的利弊并将其限制在合理的应用场景,更能在宏观层面厘清刑法介入技术发展的程度,理顺技术发展与社会应用的关系,以刑事法治保障社会福祉的增加。

二、“深度伪造”技术的解构

当前,“换脸”已经成为合成视频应用的代名词,其所依托的“深度伪造”技术也进入大众视野。“深度伪造”作为“深度学习”(deep learning)与“伪造”(fake)的融合,通常是指基于人工智能合成技术,将已有的视频、音频、图片叠加至目标影片或图像上,从而创制出新的影音图像的技术。③随着技术的革新,“深度伪造”技术已经不限于视频、音频、图像的合成与伪造,而成为涵盖现实伪造与虚拟现实创制的应用技术。“深度伪造”技术的应用所涵盖的领域包括:第一,通过对视频中的人脸进行更换,使被替换者能够实施自己从未有过的行为;第二,通过对目标对象的口型、语速和面部表情进行重塑,使目标人物表达出非真实性的言论;第三,通过对目标人物的声音进行学习,创制出目标人物的声音模型并进行非真实性的语义表达;第四,通过软件创建现实中并不存在的人物形象。④日本人工智能公司数据网格(Data Grid)开发的人工智能软件已经能自动生产虚拟人物的全身模型,并将之应用于未来的服装行业。⑤由于音频图像合成技术发展得较为成熟,“深度伪造”技术在实践中主要被用于对视频、音频的合成。上述技术也可以叠加应用,如通过人脸和语音的双重伪造,创制出某一政治家从未发表过的演讲视频。厘清“深度伪造”技术通过何种运作机理完成上述合成过程,无疑是理解该技术的基础,也是确立刑法在何种程度上介入这一新兴技术领域的前提。

“深度伪造”技术合成的视频、音频之所以能起到以假乱真的效果,主要缘于“学习—监督”算法的设定与应用:一方面通过“深度学习”技术的应用,使程序具有对样本进行深度分析、转化、重构的能力,能在短时间内提炼出样本视频、音频、图像的共性,合成新的密钥,在此基础上创制学习成果并予以输出;另一方面通过“生成式对抗网络”算法模型的设定,使“深度伪造”技术能够自动、反复地对已生成的视频、音频、图像进行修正、重构,并在自动学习的过程中不断提升成果质量,达到“温故而知新”的效果。

(一)“深度学习”技术为“深度伪造”提供算法支撑

“深度学习”作为人工智能在智能化进程中的核心技术,为“深度伪造”技术的发展提供了算法支撑。“深度学习”技术通过相互联结的多层次神经网络,对输入的数据进行分布式信息储存、大规模并行处理,并根据设定的程序对数据进行计算分析,完成设定的任务。“深度学习”技术能在短时间内完成对视频、音频、图像的学习并输出新合成的影音图像,这成为“深度伪造”走向智能化的技术基础。以换脸程序的应用为例,当海量的样本图片被“喂”给“换脸算法”后,“换脸算法”会对不同场景、角度、光线图片下的目标人脸进行识别,转换为不同的二进制编码,并对上述二进制编码的共性进行提炼,形成新的目标人物的二进制编码。此后,“换脸算法”会将新合成的目标人物逐帧映射到原有视频人物的面部,形成新的视频影像,以此完成对样本图片的转化、分析、重构。前文提及的“琴鸟”公司创制的“语音合成”软件也是基于同样的运作原理,即将目标人物的声音进行输入并转换为二进制编码,在长时间的语音编码的基础上将目标人物的语音特点进行提炼,形成新的二进制编码,他人只要利用这套编码进行语音转换,就可生成目标人物的语音内容。

“深度学习”技术通过模拟人脑的运行规律,在此基础上构建一套能够不断优化学习的神经网络,连续对目标数据进行转换学习,在大量目标数据的基础上提炼出目标数据的共性。随着算法训练数据量的增大,“深度学习”技术所提炼出的目标数据的共性特征愈发精确,最终输出的算法结果也愈发接近目标数据。只要给予“深度学习”技术足够的样本数据,其合成高质量的视频、音频内容就只是时间问题。伴随着云计算在人工智能领域的广泛运用,在短时间内合成视频、音频不再有技术障碍。

（二）“生成式对抗网络”算法模型不断优化“深度伪造”合成效果

“深度学习”技术为“深度伪造”创制合成视频、音频提供了基础算法，但真正促进“深度伪造”技术向高质量迈进的是“生成式对抗网络”算法模型的构建。不同于传统“深度学习”技术所采用的单链条学习网络，“生成式对抗网络”采用两条并行的学习神经网络，在互动式对抗中逐步优化学习结果。其中，第一条学习神经网络通过对“深度学习”技术的应用，初步创制出合成的影音图像；第二条学习神经网络通过将合成的视频、音频与样本数据进行对比，对合成效果进行二次验证；如果合成数据与样本数据之间的差异达到一定比例，则新合成的视频、音频会被退回到第一条学习神经网络中重新进行学习、分析、提炼，直到检测结果符合设定的误差范围。正是在这种“对抗式”学习过程中，“深度伪造”技术不断进行自我学习、自我优化，也正是在这种“双向互动式”学习过程中，两种算法通过相互监督，不断改进，直至产生高质量的影音图像。在“生成式对抗网络”模型设计中，两条学习神经网络中的一条担负着生成功能，另一条担负着鉴别功能，两者能在大数据的“喂食”下不断进行无监督学习，各自在进行算法训练的同时督促对方算法的提升。因此，当训练数据足够充分时，“深度伪造”技术创制出的视频、音频完全能起到以假乱真的效果。

三、对“深度伪造”技术进行刑法规制的证成

“深度伪造”是人工智能技术的综合运用，其为现代生活带来娱乐体验的同时，也会带来诸多风险。当风险演化为现实危害时，就需要刑法介入，将该技术限制在合理的应用场景。

（一）刑法规制的必要性：对“深度伪造”技术造成现实危害的隐忧

影音图像的合成并非新兴技术，其关注度空前高涨只是由于此前合成技术的应用具有较高的壁垒而与普通大众无缘。令人遗憾的是，合成技术借助于“深度伪造”技术“走下神坛”，却以“异化”的方式进入公众视野。随着 2017 年红迪网上虚假成人情色视频的出现，“深度伪造”技术以一鸣惊人的方式完成了首秀，此后其实践应用日益呈现出“异化”的趋势。继前文提及的加尔·加多之后，《哈利·波特》中赫敏的扮演者艾玛·沃特森（Emma Watson）、《这个杀手不太冷》中女主角的扮演者娜塔莉·波特曼（Natalie Portman）甚至美国总统特朗普的女儿伊万卡·特朗普（Ivanka Trump）都曾成为合成情色视频的受害者。在我国，部分网站上兜售的明星情色视频也使众多女性成为“深度伪造”技术被滥用的受害者。⑥这些合成视频本已严重侵犯被害人的名誉权、肖像权等基本权利，再被用来骚扰、侮辱、勒索被害人时，更会给被害人带来二次伤害。⑦此外，当运用“深度伪造”技术制作的合成视频被用于“人脸验证”领域时，又会造成他人财产损失，特别是在支付型应用程序中，行为人通过制作“眨眼”“摇头”等视频以模拟被害人，骗过应用程序对活体认证的要求，便可实施贷款、盗刷资金等侵财行为，“深度伪造”技术由此沦为犯罪工具。⑧

如果说“深度伪造”技术被应用于上述场景仅会对个体法益造成侵害，尚不足以引起重视，当该技术借助于对他人身份的冒用而影响社会稳定时，就会造成更大的危害。特别是当虚假视频、音频配合着谣言在网络上进行扩散时，不仅会突破时空的限制，加剧控制难度，还极有可能导致社会秩序陷入混乱与无序。⑨例如，当“深度伪造”技术被应用于金融领域时，可能引发金融体系整体性动荡。由于金融市场具有对信息高度敏感的特性，任何一条敏感信息的传递都会迅速引发金融市场波动，当行为人制作“黑天鹅”事件的虚假视频并通过网络进行传播时，在金融市场缺乏预期的情况下，这一虚假视频无异于重磅炸弹，资金出于避险需求必然在极短时间内大幅流出金融体系，继而引发证券、债券、外汇市场的连锁反应，严重影响金融市场的稳定和预期。⑩同理，在我国全力迎战新冠肺炎疫情期间，如果有人利用“深度伪造”技术制作出某权威专家发表“疫情不可防、不可控”的虚假视频，就极有可能引发全社会的不稳定。在信息爆炸时代，民众对信息的真伪虽然有一定的鉴别能力，但当谣言以视频的形式出现时，不仅深度契合“眼见为实”的认知共性，更在权威专家“背书”的基础上大幅增加谣言的可信度，由此导致虚假视频、音频的影响范围越广、影响深度和危害性越大。当“眼见不为实”成为社会常态时，必然会对长期延续下来的认知体系产生强烈冲击，激起人们对一切事物的怀疑，直接冲击社会信任体系。⑪由此导致在公共场合需要使用视频、音频以佐证相关行为时，出现无法进行有效证明的

窘境,社会信任的基石受到侵害。

上述情况足以引发社会秩序的动荡与不安,但“深度伪造”技术被滥用的危害还不止于此。如果该技术被用于挑拨国家关系,则世界范围内地区稳定、国际关系稳定都会陷入不确定的风险之中。当“深度伪造”技术被用于制造国家间宣战画面或者创制并不存在的恐怖袭击画面时,极有可能引发国家间的过激反应或者将国家间的对立、冲突推向无法挽回的境地。这种假设并非危言耸听。当前,“深度伪造”技术已经能通过对 2 小时语音资料的深度学习,在 5 天内合成特朗普代表美国向世界任何国家宣战的语音,语音合成效果真假难辨。⑫德国有研究者在“深度伪造”技术的应用上已经能将其欲表达的语义及相应的表情变化复制到一些国家领导人的讲话视频中。⑬难以想象,这样的音频或视频的传播会对本已摩擦不断的世界造成何种影响。

正是由于“深度伪造”技术的滥用会引发对个人人身权利、财产权利的侵害,引发社会秩序混乱甚至对国家安全、世界安全造成威胁,所以对“深度伪造”技术应用的担忧并不是空洞的设想。这为刑法介入技术发展以防范技术被滥用奠定了现实基础。

(二)刑法规制的正当性:对“深度伪造”技术的特征及其隐患的审视

单纯的影音图形合成并非新兴技术,“深度伪造”技术之所以引发社会的担忧,主要源于该技术带来的视频、音频合成方式变化为其被滥用提供了极大的便利。较之传统的视频、音频合成技术,“深度伪造”技术的特征主要表现为以下三点。

第一,视频、音频合成的“自主性”。“深度伪造”技术改变了传统视频、音频合成路径,使系统生成视频、音频具有自主性,不仅大大提升合成效率,还极大地降低合成成本。“虽然电影特效和深受广大网民喜爱的‘PS’技术都是‘无中生有’的伪造,但它们是被动性伪造,依赖个体参与,是单一的、独立性的伪造。而深度伪造借助人工智能技术,可以实现机器的自主学习,伪造的行为脱离了人的参与,通过人工智能的计算机程序自动完成。它将伪造从被动性伪造推入了自主性伪造的全新阶段。”⑭传统的视频合成技术不仅需要大量工作人员对目标人脸的特征进行提取、重构并在视频中逐帧进行替换,还需要工作人员根据讲话内容再对目标人脸的嘴型进行修正。这一过程不仅需要专业团队的协作、耗费大量的时间,还难以保证视频合成效果的流畅性。但是,“深度伪造”技术使实践中对目标人脸特征的提取、转化、重构工作完全基于“深度学习”后的自动化运作,极大地降低了各项成本投入。

第二,视频、音频合成的“便捷性”。在“深度伪造”技术被广泛应用之前,视频、音频合成不仅具有专业技术门槛,还需要高昂的费用,普通民众对此望而却步。因此,在较长时间内,高质量的视频、音频合成技术的应用场景主要限于商业影视剧制作领域,难免陷入“曲高而和寡”的境地。但是,“深度伪造”技术的兴起打破了技术、成本壁垒,视频、音频合成成为大众科技。伴随着网络公司通过应用程序对“深度伪造”技术进行推广,“深度伪造”技术的受众面愈发广泛,其技术门槛进一步降低。只要自己的电脑、手机有足够的运行程序,每个人都可以享受“换脸”带来的愉悦感。

第三,视频、音频合成的“逼真性”。“深度伪造”技术基于“对抗式生成网络”,使应用程序在“生成—鉴别”模型中能够不断对样本数据进行分解和重组,而每一次“学习—监督”过程的淬炼都会使目标人脸的替换效果更上一层楼。正是在这种无监督学习的模型设计中,“深度伪造”技术能不断对合成过程进行打磨,不断对合成效果进行升级。如果样本数据足够充分,目标人物表达不同感情时所对应的面部表情、在不同光照下面部影像的成像规律甚至在不同场景中动作习惯的细微变化,均能被“深度伪造”技术捕捉并予以演绎。“深度伪造”技术合成比人工识别更加精准,制作效果更加逼真。

由于“深度伪造”技术具有“自主性”“便捷性”“逼真性”特征,使得高质量的影音图像合成成为易于掌握的技能,但受众群体过于广泛易引发各种后果,加剧了人们对“深度伪造”技术的隐忧。如果视频、音频合成被用于服务社会,自然能增进全社会的福祉;但事实是,网络的互联互通已经将世界联系为一个整体,数据的跨境流动已成为常态,获取任何人的视频、图片都不再是难事,通过“深度伪造”技术的应用能够对视频、图片进行深度学习、完成伪造并将之应用于任何场景。特别是对公众人物而言,涉及其面部特征、语音特征的样本数据较多且易于获取,通过“深度伪造”技术进行大样本的学习和模仿,所制造出的合成视频、音频不经过专业性的技术鉴别就难判真伪,由此引发的危害后果实难控制和

预防。这更加决定了刑法应当有所作为,对“深度伪造”技术的应用进行规制。

(三)刑法规制的可行性:对“深度伪造”技术相关域外立法的借鉴

针对“深度伪造”技术被滥用所引发的危害后果,有的国家出台了专门性立法对该技术的应用场景进行限制,并将部分滥用该技术的行为视为犯罪行为予以规制。这为我国对“深度伪造”技术的滥用进行刑法规制提供了经验借鉴。

作为“深度伪造”技术的发源地,美国尝试在联邦和州两个层面通过立法规制该技术的滥用。在联邦层面,2019 年 6 月,由众议员 Clarke 提出的《深度伪造责任法案》(Deep Fakes Accountability Act)旨在防止本国及外国势力利用“深度伪造”技术对美国大选进行干预,要求合成视频创制者以在视频中添加水印及个人声明的方式对“深度伪造”技术进行应用;[15]对于违反标识义务,意图羞辱他人或者干扰政治运作、引发武力或外交冲突而发布合成视频的行为,将面临最高 5 年监禁的刑事处罚。[16]在州层面,2019 年 7 月,弗吉尼亚州对《复仇情色法案》(Nonconsensual Pornography Law)进行修正,将利用“深度伪造”技术合成他人情色视频、图片并予以传播的行为定义为“未经他人同意而将他人的色情视频、图片予以传播”的犯罪行为,对该行为可判处最高 12 个月监禁及 2500 美元罚款。[17]由此,弗吉尼亚州成为美国第一个对“深度伪造”技术进行立法回应的地区。同年 9 月,旨在防止“深度伪造”技术影响选举公正性并保障选举安全的得克萨斯州《以虚假视频干预选举结果的刑事犯罪法案》(Relating to the Creation of A Criminal Offense for Fabricating A Deceptive Video with Intent to Influence the Outcome of An Election Bill)生效,该法案将利用“深度伪造”技术在选举 30 日内制作有关候选人的虚假视频并进行传播的行为规定为犯罪行为。[18]同年 10 月,加利福尼亚州通过了《730 号议会法案》(Assembly Bill No.730),将在选举前 60 日内制造、传播经过篡改的有关政治家、选举候选人的视频、音频、图像的行为视为犯罪行为。[19]与美国的治理模式不同,欧盟并未针对“深度伪造”技术出台专门性立法,而是以《通用数据保护条例》(General Data Protection Regulation)为依据,将数据权作为独立的宪法性权利和公民基本权利予以保护,以规范数据采集、储存、保管、应用流程,应对“深度伪造”技术被滥用对公民身份的冒用和隐私的侵犯。[20]

从上述立法例可以看出,域外立法以刑事责任规制“深度伪造”技术被滥用的逻辑在于:限制该技术的应用场景,对以影响选举为目的制作并传播合成视频、音频、图像的行为予以限制或禁止,对制作、传播未经他人同意的合成色情视频、音频、图像的行为予以禁止;通过要求制作者对使用“深度伪造”技术制成的视频、音频、图像进行标记,防止该技术被用于非法场景;通过加强前端生物识别信息和数据权利保护,防范“深度伪造”技术被滥用。

我国尚未针对“深度伪造”技术出台专门性法规,但 2019 年 11 月 18 日国家互联网信息办公室、文化和旅游部、国家广播电视总局联合印发的《网络音视频信息服务管理规定》(以下简称《规定》)明确要求,网络音视频信息服务提供者、使用者在对基于深度学习和虚拟现实技术合成的视频、音频信息进行制作和传播的过程中,应当进行明显的标识,同时不得利用上述技术制作、发布、传播虚假新闻。[21]这一方面通过明确制作者的标识义务以区分视频、音频是否系合成,另一方面通过禁止“深度伪造”技术在新闻信息领域的应用而限制该技术的应用范围,确保新闻信息发布的准确性、权威性。《规定》通过具体的技术路径设置展现了国家层面对“深度伪造”技术应用要区分场景的规制态度,对于未来出台相关国家立法具有指导性意义。由于《规定》对“深度伪造”技术应用场景的区分不够细化以及缺乏配套的规制措施,使得其缺乏可操作性。例如,个人基于娱乐的目的,利用“深度伪造”技术将自己或朋友的头像与经典影片中的情节合成新的视频并在社交网站上与他人分享,但并未履行标识义务,对此行为该如何规制?《规定》中并没有具体的应对措施。再如,有人利用“深度伪造”技术对央视新闻联播的内容进行拼接、合成并在网络上传播,但系以幽默、诙谐的方式赞扬国家发展成就,该行为是否属于《规定》禁止的伪造新闻行为而应一律禁止?这恐怕也值得商榷。对于视频、音频的制作者未履行《规定》中的标识义务和禁止适用义务,《规定》中并没有相应的规制措施。这就使得《规定》的宣示意义大于实践价值,造成法律保护的缺位。

一般而言,对于技术的规制,应当先从民事法律、行政法律规制开始,最后才用刑法规制。但是,

由于“深度伪造”属于新兴技术,现行民事法律、行政法律针对该技术的规范尚不完善,在此情况下,刑法应当有所作为,直接对滥用该技术造成严重社会危害的行为进行独立评价。

四、刑法规制“深度伪造”技术的应然立场及路径展开

技术创新的刑法规制是一个永恒的难题,规制过多会限制技术的发展与应用甚至阻碍技术创新,规制缺位会使技术被滥用而导致严重的危害后果。因此,当刑法介入技术创新时,需要在顶层设计上确立规制的应然立场,在此基础上进行规制路径完善。

(一)刑法规制“深度伪造”技术的应然立场

通常而言,技术的发展与革新是沿着从技术原理向实践应用转化再向意识形态阶序发展的脉络演进的,在技术逐渐社会化的过程中,刑法介入的程度应当愈发深入,并最终形成刑法规制与技术发展的平衡状态。[22]但是,对于“深度伪造”技术,不能完全按照上述阶序演化路径进行刑法规制。因为“深度伪造”技术创立之初就呈现出“异化”的趋势,由此引发的技术被滥用的隐忧已经超过了技术红利带来的娱乐体验,刑法的介入更显急迫。这就需要首先确立刑法对“深度伪造”技术进行规制的基本立场,再以此为基础进行规制路径展开,通过逆向治理思维完成对“深度伪造”技术进行刑法规制的顶层设计。

1.宏观立场:以“两面性”视角控制规制的深度

“深度伪造”技术应用的差异化会导致两极化的结果。该技术被用于损害他人人格、危害社会管理秩序、危害国家安全,导致潜在的风险向现实危害后果转化时,需要刑法的强势介入;该技术被应用于社交、娱乐场景时又显露出“友好面相”,“可以让电影、纪录片等艺术创作突破时空限制,以更真实的方式呈现,也能实现替身演员演出等效果;可以给通过视频进行的批评、讽刺、戏仿等提供新的表达形式;可以创造虚拟主播来播报新闻、天气预报等内容;也可以给医疗、零售、娱乐等领域提供更具亲和力的人形问诊机器人、虚拟客服、虚拟偶像等”[23]。因此,刑法应承担起应有的职责,强化对技术扭曲使用行为的制裁,控制和约束技术的滥用;同时,刑法应当控制对技术行为的评价深度,以保护正常的技术应用活动。[24]正是基于“深度伪造”技术的“两面性”特征,美国社交网站 Facebook 在过去两年间并未禁止“深度伪造”视频在该网站上传播。

在全球面临第四次科技革命的背景下,实现关键领域的核心技术突破成为国家间科技竞争的主要角力场,也成为我国国家竞争力实现“弯道超车”的关键。这就要求我国刑法在对“深度伪造”技术进行规制时,既要考虑该技术被滥用的危害,又要兼顾该技术在实际应用中的有效性。回顾我国信息化发展的历程可以看出,我国对大数据、物联网等新兴产业采取普遍支持的态度,在一些国家禁止使用的人脸识别技术在我国公共场所也得到普遍使用。对于以“深度伪造”技术为代表的人工智能发展,我国应以“两面性”视角进行合理的限制。限制技术的应用场景而非完全禁止技术的应用,是刑法介入“深度伪造”技术领域的应然立场。

2.微观立场:以预防机能实现规制的效果

对于“深度伪造”技术被滥用,可以适用我国《刑法》中的若干罪名。例如,当该技术被用于侵犯公民名誉、财产时,可以以侮辱罪、盗窃罪、(信用卡)诈骗罪甚至传播淫秽物品(牟利)罪对相关行为予以规制;当该技术被用于侵犯公共法益时,可以根据实行行为所侵害的具体法益类型,以编造并传播证券或期货交易信息罪、损害商业声誉罪、编造或故意传播虚假信息罪、伪证罪、辩护人或诉讼代理人伪造证据罪予以规制;当该技术被用于侵犯公共安全、国家安全时,可以以煽动分裂国家罪、煽动颠覆国家政权罪、宣扬恐怖主义或极端主义或煽动实施恐怖活动罪、战时故意提供虚假敌情罪、战时造谣扰乱军心罪、战时造谣惑众罪对危害行为予以规制。上述罪名为规制滥用“深度伪造”技术的行为提供了相对完整的刑事制裁体系,但这套产生于农业社会、成熟于工业社会的刑法理论与刑罚观念过度强调事后惩罚而忽略事前预防,在信息社会呈现出规制的无效性和适用的滞后性。[25]这种缺陷突出表现在:如果说“深度伪造”技术被用于侵犯个人法益时,刑法的事后规制尚能使被侵害的社会关系得以恢复,当“深度伪造”技术被用于危害国家、公共安全时,所造成的严重危害往往难以通过事后惩治予以补救。特别是网络的无限延展性和无国界性使得“深度伪造”技术被滥用的危害后果传播具有不可控性、危害性呈数量级增长,行为的社会危害性远远超过刑罚的严厉性,更加凸显刑法的事后惩戒机制在网络

治理领域的无效性。因此,近年来,刑法预防机能的发挥得到更多提倡,“预备行为实行化”“帮助行为正犯化”“民事、行政违法行为犯罪化”等理论在刑事立法中得以彰显,刑法规制的行为类型愈发向前端延伸,旨在将危险后果控制在萌芽状态。[26]对于“深度伪造”技术的刑法规制同样应强化对前端行为的制裁,突出刑法的预防机能,将实然性危害消灭于萌芽状态。

(二)以“信息保护+应用治理+平台监管”模式构建刑法规制路径

“深度伪造”技术的刑法规制是一项系统性工程,从底层数据收集到中层对数据加工处理的算法指令再到上层数据和算法的最终应用,构成规制的重要维度。[27]在多层次规制维度中,算法治理属于技术发展的范畴,刑法对此不宜过多介入;数据治理和应用场景限制是刑法规制“深度伪造”技术的重点领域。此外,随着网络平台在网络社会结构中发挥枢纽作用、扮演着“准政府”的角色,不断强化网络平台的监管责任成为刑法规制“深度伪造”技术的重要抓手。[28]由此,可以构建“信息保护+应用治理+平台监管”的刑法规制路径,充分发挥刑法的预防机能,确保“深度伪造”技术在刑事法治体系下有序发展。

1.强化个人生物识别信息的刑法保护

滥用“深度伪造”技术行为的本质是借助于他人的生物识别信息实现身份冒用,这就要求刑法特别关注该技术应用的前端保护,强化对侵犯个体人脸特征、语音特征等生物识别信息行为的规制,减少个人身份信息相关数据被滥用的可能性。事实上,换脸软件事件中直接导致 ZAO 应用程序下架整改的原因并非单纯的对换脸技术本身的担忧,还包括该款应用程序在用户协议中要求用户或肖像权人同意 ZAO 及其关联公司永久保留、存储、转让、使用用户肖像,从而引发了对个人生物识别信息泄露的担忧。[29]

当前,“由于社会数据化进程的加快,各种之前不被关注的个人信息所具有的潜在价值被逐步挖掘,以往单纯的、没有意义的‘碎片’变为了对主体自身具有重要价值的‘信息’,越来越多的个人信息需要被纳入法律的保护范围”[30]。特别是随着对生物识别信息进行研究和应用的社会探索的加快,生物识别信息已经成为全方位“解锁”个人信息的密钥。以人脸信息为例,由于人脸具有稳定性、易于识别、易于采集的特征,使得人脸识别技术成为人工智能迅速扩展的领域,已经在手机解锁、智能支付、门禁安防、交通出行、金融认证甚至公共安全监控识别领域得到全场景推广应用。当各个应用场景都对人脸信息进行识别和验证时,不仅完成对个体人脸信息的抓取与保留,还能通过人脸这道“门”与既有数据库中的个体数据进行关联、对比,于是个体的身份信息、行踪轨迹、电话号码、收入状况、消费习惯等信息均可能遭到泄露。人脸的功能从体现个人外在形象的社会属性逐渐呈现出具有个体特质的私人属性,人脸信息也从公开信息逐渐演化为个人隐私信息。最关键的是,由于人脸、声音等外在生物特征能被以非接触的方式予以采集和应用,个体极有可能在无意间被第三方获取个人生物识别信息,该信息作为唯一身份标识一旦泄露,身份被冒用的风险极高且由于个人难以追索控制,最终会对个体的人身权利和财产权利造成不可逆的损害。[31]因此,人脸信息已经不是传统技术条件下的公开信息,其已经成为事关个人身份、财产、隐私的个人机密敏感信息。“深度伪造”技术无疑能放大个人生物识别信息被滥用的危害性,因而通过刑法规制“深度伪造”技术也是信息化时代强化个人信息保护的必然要求。

在我国,《全国人民代表大会常务委员会关于加强网络信息保护的决定》《网络安全法》《刑法》以及《最高人民法院、最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》(以下简称《解释》)共同构成规制侵犯公民个人信息行为的刑事制裁体系。其中,《全国人民代表大会常务委员会关于加强网络信息保护的决定》《网络安全法》作为前置法明确将个人生物识别信息作为公民个人信息的应有之义予以保护[32],《解释》再次确认了对具有身份识别功能的生物识别信息的刑法保护[33]。但是,由于司法机关对公民个人信息的认识仍停留于对信息隐私特性的保护,导致实践中司法资源过多投放到对个人身份信息及个人活动信息这类具有较强隐私属性的信息的保护上,对人脸、声音这些具有较强外在社会交往属性的生物识别信息却有意无意地忽视了,从而造成刑法保护的缺位。因此,在现有刑法体系下,激活刑法对个人生物识别信息的保护显得尤为迫切。这要求司法机关深刻认识到,不同于一般的个人信息,生物识别信息是与个

人的身份、财产密切相关的个人机密敏感信息，需要投入更多司法资源加以保护。司法机关应当通过侵犯公民个人信息罪的实践扩容，加大对个人生物识别信息的刑法保护，从源头减少人脸、声纹等个人生物识别信息数据被通过“深度伪造”技术进行滥用的可能性。

2.通过增设罪名规制身份冒用行为

在“深度伪造”技术被滥用的过程中，法益侵害直接指向对他人身份的冒用。因此，在强化前端个人信息保护的同时，刑法应着重规制“深度伪造”技术的使用过程，通过限定该技术的应用场景、明确技术使用者的责任，将该技术的应用规范在有序状态下。一方面，当使用者基于娱乐的目的，制作自己、同学、亲友及他人的合成影音图像时，由于该类行为并不具有社会危害性，故不应以刑法进行评价。但是，由于“深度伪造”技术的高度模仿性使得合成的影音图形有可能被他人用于违法犯罪场景，所以制作者、传播者仍应谨慎地使用“深度伪造”技术。制作者、传播者可通过履行如下义务，作为免责的依据。首先，使用者不得将“深度伪造”技术应用于色情、暴力、恐怖主义等违法犯罪场景；其次，使用者应当履行标识义务，在利用“深度伪造”技术合成影音图像时进行明显的标记；最后，对于法律、法规、行政规章在现阶段及未来列举的“深度伪造”技术应用场景“负面清单”，制作者、传播者应当恪守。由此，可以充分发挥“深度伪造”技术在商业、娱乐领域的正向价值。另一方面，当使用者未经他人同意而使用“深度伪造”技术制作他人的影音图像时，由于合成的影音图像可以被用于后端的违法犯罪行为，所以即便合成的内容不违法，合成行为也构成对他人身份的冒用。特别是随着个人生物识别信息在“识别”“认证”中的作用愈发凸显，利用“深度伪造+个人生物识别信息”更易完成对他人身份的冒用，使得该类行为的危害性愈发明显，应当将该类行为纳入刑法规制体系。

我国《刑法》中存在一批针对身份冒用行为的罪名，招摇撞骗罪、冒充军人招摇撞骗罪以及使用虚假身份证件、盗用身份证件罪均是直接规制身份冒用行为的罪名，合同诈骗罪、信用卡诈骗罪、妨害信用卡管理罪中也附随对身份冒用行为的评价。但是，上述罪名体系无法完全涵盖利用“深度伪造”技术实施的身份冒用行为。招摇撞骗罪的冒充对象是国家机关工作人员，冒充军人招摇撞骗罪的冒充对象是军人，此两个罪名均是针对特殊群体的立法保护，利用“深度伪造”技术对其他群体身份的冒用不在其规制之列。使用虚假身份证件、盗用身份证件罪的犯罪对象是法定的身份证明文件，所规制的是行为人通过伪造、变造、盗用他人居民身份证、护照、社会保障卡、驾驶证等法定证明证件进行身份冒用的行为等行为类型，对利用他人生物识别信息进行身份冒用的行为无法予以规制。合同诈骗罪、信用卡诈骗罪、妨害信用卡管理罪均是将身份冒用行为作为手段行为进行评价的，如果行为人未实施后续的取财、领卡行为，则身份冒用行为不再进行单独评价。因此，现有的刑法罪名无法对滥用“深度伪造”技术进行身份冒用的行为进行评价。“在信息高度发达的当今社会，身份承载了更多的价值和内容，各种身份信息成为电子交易、入学就业的凭证，例如身份证号码、姓名、性别、出生日期、职业、银行户名及账号、家庭住址等，这些信息在某种程度上成为确认一个人身份的手段和凭证。一方面，信息社会中获取他人身份信息相对简便和成本低廉，盗窃身份的违法犯罪行为开始不断增多，另一方面，伴随着个人身份信息承载了越来越多的作为谋取便捷、舒适生活的手段，甚至可以作为实施犯罪的‘通行证’，身份犯罪开始爆发式增加。”[34]身份冒用行为的危害性增加与相应的刑法规制缺失的状态日益凸显。身份冒用不是新型危害行为，近几年来被媒体曝光的冒名顶替上大学事件就是该行为危害性的真实写照。在这些事件中，身份被冒用者不只是失去了公平的求学机会，甚至整个人生被改写。在互联网背景下，借助于“深度伪造”技术的身份冒用行为将成为一种更加常态化也更具有社会危害性的行为，刑法有必要通过增设“身份冒用罪”对该行为进行规制[35]。

3.强化网络平台的刑事责任，预防危害后果传播

互联网历经 Web1.0“人机互动”到 Web2.0“人人互动”再到 Web3.0“空间互动”的迭代升级，已经逐渐摆脱工具属性，成为人们互动交流的秩序空间。在现实空间与网络空间并存的双层社会体系下，网络平台已经逐渐摆脱被管理者的角色而越来越多地承担网络空间的管理、自治义务，扮演着管理者的角色。[36]作为信息交互、数据留存、秩序维护的中枢场所，网络平台既能导致“深度伪造”技术被滥用的危

害后果呈数量级增长，也能将这种危害后果控制在“襁褓”之中，因而不断强化平台监管的刑事责任是规范“深度伪造”技术的有效方式。

我国《刑法》以拒不履行信息网络安全管理义务罪、非法利用信息网络罪、帮助信息网络犯罪活动罪三个罪名构建了网络平台运营的刑事责任，然而，由于“深度伪造”技术的应用成果、应用场景兼具违法与合法的可能性，导致实践中网络平台主体对合成影音图像的内容存在认识错误。在现行司法解释尚未对上述罪名的类型化特征进行总结的前提下，笔者认为当前应以“通知—删除”义务的履行作为规范罪名适用的标准，在未来技术成熟时以“发现—标识”“发现—删除”义务的履行作为评价网络平台主体刑事责任的依据。

从科技发展趋势来看，对“深度伪造”技术及其传播内容的审查必然成为网络平台审查义务的应有之义。[37]但是，由于“深度伪造”技术的识别尚不成熟且通过技术合作的方式识别“深度伪造”技术的应用会增加网络平台的运营成本，所以真正能使用“深度伪造”识别技术并对“深度伪造”技术的内容进行实质审查的仅限于少数互联网头部平台。从刑法适用的平等性、均衡性的角度出发，目前应以“通知—删除”义务的履行作为评价网络平台主体罪与非罪、此罪与彼罪的标准，以预防“深度伪造”技术被滥用后果的传播。具体而言，网络平台主体应当履行如下义务：其一，负有对合成的影音图像内容进行审查的义务，对有涉黄、涉暴、涉恐等内容的影音图像，应当予以删除以防止其传播。其二，除了网络平台自查，当利用“深度伪造”技术合成的影音图像被相关监管部门指出其应用场景违法或者被权利人主张侵权时，网络平台应当对该影音图像进行内容审查，若其涉嫌违法则应予以删除。如果上述“通知—删除”义务得以积极履行，就表明网络平台运营者对“深度伪造”技术被滥用的危害后果持排斥态度，从而不以刑法予以评价；反之，根据网络平台运营者的主观认知，分别以拒不履行信息网络安全管理义务罪、非法利用信息网络罪、帮助信息网络犯罪活动罪进行具体评价。

在将来反向识别技术发展得相对成熟时，对“深度伪造”技术的审查将成为网络平台的应然义务，届时应当将“发现—标识”“发现—删除”义务作为平台主体刑事责任的内容。对于通过“深度伪造”技术识别出的合成影音图像，网络平台运营者应当区分以下情形予以处置：首先，将内容违法的影音图像予以删除；其次，对属于“深度伪造”技术创制的影音图像未加标记的，通过添加水印、标识的方式提醒公众知悉；最后，对利用“深度伪造”技术创制的影音图像属于应用场景违法的，予以删除。2019 年 11 月国家互联网信息办公室等部门印发的《网络音视频信息服务管理规定》明确禁止“深度伪造”技术在新闻信息领域的应用，随着后期相关立法的出台，必将有“深度伪造”技术应用场景“负面清单”为网络平台附加更多审查义务，网络平台不仅要基于破解技术识别出“深度伪造”视频，还要对属于“负面清单”管制范围、违反“负面清单”应用场景的合成影音图像及时予以删除。通过不断强化网络平台的实质审查义务，有助于预防“深度伪造”技术被滥用的危害后果传播。

五、结语

在人类进入智能化时代的征途上，新兴科技的突破性发展带来欢喜与隐忧的情况是并存的。这也激发人们对新兴科技治理规则的探索热情。在抒发热情之余，更应秉持一份理性，既要通过规则的设定与落实，使新兴科技为我所用，又要避免科技发展走向失序的深渊，要让科技的归科技、法律的归法律。

注释

①㉙参见蔡淑敏：《从爆红到被整改，换脸 App“ZAO”降温》，《国际金融报》2019 年 9 月 9 日。②⑭参见李怀胜：《滥用个人生物识别信息的刑事制裁思路》，《政法论坛》2020 年第 4 期。③See Brandon J. *Terrifying High-tech Porn: Creepy ‘Deepfake’ Videos are on the Rise*, http://www.foxnews.com/tech/terrifying-high-tech-porn-creepy-deepfake-videos-are-on-the-rise.④参见苗争鸣：《可怕的“深度伪造”技术》，《世界知识》2019 年第 22 期。⑤See DJ PANGBURN. *You've been Warned: Full Body Deepfakes are the Next Step in AI-based Human Mimicry*, http://www.fastcompany.com/90407145/youve-been-warned-full-body-deepfakes-are-the-next-step-in-ai-based-human-mimicry.⑥参见揭书宜：《不雅视频将女明星换脸成主角售卖：涉多重违法》，澎湃新闻网，https://www.thepaper.cn/newsDetail_forward_4326898，2019 年 12 月 24 日。⑦See Douglas Harris. Deepfakes: False Pornography is Here and the Law cannot Protect You, *Duke L. & Tech. Review*, 2019, Vol.99.⑧参见明乐齐：《网络黑产犯罪的趋势与治理对策研究》，《山东警察学院学报》2019 年第 4 期。⑨参见刘宪权：《网络造谣、传谣行为刑法规制体系的构建与完善》，《法学家》2016 年第 6 期。⑩参见王帅：《“深度伪造”的法律风险与防范》，《中国城乡金融报》2019 年 12 月 20 日。⑪参见王禄生：《论“深度伪造”智能

技术的一体化规制》,《东方法学》2019年第6期。⑫See Micheal Chertoff and Anders Fogh Rasmussen. The Unhackable Election: What It Takes to Defend Democracy, 98 *Foreign Aff.*, 160(2019).⑬⑮See Mary Frost. *Clarke Introduces Bill to Combat High-tech Altered Videos*, https://brooklyneagle.com/articles/2019/06/14/clarke-introduces-bill-to-combat-high-tech-altered-videos/.⑯See *H.R.*3230-*Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of* 2019, http://www.congress.gov/bill/116th-congress/house-bill/3230.⑰See Waldman Ari Ezra. A Breach of Trust: Fighting Nonconsensual Pornography, *Iowa Law Review*, 2017, No.2.⑱See *Bill Title*: *Relating to the Creation of A Criminal Offense for Fabricating A Deceptive Video with Intent to Influence the Outcome of An Election*, https://legiscan.com/TX/text/SB751/id/1902830.⑲参见李桐佑:《美加州立法禁"深度伪造"视频》,环球网,https://news.ifeng.com/c/7qbaZk1sS4u,2020年1月22日。⑳参见商希雪:《生物特征识别信息商业应用的中国立场与制度进路》,《江西社会科学》2020年第2期。㉑《网络音视频信息服务管理规定》第11条规定:"网络音视频信息服务提供者和网络音视频信息服务使用者利用基于深度学习、虚拟现实等的新技术新应用制作、发布、传播非真实音视频信息的,应当以显著方式予以标识。网络音视频信息服务提供者和网络音视频信息服务使用者不得利用基于深度学习、虚拟现实等的新技术新应用制作、发布、传播虚假新闻信息。转载音视频新闻信息的,应当依法转载国家规定范围内的单位发布的音视频新闻信息。"㉒参见于志刚:《网络安全对公共安全、国家安全的嵌入态势和应对策略》,《法学论坛》2014年第6期。㉓曹建峰:《深度伪造技术的法律挑战及应对》,《信息安全与通信保密》2019年第10期。㉔参见于志刚:《网络空间中培训黑客技术行为的入罪化》,《云南大学学报》(法学版)2010年第1期。㉕参见于志刚:《中国网络犯罪的代际演变、刑法样本与理论贡献》,《社会科学文摘》2019年第5期。㉖参见孙万怀:《违法相对性理论的崩溃》,《政治与法律》2016年第3期。㉗参见石婧、常禹雨、祝梦迪:《人工智能"深度伪造"的治理模式比较研究》,《电子政务》2020年第5期。㉘参见于志刚:《中国互联网领域立法体系化建构的路径》,《理论视野》2016年第5期。㉚于志刚、吴尚聪:《我国网络犯罪发展及其立法、司法、理论应对的历史梳理》,《政治与法律》2018年第1期。㉛参见付微明:《个人生物识别信息的法律保护模式与中国选择》,《华东政法大学学报》2019年第6期。㉜《全国人民代表大会常务委员会关于加强网络信息保护的决定》第1条规定:"国家保护能够识别公民个人身份和涉及公民个人隐私的电子信息。"《网络安全法》第76条第5款规定:"个人信息,是指以电子或者其他方式记录的能够单独或者与其他信息结合识别自然人个人身份的各种信息,包括但不限于自然人的姓名、出生日期、身份证号码、个人生物识别信息、住址、电话号码等。"㉝《最高人民法院、最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》第1条指出,"能够单独或者与其他信息结合识别特定自然人身份"的信息属于公民个人信息。㉞于志刚:《关于"身份盗窃"行为的入罪化思考》,《北京联合大学学报》(人文社会科学版)2011年第1期。㉟笔者的设想是:将"身份冒用罪"作为我国《刑法》第280条第2款进行规定,同时将该条第1款即使用虚假身份证件、盗用身份证件罪的法定刑提升为"情节严重的,处3年以下有期徒刑、管制或拘役","情节特别严重的,处3年以上7年以下有期徒刑"两档,与该条规定的其他罪名相协调。"身份冒用罪"的具体内容可设计为:"冒用他人身份,情节严重的,处三年以下有期徒刑、拘役或管制,并处罚金;情节特别严重的,处三年以上七年以下有期徒刑,并处罚金。"㊱参见于志刚:《"双层社会"中传统刑法的适用空间》,《法学》2013年第10期。㊲See Jack M. Balkin. Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation, 51 *U.C.D.L. Review*, 2018, Vol.51.

责任编辑:邓　林

Construction of Criminal Law Regulation System of "Deep Forgery" Technology

Li Teng

Abstract: As an important field of artificial intelligence application, "deep forgery" technology can synthesize video and audio images with high precision by means of "deep learning" technology and "generative countermeasure network" model. The necessity of criminal law regulating the technology of "deep forgery" lies in the fact that it is easy to be abused; the legitimacy comes from the examination of the characteristics and hidden dangers of the technology, such as autonomy, convenience, and verisimilitude; and the feasibility is based on the legislative experience of regulating the technology in foreign countries. The criminal law of our country regulates the technology of "deep forgery" from the following standpoint: examine the risks and benefits of the technology objectively and neutrally, reasonably control the depth of criminal law intervention, and take the prevention function of criminal law as the goal. There are three specific regulatory paths: firstly, we may prevent the abuse of "deep forgery" technology from the front by strengthening the criminal protection of personal biometric information; secondly, we may strengthen the criminal protection of personal identity by adding the crime of identity fraud and regulate the rational use of "deep forgery" technology; thirdly, we may strengthen the criminal responsibility of network platform and prevent the spread of harmful consequences of "deep forgery" through the obligations fulfillment of "notice-delete", "discover-mark", and "discover-delete".

Key words: "deep forgery"; social harm; criminal law intervention; position of criminal law; regulation paths