

AI换脸行为的刑法规制路径

文/吕蓉蓉 郑雨欣

内容提要 AI换脸技术作为深度伪造技术的典型代表,在释放创意潜能、推动产业革新的同时,亦被滥用于诈骗盗窃、商业侵权、淫秽传播、侮辱诽谤等违法犯罪活动中,严重侵害了个人法益与公共法益。当前刑法在规制AI换脸行为滥用方面,存在个人法益与公共法益保护失衡、非法使用行为规制缺失、网络服务提供者责任边缘化等问题。为解决上述困境,刑法应秉持谦抑性的立场,扮演好“二次法”的角色,从实务层面积极探索涵盖入罪标准扩充、罪数关系厘清及网络服务提供者责任形态细化的具体治罪方案,以此为司法实践提供兼顾法益全面保护与技术风险防控的刑法规制路径。

关键词 AI换脸 深度伪造 刑法规制 侵犯公民个人信息罪 网络服务提供者责任

随着生成式人工智能和深度合成技术的深入发展,AI已从专业应用走向大众普及,并作为一种变革性力量,在经济、文化等领域深刻重塑生产模式、释放发展新动能。AI换脸技术以其便捷性、娱乐性和创造性,为影视制作、社交娱乐等领域注入了新的活力,展现出技术本身的中立价值。然而,中立的技术一旦被滥用,便会引发侵权与监管的难题,甚至被用于电信诈骗、淫秽物品制作、伪造证据等违法犯罪活动。从全国人大代表张伯礼院士指出自己被冒用形象讲述虚假内容,到演员温某某分身直播带货的奇特现象,AI换脸技术正被不法分子滥用,严重侵害公民权益、扰乱社会秩序和损害社会信任体系。在此背景下,如何在刑事法律层面规制AI换脸乱象,成为亟待解决的问题。即刑法应以何种角色介入AI换脸技术的治理?面对AI换脸技术滥用所引发的违法犯罪行为,法律规范该采取何种路径全面精准地保护各类法益?以及网络服务提供者责任的边界又该如何界定?只有回应好上述议题,才能通过构

建兼顾技术创新与法益保护的刑事治理体系,推动人工智能技术健康有序地发展。

一、AI换脸行为的犯罪样态和侵害法益分析

(一) AI换脸行为的犯罪样态

AI换脸技术是一种基于深度学习的人脸图像识别伪造技术,能够借助生成对抗网络(GAN)实现对人物肖像的超逼真模拟与场景重构。因技术发展日益成熟且操作门槛降低,多数AI伪造工具已实现高度流程化和自动化,用户无需理解底层算法,仅需准备人脸图像或人声样本即可快速合成逼真内容。这种较低的技术门槛,在催生广泛娱乐应用的同时,也导致了灰色产业化的苗头,本文通过对近年来司法案例的梳理,将AI换脸不法行为归纳为以下几种常见典型行为模式:

1. 侵犯犯罪型模式

在该类行为模式中,行为人以非法占有为目的,利用AI换脸技术伪造他人身份或生成动态人脸视频,进而实施诈

骗、盗窃等侵犯公民财产的犯罪行为。其行为方式包括但不限于以下两类：其一，冒充特定关系人实施欺诈。行为人通过AI技术实时合成被害人亲友、同事等特定关系人的面部影像，并发起即时通话，以突发疾病、紧急转账等虚假事由骗取被害人信任，诱使其处分财产。^①该类行为利用了人际信任机制与视觉验证的心理依赖。其二，非法突破身份认证系统，转移他人账户资金。行为人通过非法获取的公民个人信息与AI生成的动态人脸视频，绕过金融机构、支付平台等所设置的人脸识别验证机制，非法登录或者注册他人账户并完成身份认证，进而实施资金转移^②或盗刷账户^③。此外，实践中还出现利用AI换脸技术冒充军人、警察或者其他公职人员实施招摇撞骗的行为，相较于传统方式更具欺骗性和隐蔽性，不仅侵害财产权益，更损害国家机关形象和公信力。^④

2. 商业利用型模式

在该类行为模式中，行为人以营利为目的，未经权利人授权，擅自使用他人肖像或者作品，通过AI换脸技术进行篡改、合成，并将其应用于电商直播、广告推广或其他营利性活动，构成对权利人人格权与知识产权等权益的侵害。其行为方式主要包括以下两种情形：其一，盗用公众人物形象进行虚假直播或者商品推广。例如，演员温某某形象被多家电商盗用于不同带货直播间。该行为可能涉嫌侵犯公民个人信息罪、诈骗罪以及销售伪劣产品罪。其二，平台方擅自将他人肖像或作品设为付费换脸模板。AI换脸程序运营者在未授权情况下，将网红或者明星的肖像视频上传至其运营的换脸软件中，作为AI换脸模板供用户付费使用，此行为不仅侵犯了权利人的肖像权，还可能构成侵害他人作品的信息网络传播权，^⑤违法所得数额较大或有其他严重情节，则构成侵犯著作权罪。

3. 淫秽传播型模式

该类行为模式是AI换脸技术滥用中最常见的不法应

用行为。行为人利用AI换脸技术，将他人面部特征非法合成到淫秽视频或图片中，进行网络传播甚至以此牟利。一是技术开发者制作具有一键脱衣、虚拟换脸等功能的软件或算法模型。^⑥二是内容制作者直接从事淫秽视频的伪造与传播，如虞某某一案中，行为人通过互联网学习掌握了AI换脸软件的使用方法，并从互联网等渠道收集百余人的脸型信息，伪造生成大量不雅视频、图片，并通过收入群费、销售软件、提供定制服务等方式盈利。^⑦这种行为严重侵害了被换脸者的多项人格权益，破坏社会公序良俗，涉嫌制作、传播淫秽物品牟利罪。

4. 侮辱诽谤型模式

在该类行为模式中，行为人利用AI换脸技术制作虚假视频或图像，并公开传播，旨在损害特定对象的社会评价与人格尊严。一是以恶搞鬼畜为名的侮辱行为，通过将他人面部合成至具有丑化、诅咒、贬低意味的视频情景中进行人格羞辱。二是以合成造假为手段的诽谤行为，伪造他人参与违法活动、发表不当言论的虚假视频，向社会公众散布虚假信息。“雷军语音包”和印度总理莫迪的跳舞视频即属此类典型行为。此类行为的频发会加剧网络信息的复杂性和不确定性，损害公众的知情权和媒体新闻传播公信力，反映出AI技术滥用对人格权保护与公共言论环境所带来的新型挑战。

（二）AI换脸行为的多元法益侵害

AI换脸行为的危害性不仅体现在技术手段的隐蔽性和欺骗性，与传统犯罪相比，其侵害的法益具有多元性，一个犯罪行为可能同时侵害个人法益与公共法益，这也构成了进行刑法规制的正当性基础。

1. 个人法益的侵害

AI换脸行为首先侵害的就是被换脸公民的个人法益，主要包括：第一，对人格权的侵害。民法典明确规定，自然人

① 张天培、苏滨：“以假乱真的AI诈骗，如何防范？”，载 <https://www.chinacourt.cn/article/detail/2025/04/id/8780246.shtml>，2025年12月25日访问。

② “男子用‘AI换脸’登录23人账户，盗刷信用卡获利4年半！”，载 <https://news.cctv.com/2025/07/20/ARTIiOURgyYkuxEbw-257c1A250720.shtml>，2025年12月24日访问。

③ 参见重庆市渝中区人民法院（2024）渝0103刑初395号刑事判决书。

④ “AI换脸织就的诈骗迷网”，载 https://www.spp.gov.cn/spp/zdgg/202508/t20250826_704627.shtml，2025年8月26日访问。

⑤ 参见上海市金山区人民法院（2022）沪0116民初13225号民事判决书。

⑥ “‘一键脱衣’牟利 AI技术岂能乱用”，载 <https://news.cctv.com/2024/04/03/VIDEK1kfOQ9BoE1NciHFzNXx240403.shtml>，2025年12月25日访问。

⑦ 参见杭州互联网法院（2023）浙0192民初4563号民事判决书。

应用

享有肖像权、名誉权等人格权。未经授权使用他人肖像进行AI换脸,侵犯肖像权;若移植到不雅、违法视频中,还会进一步构成对名誉权的侵害。第二,对财产权的侵害。AI换脸技术被应用于电信诈骗、金融盗窃等财产犯罪中。通过伪造他人身份或动态人脸视频,实施欺诈或绕过身份验证机制,非法获取他人财产。比如张某、余某利用非法取得的公民个人信息,借助软件生成3D头像通过支付宝人脸识别验证,骗取新用户红包奖励,最终被以侵犯公民个人信息罪、诈骗罪定罪。^①第三,对个人信息权益的侵害。人脸信息作为敏感个人信息,受到法律的严格保护,未经同意收集、使用进行AI换脸,严重侵犯了公民的个人信息安全。

2. 公共法益的侵害

国家法益是指以国家作为法律人格者所拥有的公法益,社会法益是指以社会整体作为法律人格者所拥有的社会共同生活之公共利益,国家法益与社会法益合称为公共法益。^②AI换脸行为对公共法益的侵害同样不容忽视:一是对社会公共秩序的破坏。利用AI换脸技术伪造重大公共事件、商业事件的不实视频,可能引发舆论混乱或扰乱市场秩序,制作的淫秽视频则违反善良风俗。二是对社会信任体系的侵蚀。当公众意识到互联网所见存在伪造可能性时,在长期面对真假难辨的信息环境时,将引发社会焦虑。三是对司法秩序的挑战。AI换脸技术如若被用于视听资料等证据材料的伪造,将干扰司法公正,给司法取证与证据审查带来挑战。四是对国家形象的损害。如果滥用AI换脸技术生成涉及国家机关的不实视频,不仅可能误导公众,还会损害国家形象和权威性,影响政府公信力。

综上,AI换脸行为侵害的法益呈现出多元化、复杂性的特点,这种特征决定了刑法规制路径应当坚持全面保护与重点打击相结合的原则,衔接好民事、行政相关法律法规,根据行为的社会危害性程度,构建层次分明的法律治理体系。

二、AI换脸行为刑法规制的现实困境

AI换脸行为越来越受到司法关注,虽然暂未出现大规模的刑事案件,但是将相关技术作为犯罪工具的案件有所增加。分析AI换脸行为的刑法规制现状,可知存在以下问题。

(一) 保护失衡: 社会法益与个人法益的救济失当

当前,刑事司法对AI换脸行为的规制呈现出“重秩序、

轻个体”倾向,其评价重心普遍置于对社会公共法益的维护,而相对忽视了对受害者个人法益的充分救济^③。这一救济失当问题在“淫秽传播型”犯罪模式中体现得尤为典型。司法机关在处理此类案件时,通常援引刑法第三百六十三条,将行为主要定性为“制作、传播淫秽物品牟利罪”,认为该行为侵犯了社会公众的性羞耻心,是典型的侵害社会健康风俗和管理秩序的犯罪。

然而,这种单一的法益评价视角,在AI换脸技术背景下具有局限性,相较于主动拍摄并传播牟利的传统淫秽物品犯罪,利用AI换脸技术生成的淫秽内容,对特定被换脸个体造成的侵害更为深远。行为人未经同意,将公民肖像与淫秽信息进行非法合成与广泛传播,直接导致受害者的人格尊严遭到践踏、社会评价被恶意贬损,并承受持续性的精神痛苦与隐私焦虑。若仅以传播淫秽物品罪论处,则刑事评价体系仅回应了行为对社会风尚的抽象危害,遗漏了对被换脸者具体人格法益的独立、充分评价,导致刑罚的惩治与救济功能出现疏漏。

在现行刑法体系下,受害者的人格权益损害虽然可以通过刑法第二百四十六条的诽谤罪寻求救济,但司法实践中的入罪门槛较高。根据最高人民法院、最高人民检察院《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》第2条的规定,一般情况下,同一诽谤信息实际被点击、浏览次数达到5000次以上,或者被转发次数达到500次以上的才构成刑法第二百四十六条第一款规定的“情节严重”,才符合入罪条件。然而这一入罪标准在应对AI换脸诽谤时面临双重困境:其一,危害的即时性与扩散的隐蔽性。一段深度伪造视频在极短时间内可能就会对受害者的声誉造成毁灭性打击,但其在私密社群、境外平台的传播次数却难以精确统计并达到法定数额门槛。其二,损害后果难以量化。视听伪造带来的心理创伤与名誉贬损更为严重,却无法在次数量化中得到充分体现。因此,即便通过诽谤罪路径保护个人法益,也可能因情节严重证明困境致使大量严重侵害人格尊严的换脸行为无法受到刑法的有效规制。

(二) 规制空白: 非法使用行为的定罪缺失

AI换脸技术实施与危害的核心客体是公民的人脸信息。此类信息在个人信息保护法中已被明确定性为敏感个

① 参见浙江省衢州市中级人民法院(2019)浙08刑终333号刑事裁定书。

② 张明楷:“公共法益与个人法益混同立法例下的刑法解释”,载《比较法研究》2024年第4期。

③ 刘艳红、姜文智:“AI换脸行为刑法规制的纠偏:法益与罪数的双重路径”,载《中国社会科学院大学学报》2024年第5期。

人信息中的生物识别信息,具有唯一性、专属性、人身依附性,^①应被视为一种承载人格尊严、人身安全与财产权益的复合型新型法益。

我国对个人信息的法律保护呈现出刑法先行的特点。^②自2009年刑法修正案(七)设立侵犯公民个人信息罪相关条款,并经后续修订完善后,该罪名已成为刑事保护的主要依据。随着网络安全法、个人信息保护法等法律法规陆续出台,我国已初步构建起前置法与刑法相互配合的个人信息保护体系。然而,现有法律框架在应对生物识别信息滥用等问题上仍面临挑战。^③当前刑法规定的侵犯公民个人信息罪主要规制的是非法出售、非法提供、非法获取公民个人信息的行为,该罪名似乎存在罪名设定太大、行为方式规制太小的问题。^④在AI换脸犯罪中,并非所有的人脸信息都是通过非法手段获取,而是有可能利用他人已经自行公开的个人照片或者视频等信息进行不法使用,这就导致“合法获取+非法使用”的AI换脸行为游离于刑法规制之外。此规制困境引发了学理上的争议:一种观点主张,鉴于人脸信息的重要性与现有罪名的局限性,应考虑在刑法中增设独立的身份冒用罪,精准打击身份信息的非法使用行为。^⑤反对观点则认为,频繁增设新罪可能破坏刑法的稳定性,主张解释论优于立法论,可通过扩张解释现有罪名予以应对。

(三)归责困境:网络服务提供者法律责任的边缘化

在AI换脸技术滥用的治理链条中,提供算法模型、工具软件或内容传播渠道的网络服务提供者居于关键环节。相较于传统犯罪,AI换脸行为的实施高度依赖数字平台,其伪造内容的生成、传播与危害扩大,与平台的技术防控措施、审核机制及合规意愿直接相关。因此,将网络服务提供者纳入刑事责任的评价体系具有现实必要性。然而,当前刑法规制仍主要聚焦于直接实施换脸、诈骗等行为的前端不法分子,对具有技术控制与内容管理能力的平台,其刑事责任存在追责缺失与边缘化倾向。若平台未充分履行用户个人信息保密义务,可能导致信息大

规模外泄;若未有效履行对违法内容的审查与及时移除责任,则会助长不实信息与侵权内容的传播。在此类情形下,若平台未充分履行用户个人信息保密义务或未有效履行对违法内容的审查与移除责任,理论上可以适用刑法第二百八十六条之一的拒不履行信息网络安全管理义务罪。该罪名虽已明确网络服务提供者的网络安全保障义务,但司法实践中适用极少。^⑥

这一适用困境部分是由于传统网络治理规则与AI技术发展之间的矛盾所引发的。在AI换脸等技术背景下,传统的避风港原则是否仍可无条件适用。该原则以不知情且未直接获利为前提,赋予平台事后删除义务即可免责。AI换脸内容具有高度伪造性与识别困难性,权利人与普通用户难以有效发现并通知,并且其传播速度的迅速性与危害程度的严重性,使得事后救济往往为时已晚。特别是,当平台自身提供或深度集成了换脸生成工具、算法接口时,其技术角色已从中立的信息通道转变为风险的积极参与者。因此,学界与实务界愈发主张,对于涉及生物识别信息滥用与人格权侵害的高风险场景,应要求具备技术优势地位的网络服务提供者化被动为主动,积极履行审查与监管义务。^⑦

三、AI换脸行为的治罪方案

(一)总体原则:刑法“二次法”角色的尊重与调适

1.以解释论完善既有罪名规制体系

AI换脸不法行为本身即是对公民个人信息权的侵犯。人脸信息是生物识别信息,属于个人信息范畴,这在民法典第一千零三十四条第二款、网络安全法第七十六条、个人信息保护法第二十八条第一款中已有明确指引。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》)也将侵犯“健康生理信息”作为打击对象,相当于肯定了生物识别信息(包括人脸信息)的个人信息地位。AI换脸是一种新型技术,是行为人用以实施犯罪的辅助工具,本质上仍是利用技术手段,对现有刑法体系的个人法益与公共法益的侵

① 李怀胜:“滥用个人生物识别信息的刑事制裁思路——以人工智能‘深度伪造’为例”,载《政法论坛》2020年第4期。

② 劳东燕:“通过指导性案例推进个人信息的刑法保护”,载<https://mp.weixin.qq.com/s/YxwRDqgEYaMsztMmcPlO6w>,2026年3月5日访问。

③ 同本页注释①。

④ 刘宪权、何阳阳:“个人信息保护法视角下侵犯公民个人信息罪要件的调整”,载《华南师范大学学报(社会科学版)》2022年第1期。

⑤ 李腾:“‘深度伪造’技术的刑法规制体系构建”,载《中州学刊》2020年第10期。

⑥ 张婷:“人工智能‘深度伪造’技术风险的刑法应对”,载《昆明理工大学(社会科学版)》2025年第5期。

⑦ 肖姗姗:“‘三端共治’模式下滥用深度伪造技术的刑法规制”,载《湖北警官学院学报》2024年第3期。

应用

害,没有超出现行刑法保护法益范围。即使有刑法规定存在滞后性的部分,也应当优先以解释论对其予以完善。

2. 在谦抑性的基础上完善与前置法的衔接

AI 换脸行为涉及的个人信息保护规范,除民法典、网络安全法、个人信息保护法等前置法外,还有国家互联网信息办公室联合多部门出台的一系列部门规章。如《网络音视频信息服务管理规定》指出,网络音视频信息服务提供者和使用者在制作、发布、传播非真实音视频信息的,应当以显著方式予以标识;《互联网信息服务算法推荐管理规定》要求算法推荐服务者对于未作显著标识的算法生成合成信息的,应当作出显著标识后才能继续传输;《互联网信息服务深度合成管理规定》明确深度合成服务提供者应履行用户身份认证、内容管理、显著标识义务;《生成式人工智能服务管理暂行办法》对生成式人工智能的技术治理、服务规范等划定责任红线。对于 AI 换脸这类深度伪造技术,我国采取的是以标识义务为中心的开放性治理模式,在肯定技术合理存在于规范使用的前提下,推动技术应用与发展。^①所以,在面对新兴技术带来的风险时,刑法应秉持谦抑性原则,避免动辄将增设新罪当作打击犯罪的不二法门。在现有民事法律、行政法律规范丰富发展的形势下,刑法作为“二次法”,不必对前置规定亦步亦趋,而应为民事处理、行政处罚留有一定空间,刑法判断也应当坚持一定的独立性,在前置规定的基础上再行限制入罪范围,为介入社会治理选择适当的切入时机,确保处罚适当。^②

（二）实务进路：入罪标准的回应与罪数选择

1. 情节严重的解释扩充与非法使用的入罪必要

为破解个人法益与公共法益救济失当的问题,可对“情节严重”要件作出符合AI换脸技术特性的扩大解释。侵犯公民个人信息罪、诽谤罪等罪名在入罪条件上都设有情节严重的兜底,可结合司法实践和个案情况,允许运用解释的方式对此内涵适度扩充后再行适用,无需机械地受制于数量、次数、金额的入罪门槛,以畅通刑事司法领域的个人权利救济,实现个人法益与公共法益的全面保护。对于AI换脸案件情节严重的评判角度,可以从被侵犯的主体、传播范围、逼真程度、影响后果等因素综合认定,也可参照

适用规范领域对相关罪名达到严重的要求。如最高人民法院、最高人民检察院《关于依法惩治网络暴力违法犯罪的指导意见》中,对侮辱诽谤自诉转公诉的范围进行了扩张解释,即在判断是否达到严重危害社会秩序时,除了结合“侵害对象、动机目的、行为方式、信息传播范围、危害后果等因素作出判定”外,还于第12条增设了“严重危害社会秩序”的5种情形,并迅速在司法实践中得到了回应。^③

为填补人脸信息“合法获取+非法使用”的规制空白，可以通过解释将“非法使用”行为纳入侵犯公民个人信息罪的规制范畴。以合法的方式获取信息后，再用非法手段进行使用，是目前利用深度伪造技术侵权的主要行为方式。^④而此行为并未包含在侵犯公民个人信息罪及相关司法解释规定的出售、提供、窃取等以其他方法获取之列。为填补“合法获取+非法使用”场景下个人信息的保护缺失，宜将非法使用纳入侵犯公民个人信息罪的行为方式。有人或许会认为这是对刑法的不当扩大解释，但鉴于人脸信息一旦被非法使用，极易引发诈骗、敲诈勒索、盗窃等下游犯罪。因此对于此类属于影响人身、财产安全的敏感信息，在入罪条件上应当举轻以明重，与轨迹信息、通信内容、征信信息、财产信息同等适用《解释》中第5条第3项关于50条以上的数量规定。

2. 从一重罪与数罪并罚的区分与取舍

AI 换脸不法行为通常侵犯多个法益,涉及多个行为,触犯多个罪名,应当予以充分评价,避免遗漏。单纯将 AI 换脸信息获取、出售、提供、使用,达到入罪标准的,构成侵犯公民个人信息罪,上述不法行为同时构成其他犯罪的,按照想象竞合犯的处理规则,从一重罪处罚。在判决的表述上应对行为人触犯的数个罪名予以明示,以彰显刑法保护多个法益的机能,发挥刑法行为规范的功能,实现特殊预防与一般预防的效果。如将合法获取的人脸信息利用 AI 换脸制作淫秽视频用以牟利,构成侵犯公民个人信息罪与制作淫秽物品牟利罪的想象竞合,择一重罪处罚。

如果行为人获取、出售、提供、使用人脸信息 AI 换脸后又实施其他犯罪,存在数个行为,同时触犯数个罪名,在没有法律特别规定时,不应适用牵连犯的从一重罪的处

①姜瀛：“人工智能‘深度伪造’技术风险刑法规制的向度与限度”，载《南京社会科学》2021年第9期。

②喻海松：“‘刑法先行’路径下侵犯公民个人信息罪犯罪圈的调适”，载《中国法律评论》2022年第6期。

③ 参见浙江省遂昌县人民法院(2024)浙1123刑初9号刑事判决书。

④ 刘宪权、何阳阳：“个人信息保护法视角下侵犯公民个人信息罪要件的调整”，载《华南师范大学学报（社会科学版）》2022年第1期。

理规则，应数罪并罚。因为成立牵连犯要求手段行为与目的行为在社会生活中具有通常性，而现实生活中利用 AI 换脸实施犯罪的形态复杂多变，行为模式多元，难以论断其与侵犯公民个人信息罪构成紧密的牵连关系，即便前后行为存在手段与目的的关系，但在数行为不常并发的情况下，亦不构成牵连犯。^①该精神也在最高人民法院、最高人民检察院、公安部《关于依法惩处侵害公民个人信息犯罪活动的通知》第 2 条中得到了体现。如果行为人实施 AI 换脸不法行为后又实施诈骗行为，同时触犯侵犯公民个人信息罪与诈骗罪，应当数罪并罚；如果行为人 AI 换脸行为不能单独构成侵犯公民个人信息罪，此后利用该信息实施诈骗行为构成诈骗罪的，仅成立诈骗一罪，但侵犯公民个人信息的行为可以作为量刑情节予以考虑。

3. 网络服务提供者的归责样态

技术中立但行为有责。如何在鼓励科技创新与风险控制之间寻求平衡，是法律面对的时代课题。在刑事司法领域，一般的网络技术服务提供者并非其管制对象，刑法对此应坚守谨慎克制的态度，辩证看待新兴技术的风险，厘清网络服务提供者的刑事责任形态与边界，防止打击扩大。当网络服务提供者在收到用户权益被侵害的通知后，及时移除侵权链接或者积极采取措施减少用户损失的，推定其主观上属于过失，不构成刑事犯罪。但当网络服务提供者以履行了避风港义务为由怠于履行平台审查义务的，致使不法侵害结果扩大或者侵权行为频发的，应当推定其主观上属于间接故意，应当追究其法律责任。^②

提供 AI 换脸技术的网络服务者，若直接参与非法出售、提供、使用、窃取或以其他方式非法获取人脸信息的，应视为直接实施侵犯公民个人信息犯罪的主体，需承担相应的刑事责任。若网络服务提供者与行为人事前或事中存在明确共谋并共同实施犯罪，则应按照共同犯罪中的主犯或从犯论

处。为 AI 换脸提供播放、传播平台的网络服务提供者，如违反网络信息管理等法定义务，则可构成参与和帮助犯罪的共犯、拒不履行信息网络安全义务罪的正犯、帮助信息网络犯罪活动罪的正犯。具体而言，对于平台单纯不作为，即明知行为人在平台传播的内容涉嫌违法犯罪，但不履行犯罪阻止义务，此种帮助行为存在刑法上的可罚性，具有帮助犯罪共犯责任的客观事实基础，在平台与行为人具备意思联络时，具有帮助犯共犯责任的主观要件，可以按帮助犯共犯追究责任。对于平台经责令改正后仍不作为，致使违法信息大量传播、用户信息泄露，造成严重后果时，应承担拒不履行信息网络安全管理义务罪的正犯责任。对于平台符合帮助信息网络犯罪活动罪正犯时，原则上以该罪正犯科处，但当受帮助的犯罪的正犯责任，轻于帮助信息网络犯罪活动罪的正犯责任时，例外地承认平台承担所帮助犯罪的共犯责任，是贯彻罪刑相适应原则与追求刑罚报应正义的需要。^③

结语

AI 换脸技术并非原罪，其滥用才是风险之源。面对技术迭代带来的治理挑战，刑法既不应缺位，也不能越位。笔者认为，在现有法律框架内，通过解释论路径即可回应大多数 AI 换脸行为的规制需求，无需急于立新罪。未来，应在“技术——平台——用户”三端协同治理的基础上，强化前置法与刑法的衔接，对“情节严重”进行贴合技术特性的实质解释，将“合法获取+非法使用”行为纳入刑法评价范畴，明确平台审查义务与刑事责任边界，真正实现对个人权益的充分救济与对公共秩序的有效维护。当然，刑法仅是治理体系中的最后防线。AI 换脸技术带来的挑战本质上是社会性的，其有效治理有赖于技术标准、行业伦理、行政监管与公民素养的协同共进，唯有如此，方能保证技术在造福社会的同时，始终运行在法治与善治的轨道之上。

（作者单位：江苏省南京数据资源法庭）

① 郑高健：“深度伪造技术滥用行为的刑法回应”，载《法律科学（西北政法大学学报）》2025 年第 3 期。

② 肖慧、彭拓、龚正良：“‘深度伪造’行为司法认定规则探究”，载《人民检察》2024 年第 13 期。

③ 霍俊阁：“网络直播平台帮助犯罪行为的处罚边界”，载《重庆大学学报（社会科学版）》2021 年第 2 期。