·公共管理与社会治理·

AI换脸技术的应用风险及法律规制

□刘文涛

[四川轻化工大学 自贡 643002]

[摘 要] AI换脸源于深度伪造技术，涉及大量人脸信息，兼具数据属性。AI换脸的逼真度高、操作门槛低，获得了广泛的应用。但是该技术的滥用导致了较高的著作权、人格权等私权利侵犯风险、信息安全风险和犯罪防控风险。目前域外对AI换脸技术的法律规制主要存在分散式立法规制和统一立法规制两种模式。我国尚未建立系统化的规制体系，今后应当从以下方面进行完善：一是明确以区分应用场景前提下的合理使用为基本规制原则；二是构建事前、事中规制体系，明确研发者的技术伦理和制作者的标识义务、声明义务，保障信息主体的知情权和同意权，加强传播平台的内容审查义务；三是构建以数据、算法为核心的监管体制；四是严格民事法律责任追究，加大刑法制裁力度。

[关键词] AI换脸；深度伪造；数字治理；信息安全；个人信息保护

[中图分类号] D92; D93 [文献标识码] A [DOI] 10.14071/j.1008-8105(2023)-4006

AI换脸技术源于人工智能领域中“深度伪造”（Deepfake）技术，其原理是通过深度神经网络提取输入图像的深层信息，读取出其中隐含的深层特征来实现一些任务如风格迁移（Style Transfer）、人脸互换（Face-swap）等，是计算机视觉领域中近年来新兴的热门领域。该技术一经问世便风靡一时，并引发了激烈的争议。2017年11月，美国Reddit论坛上名为Deepfake的用户发布了多部利用AI换脸技术合成的女明星的色情视频片段，不久后遭论坛封杀。随后，该作者在Git Hub社区上开源了其技术代码，此后多个衍生代码也相继开源，AI换脸技术进入大规模传播时代。2019年，我国网络用户将女星朱茵在94版《射雕英雄传》电视剧中扮演的主角黄蓉替换成了女星杨幂，引发热议。此后，提供视频换脸服务的应用软件ZAO在我国迅速风靡，但仅3天后即被微信屏蔽访问。

AI换脸技术产生了一些负面影响。AI换脸的对象从最初的明星和影视片段扩大到新闻政要和周边朋友、同学、同事等普通人群；AI换脸技术的掌控者也从专业的技术人员扩展到普通APP用户；AI换脸的目的也从单纯的娱乐恶搞拓展到非法破解人脸识别系统、进行电信诈骗犯罪等。可以说，AI换脸技术在席卷全球，在带来可观的数字经济红利的同时也显示了对人类社会潜在的巨大破坏性，引发了广泛深刻的技术伦理担忧。鉴于此，笔者从技术逻辑切入，阐述AI换脸技术的特征及应用风险，分析现有法律规制模式之不足，在此基础上对我国的法律规制体系构建和完善方向进行探讨。

一、AI换脸的技术特征及应用现状

（一）AI换脸的技术特征

相比传统的利用Photoshop软件进行换脸操作（以下简称“PS换脸”）和计算机生成动画（Computer Generated Imagery, CGI）以及人脸识别技术相比，AI换脸技术有以下特征。

一是生成图像的高度欺骗性。PS换脸技术本质上是对图像的移动（俗称“抠图”），圈定于静态图片的处理。CGI是将一系列单个画面生成动画图像的过程。AI换脸步骤则更为复杂，涉及人脸信息

[收稿日期] 2023－06－05

[基金项目] 2023年度四川省人民检察院检察理论研究一般课题（CJ2023C30）.

[作者简介] 刘文涛，博士，四川轻化工大学法学院讲师.

[引用格式] 刘文涛. AI换脸技术的应用风险及法律规制[J]. 电子科技大学学报（社科版）, 2024, 26(2): 60-69. DOI: 10.14071/j.1008-8105(2023)-4006.

[Citation Format] LIU Wen-tao. Application risk and legal regulation of AI face-changing technology[J]. Journal of University of Electronic Science and Technology of China(Social Science Edition), 2024, 26(2): 60-69. DOI: 10.14071/j.1008-8105(2023)-4006.

的定位和侦测、提取特征值以及计算变换矩阵，生成图像更加逼真自然，实现了“颜值”和“表情”的分离。

二是应用目的不同导致的欺骗攻击。从技术手段而言，AI换脸和人脸识别并无本质差异。但人脸识别目的是在海量的图像中准确识别出特定人；而AI换脸目的则是“换脸”。因此，提高换脸效果的真实性和提高人脸识别中“反换脸”能力容易演化为“矛和盾”的关系，例如利用AI换脸技术“欺骗”“攻破”人脸识别系统进行不法活动。AI换脸技术欺骗攻击与人脸识别技术反欺骗攻击的算法较量将长期存在。

三是AI换脸技术臻于成熟。自Deep Fakes算法作者开源其技术代码后，来自全球的开发者持续对其优化迭代，十数个衍生代码（如faceswap，deepfacelab）相继开源。AI换脸技术已经实现了从正脸到侧脸、从静态图像到动态视频的技术突破，甚至可以实现实时视频聊天换脸。

（二）AI换脸的数据属性

AI换脸技术数据信息主要来源有：

一是图像的采集过程。摄像头采集到的每一帧人脸的图像、位置、表情等都可以用于人脸特征值的识别以及用于人脸替换所需的矩阵。而此类图像采集具有非接触性和非强制性特点，只要特定人乃至不特定人群在采集设备的拍摄范围内时，采集设备会自动搜索并拍摄用户的人脸图像。

二是APP收集的用户信息。当前APP通过强制授权、过度索权和超范围收集用户个人信息的现象普遍存在，用户授权APP访问设备储存以获得特定图像的同时，其他图像数据、地理位置信息、设备信息、通讯录甚至金融账号信息等可能被一并获取。

三是算法优化的驱动。AI换脸的基本原理决定了只有在拥有大量目标图片作为基础数据的情况下才能达到较好的效果。在AI换脸模型建立的早期，通常以明星等公众人物为目标，以方便在网络上收集大量照片作为训练数据。此后，要开发一套新的换脸模型或者对原有模型进行优化，还需要大量的数据案例，这就催生了对海量数据信息的需求。

（三）AI换脸技术的应用现状

一是应用方便、操作门槛低。由于技术代码开源，AI换脸对技术研发者和使用者的要求并不高。对研发者而言，换脸技术对硬件配置要求并不高，结构简单易于使用以及方便修改；对用户而言，只需借助APP上传本人照片即可完成“一键换脸”，操作实现“零门槛”。

二是应用场景不断扩展。整体上，AI换脸技术呈滥用状态：合成明星色情视频、合成政治家的演讲视频成为该技术最主要的应用场景，并开始与电信诈骗相结合成为犯罪工具，而影视、医疗美容和新闻传播等行业反而尚未实现大规模应用。

三是监管难度大。AI换脸的各类作品主要通过短视频平台、网络社交平台等进行发布，第三人转发、点赞等不受限制，传播速度极快，除非是明显涉嫌色情、恐怖、暴力等平台可以迅速识别的内容以外，被侵权人难以第一时间发现和制止。同时AI换脸技术不断更新，相关APP研发成本低、周期短，对监管部门的监管手段与监管方法提出了极大挑战。

二、AI换脸技术应用的现实风险

（一）AI换脸技术应用产生的私权利侵权风险

一是著作权侵权风险。具体包括：第一，APP提供给用户免费使用的模板素材可能并未获得版权方的使用许可，而且该类APP大多允许甚至鼓励用户个人上传素材，对影视综艺作品的信息网络传播权等造成了侵犯。以换脸软件“ZAO”为例，虽然版权声明中提到“除了特别声明是ZAO跟合作方进行版权合作的之外，均来源于ZAO用户自发的上传，ZAO不享有素材的商业版权”，也提到“禁止用户随意上传和使用素材，并通过人工审核环节尽力保证ZAO上的内容不侵犯相关权利人的合法权益”，但其规避法律风险的用意远高于实际意义。第二，经过AI换脸技术处理过的作品，上传该作品的用户与服务平台可能涉嫌共同侵犯原作品著作权人的修改权、保护作品完整权、改编权和翻译权。第三，用户的著作权可能遭受APP平台侵权。用户上传、制作的换脸图像、短视频、摄影集等，本身就具有成为著作权意义上“作品”的可能性，因而受《著作权法》的保护。根据“ZAO”软件用户协议，用户上传照片后，“ZAO”可以任意使用和修改用户或者其他肖像权利人的肖像，试图通过强制性协议的方式免除自身法律责任，也可能对用户作品的修改权、改编权、信息网络传播权等造成侵犯。其四，APP研发者的著作权也面临着被侵权的风险。例如各类破解版软件的存在，使得使用者只需用极低的价格便可获得原版软件收费较高的会员功能，或者研发出相近名称、类似图标的高仿APP。

二是人格权侵权风险。人脸图像不仅包含着公民个人的生物识别信息等敏感信息，更是公民展示

个人形象、进行对外交流的重要载体，系个人“颜面”之所在。而“换脸”极易对被换脸者的肖像权、名誉权、隐私权、姓名权等人格权造成损害乃至同时造成多重损害。常见的侵权场景有：第一，将明星、政要等公众人物的人脸图像转移、剪辑到其他视频片段中去。第二，在视频直播时将主播的脸换成明星的脸，以吸引流量；甚至人为合成明星代言广告等用以商业宣传。第三，利用合成图像对他人进行侮辱和诽谤。例如恶意捏造他人八卦新闻，故意贬损被换脸者的名誉，严重的可能构成侮辱罪、诽谤罪。第四，利用AI换脸技术制作和传播淫秽色情视频或图片，在构成制作、复制、出版、贩卖、传播淫秽物品牟利罪的同时严重侵犯被换脸者的肖像权和名誉权。

三是财产权利侵权风险，集中体现在对各类自动刷脸系统的破解上。如前所述，人脸识别技术很容易成为AI换脸技术欺骗攻击的目标，特别是基于2D算法的人脸识别技术系统更容易被欺骗攻击成功，例如用打印的照片即可破解自助取件设备[①]。而3D算法虽然安全性更高，但随着3D打印技术和3D面具技术的发展，不断有成功破解人脸识别支付系统的新闻被报道[②]。若用于攻击信息安全系数较低的网贷平台、超市购物以及手机解锁等，极易造成直接财产损失。

（二）AI换脸技术应用产生的信息安全风险

一是数据信息存储和泄露风险。指纹、人脸等生物信息具有唯一性和不可更改性，一旦发生泄漏（尤其是与个人身份信息捆绑泄露），会对公民隐私、财产安全等产生严重威胁。当前各类软件注册用户动辄百万乃至千万量级，而不少APP在申请访问本地存储权限的同时，还会额外获取用户手机号、IMEI、IMSI权限。然而，APP研发门槛低不等于信息安全程度高，APP收集的海量人脸信息数据以及其他用户信息数据只是存储于运营方或技术提供方的数据库中，至于服务器安全性能、数据是否脱敏、存储是否规范、数据库是否与合作者共享、数据管理是否规范等，外界无从得知。即使APP声明不会存储面部识别特征信息，但实际很难予以有效监管。

二是数据信息滥用风险。有报告显示，2018年中国计算机人脸识别市场规模为151.7亿元，预计2025年将破千亿元[③]。如此庞大的市场规模催生了人脸信息被过度收集和滥用的风险。2021年3•15晚会曝光的第一个问题即为部分商家利用店内摄像头非法采集和滥用人脸信息的事件。另据调查，在某些网络交易平台购买数千张人脸照片的成本还不到10块钱，均为真人生活照、自拍照[④]。人脸信息数据已经形成利益可观的地下灰色产业链。

三是国家信息安全维护风险。AI换脸技术可视为自动决策机制，可改造成一项“自动武器”，“自主性攻击”人脸识别系统等，由此引发国家信息安全问题。据国际数据公司（IDC）预测，2021年我国网络安全市场总体支出达到102.2亿美元，预测2020～2024年平均年复合增长率为16.8%[⑤]。另有报告显示，2020年我国捕获恶意程序样本数量超4 200万个，日均传播482万余次；境内受恶意程序攻击的IP地址约5 541万个，约占总数的14.2%；未脱敏展示公民个人信息事件107起，涉及未脱敏个人信息近10万条，累计监测发现个人信息非法售卖事件203起[⑥]。在此背景下，国家存储公民面部识别信息的基础设施和重要机构可能会成为域外恐怖势力和敌对势力攻击的目标之一。

（三）AI换脸技术应用产生的犯罪防控风险

目前，利用AI换脸技术进行犯罪的主要场景有：一是合成他人色情视频或者图片进行传播牟利，构成制作、复制、出版、贩卖、传播淫秽物品牟利罪、传播淫秽物品罪等罪名。若网络主播以及网络直播平台利用实时换脸功能进行在线网络色情表演的，还可能构成组织淫秽表演罪、侮辱罪和诽谤罪。二是与电信诈骗相结合，成为网络诈骗活动的新型工具。例如，对受害者用其亲朋好友进行换脸从而骗取财物，或者换脸为特定职业的特定人物进行诈骗。三是利用该技术伪造人脸识别认证（俗称“过人脸”）以及伪造身份证等虚假证件。四是与计算机网络犯罪的上下游关联犯罪一起成为地下网络黑、灰产业利益链条的一环。例如以AI换脸为幌子非法收集、窃取、贩卖公民个人信息，提供解锁手机设备、破解人脸识别系统、注册虚假身份信息等服务。

上述种种给犯罪防控带来了一系列不利影响：一是犯罪手段更加智能化。相比传统的电信诈骗，AI换脸、换声进行的“面对面”诈骗，迷惑性更强。二是犯罪行为更加隐蔽，给案件的侦破带来阻碍。该类犯罪行为发生在网络虚拟空间中，不存在传统意义的犯罪现场和犯罪痕迹，且绝大多数非法行为人都租用境外网络服务器，很难进行追踪、抓获。三是对现有的技术鉴定提出挑战。AI换脸的滥用对“反AI换脸”鉴定提出了挑战，虽然已有机构研发出鉴定分类器，但无法做到百分百识别。若公安司法机关无法证明视听资料、电子证据等系

换脸操作而来，将会对案件证据证明力带来极大冲击[1]。

三、域外AI换脸技术应用现有法律规制模式及与我国的差异

（一）美国分散式立法规制模式

美国是人脸识别技术和AI换脸技术的主要发源地和领跑者，也是最早对上述技术进行立法规制的先行者。规制模式上，美国采取了分散立法模式。联邦和州层面均已经发布多项法案对AI换脸技术进行专门立法规制。

在联邦层面，2019年，美国政府要求美国国土安全部定期发布有关深度造假技术评估报告的《深度伪造报告法案》（Deepfake Report Act of 2019）得以通过，旨在回应美国社会对深度伪造负面影响的关注[2]。该法案要求国家情报局定期向（美国）国会情报委员会提交有关机器操纵的媒体对（美国）国家安全的潜在影响以及外国政府实际或潜在地使用机器操纵的媒体来传播虚假信息或从事其他恶性活动的详细报告，并鼓励美国私营企业在政府支持下开发和部署针对Deepfakes技术的反制技术，以检测Deepfakes图片或视频的真伪性，并防止其传播[3]。在此背景下，美国军事情报部门发起和主导了系列研发项目计划，例如（美国）国防部资助的Media Forensics计划主要目的即为识别由机器学习算法生成的虚假图片和视频，其下属的国防高级研究计划局在全球范围内率先研发出了首款“反变脸”的AI刑侦检测工具，准确率达90%⑦。此外，美国参议院、众议院也分别针对深度伪造提出法案：2018年12月，美国参议院提出《2018年恶意伪造禁令法案》（Malicious Deep Fake Prohibition Act of 2018），对制作深度伪造内容引发犯罪和侵权行为的个人，以及明知内容为深度伪造还继续分发的社交媒体平台，进行罚款和对负责人进行长达两年的监禁。如果伪造内容煽动暴力、扰乱政府或选举，并造成严重后果，监禁将长达10年。2019年6月，美国众议院提出《深度伪造责任法案》（Deepfakes Accountability Act），要求任何创建深度伪造视频媒体文件的人，必须用“不可删除的数字水印以及文本描述”来说明该媒体文件是篡改或生成的，否则将属于犯罪行为。同月美国众议院还提出《2020财年Damon Paul Nelson和Matthew Young Pollard情报授权法案》，建议针对深度伪造技术的鉴别开展相关技术竞赛，以刺激鉴伪技术的研究和商业化。同时，美国两党议员分别在众议院、参议院同时提出《2019年深度伪造报告法案》（Deepfakes Report Act of 2019），该法案明确了“数字内容伪造”的定义，规定（美国）国土安全部定期发布深度伪造技术相关报告。

美国各州也针对AI换脸技术颁布了一系列法案或禁令。在弗吉尼亚州，2019年7月1日实施的“复仇色情法”将禁止范围扩大至经过深度伪造的内容，包括制作或操纵视频和使用机器学习制作的图像等，如果违反该法则，最高可判12个月监禁，罚款2 500美元[4]。在得克萨斯州，反深度欺诈法律已于2019年9月1日生效，虽然立法中没有明文出现“deepfake”等词，但Deepfake算法显然属于深度伪造（深度欺诈）。在加州，州长加文•纽瑟姆（Gavin Newsom）签署了两项法案，其中AB 730号法案规定禁止向公众传播利用AI换脸制作的虚假视频用以破坏选举，且该行为属于犯罪。另一项法案主要保护普通人群，任何加州居民均可对自己的肖像被通过AI换脸合成色情视频的行为提起诉讼[5]。在纽约州，州议会提交了编号A08155的法案，首次明确提出禁止用AI技术制作虚假视频[6]。该法案规定“个人角色（Persona）是个人财产，是可以自由转让和继承的”，并通过创建“数字副本”的方式实现在世和去世40年以内的人的姓名、肖像、照片、声音、签名等个人角色转让和继承的财产权利。

美国之所以采用分散式立法模式，除了与实行联邦制、种族主义等政治因素有关外，还与其技术发展程度、经济政策密切相关。美国是全球AI换脸技术最发达、普及程度最高但也是受侵害程度最深的国家，故而立法反应最为迅速，范围涵盖联邦和州层面，内容直指AI换脸或者深度伪造技术的应用场景限制，具有极强的针对性。但是就措施的严厉性而言，美国不如欧盟。原因在于，当前全球的数据信息都跨境流向美国，该国通过为自己的企业发展打造自由发展的环境来促进美国信息科技的发展壮大[7]。在该模式下，“深度伪造”巨大的数据流量给网络平台带来高额的利益回报，产生了极大的数据经济红利。换言之，美国只有在技术产生明显副作用情况下才予以规制，充分显示了其“先发展、后治理”的经济发展路径[8]。以人脸识别技术规制为例，美国对政府部门人脸识别技术应用进行严格限制，但对于企业等非政府部门则限制较为宽松。但是，在分散式立法模式下，存在立法步调难

以协调、各州内部利益诉求不一，以及数据流动带来的跨地域数据治理问题。

（二）欧盟统一立法模式

相比美国，欧盟不论是AI换脸技术普及程度还是受侵害程度均较低，因此并未采取针对性立法，而是将生成对抗网络（GAN）、深度伪造（Deepfake）等深度合成技术纳入作为隐私和数据保护法律框架的《通用数据保护条例（GDPR）》（以下简称GDPR）进行统一规制。GDPR旨在特别保护自然人的个人数据保护权，对各类数据公司对个人信息和敏感数据的处理进行了严格的限制性规定。GDPR虽然并未明文出现AI换脸或者Deepfake算法等词，但由于AI行业自有的数据属性，其对AI换脸技术依然有强力的规制。一方面，GDPR在全方位实现对数据主体的保护。GDPR共包含十章，其中第三章详细规定了数据主体享有的包括知情权、访问权、更正权、删除权（另称被遗忘权）、限制处理权、持续控制权、拒绝权和自动化个人决策权等权利；第四章规定了数据控制者和数据处理者的主要义务，包括数据系统保护和默认保护、记录数据处理活动、报告数据泄露事件、数据保护影响评估、事先咨询和设立数据保护官等；第八章规定了违反上述规定的严厉制裁措施和救济渠道。该条例自2018年5月25日正式生效后，不少科技公司和互联网公司因违反相关规定遭受诉讼和重罚，如法国对谷歌公司处以5 000万欧元的罚款，该条例因此被称为史上最强的数据保护条例。另一方面，GDPR在个人数据保护的基础上特别加强了对基因数据、生物识别数据等特殊种类的个人数据的保护力度。GDPR第4条对“个人数据”“基因数据”和“生物识别数据”的概念作出了明确界定，并将人脸图像和指纹识别数据明确规定为生物识别数据；第9条更是规定了除条例明确规定的特殊情形外，对特殊种类的个人数据处理以禁止为原则。此外，第32条还规定了应当统筹考虑最先进的技术处理过程的安全性，由此对AI换脸技术等产品的研发提出了未雨绸缪的要求。

除了通过GDPR进行规制以外，欧盟还尝试将AI换脸等深度伪造技术纳入不实信息和人工智能的规制框架。前者以2018年6月欧盟理事会通过的《欧盟不实信息实践准则》为代表；后者则以2021年4月欧盟委员会公布的《人工智能法（草案）》为代表，采取的依然是统一立法模式[9]。

相比美国法案主要针对AI技术应用场景的规制且对技术发展本身持较为宽容的态度，欧盟GDPR框架下的信息治理模式更加强调对公民隐私权的保护，更侧重数据信息的处理，也不对政府部门和非政府部门实行区别对待。总体来看，欧盟对人脸识别技术管控日趋严格，呈现出从弱风险预防到强风险预防的立场转变[10]。欧盟管控虽然较为严格，但一定程度上也可能影响新技术的进步。究其原因，除了欧盟认为公民隐私安全要优于科技带来的便利性的社会理念以外，还与欧洲目前信息技术产业的发展趋势密不可分。当下欧洲信息技术产业并不发达，通过将生物信息在内的各种数据处理和数据流通纳入GDPR予以严格规制，进而增强自身信息控制力，达到压制他国发展、保护本土企业的战略目标[7]。

除欧盟外，新加坡、英国、韩国等国家，均有适用于深度合成技术相关犯罪案件审理的法律法规⑧。

（三）我国相关法律规制体系建设及与域外的差异性

相比欧美，我国并未对人脸识别技术和AI换脸技术进行专门立法，尚未建立完善的法律规制体系，相关法律法规散见于《中华人民共和国民法典》《中华人民共和国个人信息保护法》《中华人民共和国刑法》等法律法规和行政法规等。具体而言：

一是《民法典》人格权编对公民肖像权、名誉权、隐私权等的保护。《民法典》第1019条规定，“任何组织或者个人不得丑化、污损，或者利用技术手段伪造等方式侵害他人的肖像权”，成为民法层面约束AI换脸行为的直接条款。此外，《民法典》第110条原则性规定个人信息受法律保护，第1034~1038条进一步予以细化，从而奠定了民事领域个人信息保护的立法基础。

二是《个人信息保护法》较为详细地规定了个人信息处理规则、个人信息者享有的权利、信息处理者的义务以及监管部门和法律责任等，从而构成了我国个人信息保护的法律框架和制度指引。该法第28条将生物识别信息纳入敏感个人信息并对个人敏感信息的处理予以从严规定，第29条在一般知情同意规则的基础上还要求人脸识别信息主体作出单独同意。

三是《刑法》相关规定。2015年《刑法修正案（九）》确立的“侵犯公民个人信息罪”以及《关于办理非法利用信息网络、帮助信息网络犯罪活动等刑事案件适用法律若干问题的解释》和《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》等显示了我国对侵犯个人信息行为的打击力度不断增强⑨。除侵犯公民个人信息犯罪以外，利用AI换脸技术的行为还可能触犯传播淫秽物品罪、故意传播虚假信息罪、诈骗罪、帮助信息网络犯罪活动等关联犯罪。

四是相关行政法规。2019年国家网信办、文旅部和广电总局发布的《网络音视频信息服务管理规定》第11条对深度学习、虚拟现实等的新技术新应用作出了规定，这也是我国当下法律法规体系中最为直接的规定。根据该条，通过AI换脸技术制作的视频“应当以显著方式予以标识”，且不得利用该技术“制作、发布、传播虚假新闻信息”。2020年国家市场监督管理总局、国家标准化管理委员会发布的《信息安全技术个人信息安全规范》对APP收集个人生物识别信息提出了明确要求，除人脸识别信息单独同意规则外，还规定了个人生物识别信息要与个人身份信息分开存储，且原则上不应存储原始个人生物识别信息。2022年1月，国家网信办发布了《互联网信息服务深度合成管理规定（征求意见稿）》，其中第2条、第11条、第12条、第14条对于人脸生成、人脸替换、人脸操纵等予以了明确界定，该《规定》正式实施后将成为我国AI换脸等深度合成技术的基本法律框架。此外，2019年国家互联网信息办公室发布的《网络信息内容生态治理规定》第23条亦对利用深度学习等技术生成的信息内容、信息传播等作了原则性规定。

从规制模式上看，我国对AI换脸技术的法律规制更接近欧盟的统一立法模式，《个人信息保护法》是基础框架。但相比欧盟统一禁止之原则，我国对AI行业持“包容审慎”监管原则，对相关科技发展采取了类似美国“先发展后规制”的路径，只要不触碰安全底线都支持发展。这也导致该领域立法可能滞后于技术发展，对AI换脸技术的管控存在局限：第一，我国尚未建立人工智能安全的统一标准，且作为国家标准的《信息安全技术个人信息安全规范》仅为推荐性标准，并不具备强制约束力。第二，民法、刑法和行政法规等缺乏有效衔接，不同法律法规之间存在定义、用语的不一致甚至潜在冲突，难以予以一体化规制[11]。第三，法律后果规定不明确。AI换脸既可能侵犯公民财产权，也可能侵犯人格权和知识产权，严重时可能涉嫌犯罪甚至同时侵犯上述权益；涉及的侵权行为人除了图像视频合成者（即技术使用者）以外，还可能涉及技术研发者、服务提供者、网络平台和平台传播者，侵权主体的认定、侵权责任划分以及证明责任的承担均缺乏统一、明确的规定。

四、我国AI换脸技术法律规制的路径

（一）规制原则：区分应用场景前提下的合理使用

目前我国对AI换脸等深度伪造技术的治理存在加强管制和放松管制两种倾向。加强管制者认为，深度伪造技术自诞生之日起便自带威胁，且发展趋势越来越背离人类社会道德标准，对个人权利、社会秩序和国家安全已经产生了现实危害⑩。放松管制者认为，技术是中立的，可以通过正确引导降低其应用风险，但不能因噎废食、强行阻断该技术的发展创新⑪。笔者认为：一方面，我国正处于第四次工业革命的历史机遇期，2017年7月国务院印发的《新一代人工智能发展规划》明确指出到2030年我国人工智能理论、技术与应用总体要达到世界领先水平，成为世界主要人工智能创新中心。人工智能已经上升到国家战略层面，且相关技术产业正处于快速发展阶段。在此背景下，若采取严格的管制模式，势必会对科技创新、数字经济发展带来负面影响。另一方面，“深度伪造技术滥用有屡禁不止之势，甚至演变成网民的娱乐狂欢”[12]。AI换脸更是直接针对公民脸部信息，对于隐私权、人格权的侵犯远甚于一般的“刷脸”行为，若被违法犯罪行为所利用，将严重威胁社会稳定，必须予以严格规制。

基于此，应当以区分应用场景前提下的合理使用作为AI换脸技术的规制原则。具体包含两层含义：第一，对技术的规制应当区分不同应用场景，以应用场景为基本出发点选择加强管制还是放松管制更加符合现实需求。第二，虽然技术本身是中立的，但是技术的使用者总是带有一定的目的。因此，还需确定不同场景下合理使用的范围和标准。借鉴我国《著作权法》第22条可将AI换脸技术的合理使用限定在科学研究、艺术再现、课堂教学、新闻报道或经授权的商业使用等范围，同时区分不同场景合理使用之限度。例如，影视公司等利用AI换脸技术优化画面视觉效果或者对某一画面进行技术性复现的，对其放松管制有利于促进行业发展。而对于政治、宗教人物等进行换脸或者利用换脸技术进行新闻报道的，应当予以严格管制。唯有此，才能在不影响科技进步的同时最大程度降低技术滥用风险。

（二）构建事前–事中规制体系

首先是建立事前规制路径，主要包括：

一是规范研发者的技术伦理和信息素养。在以往，科技指向的是客体世界，而互联网和计算机等技术指向人类本身，通过对人类自身能力的扩展和延伸建立起“人–机”关系，将人类社会逐步演变为现实和虚拟交织、万物互联时代，进而影响整个人类社会价值观、世界观。AI换脸技术对网络安全、个人隐私等有巨大潜在威胁，甚至其算法本身就是建立在侵犯女明星人格隐私基础之上，为了制作淫秽色情视频而创建和优化的。换言之，AI换脸技术的诞生之初就已经突破了技术伦理底线。虽然其技术代码早已开源，对监管等带来障碍，但距离该技术在影视、医疗美容等行业大规模应用尚有时日，因此依然可以从源头上对技术研发者的技术伦理加以规范，引导其回到“向善”的技术伦理轨道。此外，还需要提高技术行业专业人员、产品使用者和社会民众的信息素养，通过自律方式提高对自己或他人的信息安全、人格隐私的重视和尊重。

二是强化制作者的标识义务和声明义务。我国《网络音视频信息服务管理规定》第11条仅笼统规定AI换脸视频“应当以显著方式予以标识”，但对如何标识、何为显著以及违反后果语焉不详。对此，要在落实视频制作者的标识义务的基础上进一步规定制作者的声明义务。所谓标识义务，是指AI换脸视频的制作者要在作品的显著位置和足够长的时间以水印、特殊标记等方式对作品进行标识，足以让一般观看者理解鉴别。由于个人用户发布换脸视频主要用于“吸粉”“流量变现”，并不能充分理解和遵守标识义务，可考虑赋予换脸APP强制标识义务，即凡是通过换脸APP制作的视频没有进行标识或者标识不准确的，一律由平台进行统一标识或者不予审核通过。

所谓声明义务，是指制作者在传播平台上传作品时要对视频的换脸属性进行声明，并根据对原视频图像的改变程度、视频时长、制作用途、传播次数等分别予以不同强度的声明义务，如书面方式还是口头方式、审批方式还是备案方式等。此外，标识义务和声明义务同时适用于转发者、剪辑者等二次传播者，不得对原视频的换脸标志、声明等进行删减、隐藏处理。如果制作者、传播者等未遵守上述规定，造成侵权的，应当承担民事赔偿责任，推定为“有过错”；触犯刑法的，主观上可推定为“明知”。

其次是建立事中规制路径，主要包括：

一是保障信息主体的知情同意权，最重要的是推动AI换脸技术服务提供者在收集个人信息时严格遵守知情同意规则。虽然我国《个人信息保护法》第21~23、25~27、29、31条规定了个人信息、个人敏感信息的处理必须取得个人同意或者个人单独同意，但APP通过隐私条款等方式迫使用户“强制同意”“概括同意”，无法保证自愿性。甚至在实践中，告知同意规则反而成为数据企业收集个人信息的“万能法则”，要么赋予用户信息自决权，但不遵循或者突破告知同意规则；要么“貌似遵守”这一规则，实际上故意曲解，从而达到不法收集用户信息之目的[13]。AI换脸涉及大量个人敏感信息，必须构建更为严格的可操作的知情同意规则，具体包括：第一，充分履行告知义务，保障信息主体的知情权。APP等平台必须在用户使用换脸服务前详细、准确介绍人脸信息收集处理的场景、目的、方式和限度，保证用户知情是充分、自愿的。第二，将隐私条款中的“概括同意”方式转变为“动态同意”方式，设置不同应用场景的分级同意规则，并允许用户个人进行选择和变换，提高信息主体在信息处理过程中的参与度[14]。同时规定除法律明确规定的场景外，作出同意的方式应当采取明示同意而非默示同意方式，禁止默示推定，以保障用户的个人信息自决权。第三，考虑到AI换脸视频一旦上传网络平台便会迅速传播，对被侵害人极为不利，可借鉴欧盟GDPR设立“擦除权”（被遗忘权），以限制APP对生成的换脸视频作品进行不当扩散传播。

二是加强传播平台的内容审查义务。在域外，出于社会伦理道德压力和指责，众多社交网络平台一开始就对AI换脸视频进行抵制：一方面禁止技术交流讨论，如美国Reddit社区删除了关于AI换脸技术的讨论板块，其GitHub开源代码也被清除；GIF平台Gfycat也移除了相关内容；Discord论坛对讨论制作换脸色情电影的服务器进行封锁处理。另一方面对AI换脸技术生成的视频内容进行严格审查，例如Twitter、Facebook等域外社交媒体平台对检测到的虚假视频根据具体情形作出警告、进行标记、限制推荐和删除等监管措施；即使是全球最大色情影片平台Pornhub也明确表示禁止上传复仇色情报复以及造假的内容。我国《网络安全法》规定网络经营者对其用户发布的信息负有监督管理责任，作为借鉴，今后应当从以下几个方面加强传播平台的内容审查义务：第一，网络平台（尤其是短视频传播平台）应当对换脸视频内容进行严格审查，除非明确获得权利人许可或符合法律规定的情

形外不得上传，并对AI换脸标识的显著性、视频内容是否适合推荐浏览等进行评判和跟踪。如果接到被侵权人举报申诉的，应当第一时间进行核实，采取警告、限制转发、删除等措施，严格落实《民法典》第1 195条规定的“通知–删除”义务。第二，域外不少互联网平台都推出针对AI换脸视频等虚假信息识别技术[15]，而我国网民规模、互联网普及率和社交网络市场规模已位列全球前列，提高网络平台换脸视频的识别能力是大势所趋。为此，可以考虑赋予网络平台“发现–标记”“发现–删除”义务。网络平台“需要在传统内容形式审查的基础之上更进一步，实现对伪造视频的技术审查”[16]。除了对制作者标识的换脸视频进行审查外，还要主动对其他视频进行甄别检测，一旦发现被检测视频具有深度伪造记录或者存在深度伪造嫌疑的，应当主动进行标记，并采取警告、删除视频、封禁账号等措施。

（三）构建“数据–算法”为核心的监管机制

实际上，上述规制措施只是尽可能降低AI换脸技术应用层面产生的风险，而要根本性解决这一问题，则要回归到对技术本身的监管，即所谓的“用AI打败AI”。本质上，AI换脸等深度伪造技术是基于算法的大数据分析，需要从数据和算法两方面予以监管。

一是数据安全监管。在大数据时代，数据安全更加强调以数据为着眼点的全过程安全，包括数据安全和数据防护安全。前者以保护数据完整为主要目的；后者主要保护数据载体和功能安全。目前，我国互联网网络安全面临严峻考验，人脸数据不仅关乎公民隐私权、财产权保障，还关乎社会秩序稳定和国家信息安全。因此，监管者要重点审查两方面的内容：一方面，AI换脸APP的研发者和服务提供者要能保证数据本身安全，防止数据丢失、泄露、窃取和滥用。我国《数据安全法》第21条原则性确立了数据分级分类保护制度，人脸信息作为个人敏感信息，应当纳入重要数据具体目录进行重点保护；涉及海量人脸信息数据或者重要人脸信息数据的，可以纳入“国家核心数据”实行最严格的管控制度。另一方面，数据处理者要保证信息系统的安全性，防止出现设备故障、程序缺陷、病毒或黑客攻击等问题。可考虑建立企业数据安全治理能力评估体系和标准，保证最低程度的系统安全[17]。基于人脸信息的敏感性，对相关企业应当实行强制性的数据安全认证，监管部门定期开展数据安全检查、数据安全审查和网络安全审查，加强数据治理。AI换脸技术研发者、服务提供者因未达到数据安全要求而造成用户信息泄露等损失的，应当承担赔偿责任。

二是加强算法监管。作为人工智能的基础，算法天然具有不透明性，其决策过程和结果是否公平、正当也难以受到监督[18]。算法安全是人工智能安全治理中的技术保障，2022年3月1日实施的《互联网信息服务算法推荐管理规定》为我国算法推荐服务监管提供了指引，在此基础上进一步细化：第一，依据该《规定》第23条规定的“算法分级分类安全管理制度”，将作为AI换脸的基础算法或者可能用于AI换脸的Deepfake等算法纳入“高风险”安全等级进行规制。第二，对此类算法实行严格的事前备案、安全评估机制，而非“一事一议”事后监管方式，确保监管的及时性。对用户规模大、数据量高的算法还可以建立必要的试运行机制，避免出现算法歧视等问题，如美国加州的《人脸识别技术法案》、美国华盛顿州的《人脸识别服务法案》中也有类似规定[10]。第三，网信办、市场监督管理局等监管部门可与公安部门一道建立算法检测系统，对算法的数据使用、应用场景、影响效果等进行日常监测和风险预警。

（四）严格法律责任

民事责任方面。第一，明确人脸信息的财产属性并予以保护。目前我国对人脸信息的法律定性尤其是财产属性存在争论，更多是将个人信息看做人格权益进行保护[11]。而欧美对其财产法益定位明确，例如美国纽约州议会提交的法案，明确规定个人角色（Persona）是个人财产，并通过创建“数字副本”的方式进行保护。未来我国民法应当进一步明确包括人脸信息在内的个人信息具有财产属性，实行人格和财产的双重保护。第二，明确侵权责任主体及责任划分标准。当前我国法律规定信息网络环境下的侵权主体包括网络用户和网络服务提供者两类。然而，一方面，利用AI换脸技术实施侵权行为的人往往具有匿名性甚至位于境外，这给追踪直接侵权人带来了障碍；另一方面，换脸视频可通过换脸APP或者其他用户分享、传播至其他网络社交平台，存在多个网络服务提供者，侵权主体的认定和责任划分可能存在争议。笔者认为，一方面，换脸视频的制作者、发布者作为直接侵权人承担侵权责任自不待言。另一方面，换脸APP兼具视频制作平台和传播平台属性，要承担较高标准的内容审核义务和技术防范义务，因此换脸视频APP应当与直接侵权人承担连带责任。至于第三方视频平

台是否作为侵权主体以及承担责任之大小，应当根据是否履行“通知–删除”“发现–标记”“发现–删除”等义务的具体情况认定。第三，设立惩罚性赔偿机制。目前，我国公民名誉权主要通过精神损害赔偿进行救济，法人声誉的主要通过《中华人民共和国反不正当竞争法》进行救济，但是上述救济方式存在追究成本过高而侵权人承担后果不严重的问题。笔者认为，AI换脸技术的滥用归根到底是为了追求数字经济效益，可通过设立惩罚性赔偿机制，以侵权损失额或违约所得额的数倍对被侵权人加以赔偿，以达到严惩和威慑的目的[3]。

刑事制裁方面。第一，虽然AI换脸技术属于“新科学技术”范畴，也对法律适用提出了新挑战，但总体上可以通过“解释论”方法在现有刑法框架下进行规制，对滥用AI换脸技术的行为无需通过设立新的罪名进行规制，避免刑法体系的碎片化和不当扩张。第二，当前我国《刑法》以及《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》《关于办理利用信息网络实施诽谤等刑事案件适用法律若干问题的解释》等规定了侵犯公民个人信息和转发、浏览诽谤信息入罪和“情节严重”的数量标准，但对于生物识别信息并未明确列举，故刑罚适用上存在疑问[19]。笔者认为：一方面，不论将人脸信息理解为上述规定条文中的“兜底性规定”抑或是“等”，并不影响利用AI换脸行为的入罪。第二，出于对人脸信息等个人敏感信息的超强保护以及深度伪造技术的巨大破坏性，今后应当明确：对通过AI换脸技术合成虚假信息进行诽谤等犯罪的，应当直接认定为“情节严重”，主观上推定为“明知”，无须机械地以浏览量、转发量作为入罪标准，同时将传播范围和换脸视频的逼真度等作为量刑的酌定情节加以考量[20]。

注释

① 相关案例可参见新浪网.丰巢刷脸取件被小学生“破解”刷脸支付还安全吗？[EB/OL]. [2022-03-15]. https://finance.sina.cn/bank/yhgd/2019-10-17/detail-iicezzrr2864642.d.html.

② 相关案例可参见贤集网.美国AI公司3D面具骗过包括多家人脸识别系统成功支付[EB/OL]. [2022-03-15]. https://www.xianjichina.com/special/detail_435812.html.

③ 参见中研产业研究网《2019～2025年人脸识别行业风险投资态势及投融资策略指引报告》[EB/OL].[2022-03-20]. https://it.chinairn.com/news/20191216/115627739.html.

④ 参见搜狐网.“AI换脸”骗过人脸识别！我们的人脸信息还安全吗？[EB/OL]. [2022-03-20]. https://www.sohu.com/a/427897825_120057347.

⑤ 参见CIT论坛.IDC 发布最新版全球网络安全支出指南，中国以16.8%的高增速领跑全球[EB/OL]. [2022-04-18]. http://www.ctiforum.com/news/guonei/584156.html.

⑥ 参见国家互联网应急中心发布的《2020年中国互联网网络安全报告》。

⑦ 参见新浪网.AI换脸终结者问世 美国防部推高精度“反换脸”工具[EB/OL]. [2022-03-25]. https://tech.sina.com.cn/csj/2018-08-08/doc-ihhkuskt7977099.shtml.

⑧ 例如2019年5月，新加坡议会通过的《防止网络虚假信息和网络操纵法案》（Protection for Online Falseas and Manipment Act），使政府有权要求个人或网络平台更正或撤下对公共利益造成负面影响的虚假内容，法案适用于利用深度伪造技术制作的虚假音视频。

⑨ 《刑法修正案(九)》将原《刑法》“出售、非法提供公民个人信息罪”和“非法获取公民个人信息罪”统一整合为”“侵犯公民个人信息罪”，扩大了犯罪主体和侵犯个人信息行为的范围。

⑩ 这一观点的代表性论文有宋凡：《民法典》时代下“深度伪造”科技风险与应对模式，载《中国电信业》2020年第10期；熊波：《“深度伪造”的扩张化刑事治理风险及其限度》，载《安徽大学学报（哲学社会科学版）》2020年第6期。

⑪ 这一观点的代表性论文有参见江凯帆：《智能“换脸”技术的侵权风险及其法律规制研究》，载《中共南京市委党校学报》2020年第6期；蔡士林：《“深度伪造”的技术逻辑与法律变革》，载《政法论丛》2020年第3期等。

参考文献

[1] 李天琦, 刘鑫. 深度伪造技术的证据风险与规制路径[J]. 证据科学, 2022, 30(1): 70-82.

[2] 谢进川, 唐恩思. 深度伪造的社会伤害与治理争议[J]. 新闻与写作, 2023(04): 96-105.

[3] 曹越. AI换脸技术产生的危害与应对措施[J]. 南海法学, 2020, 4(4): 69-77.

[4] 新京报. 好玩的AI换脸, 为何美国智库认定威胁国家安全[EB/OL]. [2022-03-25]. https://baijiahao.baidu.com/s?id=1643552431570255259&wfr=spider&for=pc.

[5] 加州新法案: 未经当事人同意，不得用AI换脸技术制作色情视频[EB/OL]. [2022-03-25]. https://ishare.ifeng.com/c/s/7qdCNIGndFR.

[6] 情色边缘游走, 夹缝中生存, 这个AI“换脸术”可能要被官方禁止[EB/OL]. [2018-06-18]. https://www.toutiao.com/article/6568268676615635464/?wid=1698241822319.

[7] 文铭, 刘博. 人脸识别技术应用中的法律规制研究[J]. 科技与法律, 2020(4): 77-85.

[8] 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[J]. 电子政务, 2020(5): 69-79.

[9] 张涛. 后真相时代深度伪造的法律风险及其规制[J]. 电子政务, 2020(4): 91-101.

[10] 张新宝, 葛鑫. 人脸识别法律规制的利益衡量与制度构建[J]. 湖湘法学评论, 2021, 1(1): 36-51.

[11] 王鑫媛. 人脸识别技术应用的风险与法律规制[J]. 科技与法律(中英文), 2021(5): 93-101.

[12] 孙宇, 闫雯静, 罗玮琳. 政府规制深度伪造技术应用的系统性综述及批判性反思[J]. 电子政务, 2022(1): 77-87.

[13] 林凌. 人脸识别信息保护中的“告知同意”与“数据利用”规则[J]. 当代传播, 2022(1): 108-112.

[14] 石佳友, 刘思齐. 人脸识别技术中的个人信息保护——兼论动态同意模式的建构[J]. 财经法学, 2021(2): 60-78.

[15] 智东西. 美国刮起“反AI换脸”热潮, 创企各出奇招, 联合对抗deepfakes[EB/OL]. [2019-07-29]. https://ishare.ifeng.com/c/s/7qdCNIGndFR.

[16] 王禄生. 论“深度伪造”智能技术的一体化规制[J]. 东方法学, 2019(6): 58-68.

[17] 龚诗然, 刘雪花. 数据安全治理现状研究与分析[J]. 信息通信技术与政策, 2022(2): 42-46.

[18] 刘静怡, 颜厥安, 吴从周, 等. 人工智慧相关法律议题刍议[M]. 台北: 元照出版社, 2018: 14.

[19] 王德政. 针对生物识别信息的刑法保护: 现实境遇与完善路径——以四川“人脸识别案”为切入点[J]. 重庆大学学报(社会科学版), 2021, 27(2): 133-143.

[20] 姜瀛. 人工智能“深度伪造”技术风险刑法规制的向度与限度[J]. 南京社会科学, 2021(9): 101-109.

Application Risk and Legal Regulation of AI Face-Changing Technology

LIU Wen-tao

(Sichuan University of Science & Engineering　Zigong　643002　China)

Abstract　AI face-changing is derived from deep forgery technology, which involves a large amount of facial information and has data attributes. AI face-changing has high fidelity and low operating threshold, and has been widely used. However, the abuse of this technology has led to higher risks of infringement of private rights such as copyright and personality rights, information security risks, and crime prevention and control risks. Currently, there are two main models of overseas legal regulation of AI face-changing technology: decentralized legislative regulation and unified legislative regulation. Our country has not yet established a systematic regulatory system. It should be improved from the following aspects in the future: First, it is clear that reasonable use under the premise of distinguishing application scenarios is the basic regulatory principle. Second, it is to build a preevent and in-process regulatory system to clarify the technical ethics of developers. and producers' identification and statement obligations, protect the information subject's right to know and consent, and strengthen the content review obligations of communication platforms. The third is to build a regulatory system with data and algorithms as the core; the fourth is to strictly pursue civil legal liability and strengthen Strong criminal law sanctions.

Key words　AI face-changing; deep forgery; digital governance; information security; personal information protection

编　辑　邓婧