**编者按** AI换脸技术作为一种深度合成技术，基于深度学习算法和计算机视觉技术，能够精准地进行面部特征转移。然而，这一技术的广泛应用，也诱发了一系列违法滥用行为，带来诸多安全和法律风险，亟须加以规范和治理。为了推动AI换脸技术的合法使用，促进生成式人工智能健康发展和规范应用，《人民检察》邀请专家学者对相关问题进行探讨，敬请关注。

# AI换脸的应用风险及治理


## 问题一：AI换脸作为深度合成技术的一种新发展形态，具有哪些特征，存在哪些应用风险？

**田宏杰：**AI换脸技术又称人脸深度伪造技术，是一种基于人工智能深度学习和计算机视觉的前沿技术应用，具有高仿真性、高效自动化、数据依赖性、可扩展性、即时性、大众化等特征。目前，该技术被广泛应用于影视和娱乐行业、社交媒体内容创作、教育培训、心理治疗、身份验证、科学研究等领域。伴随该技术在各领域中的加速应用，其潜在风险也逐渐展露。一是引发信任危机。当公众意识到互联网所见存在伪造可能性时，便会动摇其对数字身份认证的信任，在长期面对真假难辨的信息环境时，将引发社会焦虑。二是激化社会矛盾。使用AI换脸技术伪造的名人、公众人物的视频、图像被用于制造虚假新闻、误导舆论，其中的不当言论可能引发社会动荡。三是挑战社会伦理。AI换脸被用于不道德的娱乐内容或违反伦理的内容，可能挑战社会伦理底线。四是威胁国家安全。伪造国家领导人讲话或者涉军事视频、照片可能引发国际问题，造成国家间对信息的误判或外交冲突，引发军事对峙。

**张建忠：**AI换脸技术作为深度合成技术的一种形式，通过深度学习算法和计算机视觉技术，将一张人脸特征映射到另一张脸上。AI换脸不仅可以进行静态图像置换，还可结合动作捕捉与语音识别，实现更复杂的动态视频和交互场景。AI换脸技术具有高度拟真性，能够捕捉并复现面部细节，如表情、光照与肌肉运动等，使得最终合成结果与真实人脸极为相似，通常肉眼难以分辨。随着技术普及，许多手机应用或在线工具都能一键生成换脸内容，技术使用门槛大幅降低。AI换脸技术在很多领域展现出积极的应用潜力，但其潜在的风险也不容忽视。一是侵犯公民权利。例如，擅自使用他人肖像制作、传播恶搞视频，可能造成名誉侵权；或者通过换脸技术伪造身份实施电信诈骗，威胁个人财产安全。二是破坏社会公共秩序、商业秩序。例如，利用换脸技术伪造重大公共事件、商业事件的不实视频，引发舆论混乱或扰乱市场秩序。三是危害国家安全。例如，通过换脸技术伪造领导人的发言视频，煽动恐怖暴力情绪，教唆进行恐怖活动。四是给司法办案带来挑战。例如，AI换脸技术被用于视听资料等证据材料的伪造，给司法取证与证据审查带来挑战，等等。

**阴建峰：**AI换脸的本质是深度伪造技术的应用，核心独特性在于以数据为燃料、算法为引擎的智能生成范式，其技术特征既体现了深度学习在表征学习和生成任务上的突破，也凸显了合成媒体时代技术伦理与治理的复杂性。其具有以下特征：一是高度拟真性，二是成本低廉性，三是场景泛化性。AI换脸技术在诸多领域得到了广泛运用，在给我们的生活带来巨大便利的同时，也存在潜在的应用风险：一是存在侵犯他人人格权益、财产权益等合法权益的风险。二是存在虚假信息传播，引发社会信任危机。三是存在危害社会稳定和国家安全的风险。

**高艳东：**AI换脸技术具有以下三个特征：第一，以假乱真性。第二，高度人身性。第三，技术门槛低。新技术总有新风险，AI换脸技术存在如下风险：第一，可能导致社会、法律关系的混乱。AI换脸技术可能产生没有真实对应自然人的数字人，尤其未来智能数字人的诞生，将打破法律单一主体的局面，法律将面临“自然人+数字人”的双主体困惑。第二，侵犯个人信息和隐私。根据国家互联网信息办公室、工业和信息化部、公安部印发的《互联网信息服务深度合成管理规定》，使用深度合成服务时，使用者需要取得被编辑个人的单独同意，AI换脸使用他人的信息尤其是生物识别信息，如果没有征求对方同意，就会侵犯他人的个人信息权利。

## 问题二：违法滥用AI换脸技术面临哪些法律责任？

**田宏杰：**AI换脸技术应用门槛较低、隐蔽性强，从而常常成为违法犯罪的工具。实践中，滥用AI换脸技术的违法行为目前主要表现为以下三类：一是滥用换脸技术侵犯个体人格权益，如未经授权采集、处理人脸特征数据等；二是滥用换脸技术冲击知识产权保护机制，如非法截取影视作品、图像等受著作权保护的素材实施换脸合成视频或者图片等；三是滥用换脸技术破坏社会管理秩序，如利用深度合成技术伪造身份认证信息、传播虚假舆情或实施非接触式诈骗等。

**张建忠：**AI换脸技术在推动科技创新和社会进步的同时，在实践中衍生出了一系列违法滥用行为，需要引起高度重视。在案件办理中，我们发现此类技术被用于违法犯罪，并呈现出复杂化和多样化的特点，主要表现为侵犯人格权利、传播违法内容、冒用他人身份等。根据违法行为的性质和后果，行为人可能触犯以下罪名：一是通过伪造视频或身份实施诈骗，非法获取财物的，涉嫌诈骗罪。二是利用伪造视频散布虚假信息，严重损害他人名誉的，涉嫌诽谤罪。三是利用AI换脸技术制作、传播淫秽内容，情节严重的，涉嫌制作、贩卖、传播淫秽物品犯罪。四是制作虚假信息引发公众恐慌或扰乱社会秩序，涉嫌危害公共安全、扰乱公共秩序等犯罪。

**阴建峰：**AI换脸技术的违法滥用行为在现实中的具体表现多样且危害深远，行为人可能面临以下法律风险：一是民事侵权的法律风险。未经他人同意，擅自将他人的肖像用于视频、图片或其他形式的传播，将构成对他人肖像权的侵犯。二是行政违法的法律风险。行为人违法滥用AI换脸技术可能面临多方面的行政违法风险，如未经授权采集、使用或泄露人脸信息，违反了个人信息保护法中关于敏感个人信息保护的相关规定；未对换脸内容进行显著标识或制作、发布、传播虚假信息，违反了关于网络信息管理的相关要求，可能面临警告、罚款、责令停产停业等行政处罚。三是刑事犯罪的法律风险。行为人在使用AI换脸技术时，非法收集、使用或泄露他人的人脸信息等敏感个人信息，可能会构成侵犯公民个人信息罪。制作嘲讽或丑化他人的换脸视频，实施网络欺凌或者发布侮辱性内容的，可能构成侮辱罪。行为人利用AI换脸技术制作、传播含有淫秽内容的视频或图片，可能构成制作、传播淫秽物品牟利罪或传播淫秽物品罪。行为人利用AI换脸技术伪造他人身份，通过视频通话等方式实施诈骗，骗取他人财物的行为则可能构成诈骗罪。制作虚假视频威胁被害人支付赎金，实施敲诈勒索活动的，可能构成敲诈勒索罪。伪造视频作为法庭证据，干扰司法公正的，可能构成伪证罪。

**高艳东：**根据违法滥用AI换脸技术行为的危害程度，行为人可能承担以下法律责任：首先，违规使用者可能侵犯隐私权、肖像权、著作权、名誉权等，以及构成不正当竞争，损害消费者利益，需承担相应的民事责任。其次，AI服务提供者可能因管理不当而受到行政处罚。根据《互联网信息服务深度合成管理规定》，如果服务提供者未履行规定的安全评估、备案、标识等义务，可能会面临行政处罚。再次，违法使用者可能被追究刑事责任。具体可能构成侵犯财产类犯罪、侵犯公民个人信息罪、传播淫秽物品类犯罪等。

## 问题三：如何构建AI换脸技术应用规制体系？

**田宏杰：**为规范AI换脸技术的应用，我国已基本建立起法律层面的治理框架，通过民事、行政、刑事法律的协同作用，对AI换脸技术进行多维度规制。但是，随着技术的迭代升级和滥用行为的日渐增多，现有治理框架和规制措施的局限性亦逐渐显现出来。一是缺乏专门立法和分级监管制度；二是数据使用限制不足，溯源要求不明确；三是用户权利保护不足；四是缺乏强制性检测技术和认证机制；五是平台责任界定模糊。有鉴于此，需以“技术过程法律化”与“法律要求技术化”的双向互动为内核，围绕以下方面进一步健全完善AI换脸技术的规制体系，以达成技术赋能与法律规制的动态平衡。一是制定专门立法规范。有必要突破分散立法模式，探讨制定人工智能合成内容管理法，确立技术应用负面清单制度，区分娱乐创作、公共传播、身份认证等场景设置差异化合规标准；同时，引入“技术过程责任”概念，将训练数据合法性、算法可解释性纳入重点监管内容。二是升级技术嵌入治理。推行“防滥用”技术强制标准，要求开发者预置内容失效触发机制与不可逆数字水印；推动平台建立“技术筛查+人工复核+第三方评估”的复合审核机制，根据内容风险等级设置差异化的审核标准和响应时限；构建国家级深度伪造检测平台，整合微表情分析、虹膜纹理匹配等多模态鉴别技术，尝试建立“生成—传播—消除”的全周期响应体系。

**张建忠：**为了在保障技术发展与法治秩序之间实现平衡，实现技术发展与法律规则的和谐共生，我国已在立法、行政和司法层面进行了积极探索，但仍存在以下不足：一是现有规制体系较为分散，系统性和针对性不足；二是监管部门检测与溯源手段多依赖相关企业或机构的技术储备；三是技术开发企业的合规管理机制仍有待完善。未来可通过立法、行政、技术、社会以及国际多维度协作，建立健全具有前瞻性、系统性和针对性的规制体系，确保技术发展造福社会。一是明确技术应用的边界、使用规则以及违法责任追究机制，重点关注技术开发者、服务平台和使用者的权责边界，构建全面的责任链条。二是明确技术平台和内容平台的审查义务，要求相关平台对AI合成内容进行检测与标注，并在出现违法或高风险内容时及时预警和处置。三是推动技术创新，支持科研机构与企业开发更加高效的识别算法与溯源技术，如数字指纹、水印技术等，加强对换脸内容的真实性验证工具研发。四是加强与国际社会的交流与合作，可以在全球范围内建立信息共享和执法合作机制，借鉴域外经验，推动更广泛的国际规则对接，形成跨国打击AI技术滥用行为的联动网络。

**阴建峰：**健全完善AI换脸技术的规制体系可从以下方面着手：一是针对AI换脸技术构建系统全面的法律体系。制定统一的上位法，提高立法效力位阶，并明确规定统一的安全责任底线和基本使用标准，包括必要性原则、知情同意原则、第三方监督机制以及损害赔偿机制等内容，增强法律的权威性和系统性。优化规则供给机制，加强复合型规范治理。应对新兴技术引发的法律问题，需要特别重视软法的功用，发挥好软法规则小而精、快而准、迅捷灵活治理的优势，建立人工智能安全标准、伦理准则和行业自律规范，推动行业的自治自律。二是细化现有的法律规定，对一些关键性问题予以及时回应。确立以区分应用场景前提下的合理使用为基本规制原则，根据不同场景的风险程度和利益平衡，制定差异化的规制措施。对违法主体法律责任的承担予以明确规定。例如，明确侵权责任主体的认定及责任划分标准，在民事责任领域考虑设立惩罚性赔偿机制，在刑事制裁方面考虑适用禁止令。

**高艳东：**虽然我国出台了一系列规定，但仍存在法律位阶较低、责任主体不明、监管盲区较大等问题，未来的监管应把握以下三点：一是明确合理使用原则。区分应用场景，以应用场景为基本出发点加强管制，确保技术向善。二是建立全链条多主体的监管规则。一方面，利益链条上建立健全“事前备案—事中内容审查—事后处罚”的监管体系；另一方面，明确技术开发者、平台运营者、使用者的法律责任。三是加强国际合作。与国际社会合作，共同应对AI换脸技术带来的挑战，分享技术和经验，加强司法协作，形成全球治理合力。

## 问题四：检察机关如何在依法履职中加强对AI换脸技术应用的法律监督？

**田宏杰：**AI换脸技术带来的法律和伦理问题不容忽视，检察机关作为宪法规定的法律监督机关，可重点围绕以下方面加强对AI换脸技术应用的法律监督。一是建立刑事追责与公益诉讼并行机制。基于AI换脸技术的多重违法性特征，在依据刑法对涉罪行为追究刑事责任的同时，检察机关可视案情同步推进民事公益诉讼，以确保当事人的合法权益得到及时保护。尤其是对于侵害不特定多数人生物识别信息安全的案件，通过公益诉讼程序的及时介入，能够在保护公共利益、追究行为人刑事责任的同时，实现对当事人“停止侵权+损害赔偿”的复合救济。二是强化行刑衔接机制。检察机关可联合网信、公安等部门，建立违法线索双向移送机制，对违法案件中的犯罪线索及时立案。明确证据转化要求与电子数据固定标准，推动建立跨部门技术鉴定与评估体系。三是推动建立跨国司法协作机制。针对跨境服务器托管、虚拟货币支付等黑灰产，以检察机关为主体推进国际交流与合作，完善电子证据跨境调取与联合执法程序。

**张建忠：**作为法律监督机关，检察机关肩负着保障法律统一正确实施、保护国家利益和社会公共利益的重要职责。检察机关在AI换脸技术法律监督中应以问题为导向，充分发挥引导侦查、公益诉讼、检察建议等职能，为技术治理提供法治保障。具体可从以下方面着力：一是依法追究刑事责任，形成司法震慑。对于利用AI换脸技术实施诈骗、传播淫秽内容等犯罪行为，检察机关应严格依照刑法追责。对于重大案件，适时介入引导侦查机关精准取证。二是充分发挥公益诉讼职能。对因监管缺失导致违法内容广泛传播的平台或技术提供方，检察机关可以依法提起公益诉讼。三是强化检察建议，促进完善监管制度。在案件办理过程中，发现监管漏洞，检察机关应及时向相关部门提出完善建议。四是建立联动机制，提高监督效率。检察机关应当与网信、公安等部门建立联动机制，实现信息共享、技术共用和资源整合，提高监督效率。五是加强法治宣传教育，提升社会防范意识。

**阴建峰：**检察机关在依法履职办案中，应通过多种手段加强对AI换脸技术应用的法律监督，确保AI换脸技术在法治的轨道上运行。一是提高法律监督的精准性和有效性。一方面，积极应用大数据法律监督模型，通过数据分析、模型预测等手段，建立AI换脸技术检测机制，精准发现AI换脸技术滥用的线索，提高监督的针对性和有效性。另一方面，探索建立针对AI换脸技术滥用的公益诉讼制度，通过提起公益诉讼的方式，维护公民个人信息安全和社会公共利益。二是加强跨部门协作和信息共享。检察机关可以与公安、网信等部门加强协作，共同打击利用AI换脸技术进行诈骗、传播淫秽物品等违法犯罪行为。通过信息共享、线索移送等方式，形成合力，提高打击效果。三是针对AI换脸技术滥用的新型案件，制定司法解释，发布指导性案例，明确案件的定性、量刑标准等，为司法实践提供指导和参考。四是加强宣传教育和预防工作。通过开展法治宣传和警示教育，增强公众对AI换脸技术滥用危害的认识和防范意识，引导公众正确使用AI换脸技术，避免将其用于违法或不当用途。

**高艳东：**对于AI换脸，法律应当采用“不管私用、严管公用”的思路。检察机关应采取更积极措施，防止技术被恶意滥用。第一，加强与公安、网信等部门合作，共同构建高效的案件侦办体系。对于涉及AI换脸的刑事案件，如诈骗、诽谤等，适时介入引导取证，利用专业技术手段固定证据。同时，促进社会共治，与网信、公安等部门建立常态化的信息共享和协作机制，共同加强对AI换脸技术的监管。第二，探索人工智能技术应用领域的公益诉讼职能。针对AI换脸引发的新类型侵权纠纷，可尝试采用公益诉讼等形式维护不特定多数人的合法权益。第三，推动法律完善。检察机关在办案过程中积累了丰富的AI换脸技术相关案件经验，可根据最新技术发展特点，针对已有规定的不足，向立法机关提出完善相关法律法规的建议。

## 问题五：如何推动AI换脸问题的系统治理、源头治理？

**田宏杰：**系统治理要求将对AI换脸问题的治理融入社会系统，以多重手段确保AI换脸技术在合规框架内运行；源头治理则要求正确看待AI技术，既要将其作为AI换脸应用的出发点，也要将其作为治理AI换脸问题的落脚点。据此，AI换脸问题的协同共治可从以下几方面同步推进：一是强化平台责任。要求社交平台、AI模型平台等平台进行主动监测，强化其监督管理义务，鼓励运用技术手段识别AI换脸内容，对违规内容与账户“零容忍”，发现违规内容立即下架并报告监管平台。二是制定技术标准。创设AI生成内容强制识别标志，要求AI生成内容嵌入不可移除的数字水印或标签，确保对AI内容做到一般人可识别。三是落实追责机制。明确AI换脸技术侵权责任主体，由侵权行为人承担侵权或违约责任，由AI技术平台、社交平台承担补充责任，实现对AI技术开发者、内容制作者、传播平台、违法信息使用者的全链条追责。

**张建忠：**治理是一项复杂的系统工程，必须依靠多方协同，构建共治格局。一是完善制度体系，夯实立法保障。建议评估法律供给，对现行法律法规进行全面梳理，明确是否需要专门针对AI换脸或深度伪造技术制定专项法规，确保立法具有前瞻性和系统性。二是强化行政监管，提升动态监测能力。建议落实备案与审核机制，要求相关服务平台对深度合成内容先行审核和风险评估。此外，研发并部署实时监测与预警系统，及时发现并查处利用AI换脸技术实施的违法犯罪行为。三是强化技术自律，预防技术滥用。技术企业在研发阶段即应融入伦理与法律合规原则，鼓励企业建立自查制度，定期对自身算法和产品的合规性进行评估和报告。四是落实平台责任，阻断违法信息传播链条。内容平台应当完善审核机制，对用户上传的内容进行筛查、标记或备注提示，尽可能减少虚假信息对公众的误导。此外，建立便捷的举报和投诉通道，对涉嫌违法或者不良内容及时审查、删除并限制其传播。

**阴建峰：**针对AI换脸等深度伪造技术的治理，需打破单一主体治理的局限性，构建“监管执法—技术防御—行业自律—平台审核—公众监督”五位一体的协同治理体系，推动相关各方在关键环节形成合力，在鼓励创新与保障安全之间找到动态平衡，实现“技术向善、社会可信、治理可控”之愿景。一是监管部门应当加强监管执法力度。在构建“权责清晰”的主体协同框架的基础上，明确监管部门的职能分工，打造全链条治理格局。建立技术应用的备案机制，AI换脸技术的应用主体应当向相关部门备案，并获得备案编号，从源头上管理技术应用。在备案之后，相关部门应当对AI换脸技术的应用进行事前安全评估，聘请专家评估技术风险，做好风险防范。二是技术研发者包括相关的科技企业应当主动承担起社会责任。技术研发者须积极落实安全可信原则，建立AI换脸技术研发备案制度，嵌入伦理审查机制。在产品设计和算法开发过程中，将道德算法、伦理准则嵌入技术之中，使算法符合“善”这一科技伦理的要求。三是服务提供者应当加强行业自律，履行法定义务。提供AI换脸技术服务的相关行业应当建立健全行业标准、行业准则和自律管理制度，加强行业自律。四是平台应当加强对传播内容的审核，并建立快速反应机制。通过制定详细的内容审核标准，明确允许发布的范围。在内容发布前，利用AI技术开发自动化工具检测和标记使用AI换脸技术的内容，过滤掉违规伪造作品。建立快速反应机制，对违规内容进行有效监控和及时处理。鼓励用户参与内容监控，通过设置易于使用的举报工具，让用户可以快速报告可疑或明显违规的内容。对于被识别或被举报的违规内容，平台应立即下架，并对违规用户给予警告或封禁账号等处理。

**高艳东：**推进AI换脸问题的系统治理、源头治理，需要从以下方面着手：第一，构建全方位监管体系，确保违法必究。微观层面，由网信部门牵头，建立跨部门联合监管机制，加强对AI换脸技术的监督管理；中观层面，加强司法机关与行政机关联动，建立健全AI犯罪行刑双向衔接机制，提升国家治理AI换脸乱象的能力；宏观层面，整合国家机关、技术开发者、互联网平台、网民等多方资源，形成多行业管理、多部门治理、多层级参与的治理格局，让AI换脸技术滥用无处遁形。第二，强化协同共治，提升监管效能。通过媒体、教育机构等渠道普及AI换脸技术知识，增强民众识别能力和防范意识；建立有奖举报机制，公开举报渠道，鼓励民众积极举报AI换脸技术滥用线索；促进产学研合作，加强AI技术创新和成果转化，不断提升AI换脸技术检测能力；引导行业自律，鼓励行业协会等组织建立行业自律机制，如引入伦理审查机制，推动行业内部形成良好的AI换脸技术应用氛围。

（文章详见《人民检察》2025年第9期）