深度伪造技术应用中的法律风险与规制研究

陶海婷

中国计量大学，浙江杭州，310018；

摘要：深度伪造技术以对抗生成网络（GAN）为核心支撑，在影视制作、医疗康复、网络社交等领域展现出显著应用价值，却因高仿真性、快速迭代性等特征，衍生出技术异化、信息失真、信息泄露等多重法律风险，对人格权保护、社会信任体系及公共秩序构成挑战。现有法律规制存在分散化、滞后性问题，难以完全适配技术发展需求。本文将系统剖析深度伪造技术的原理与风险本质，借鉴美国分散化立法、欧盟统一监管经验，提出技术规制与法律规制的协同治理路径，为构建适配技术发展的法律防控体系提供理论与实践参考。

关键词：深度伪造技术；法律风险；协同规制；个人信息保护；对抗生成网络

DOI：10.64216/3080-1486.26.03.068

1 问题的提出

人工智能技术的迭代升级推动“深度合成”技术进入商业化落地阶段。深度伪造技术凭借高仿真性、泛用性特征在影视特效制作、医疗康复辅助、网络社交娱乐等领域实现广泛应用，但其“双刃剑”属性也因规制滞后逐渐凸显——2017 年国外 Reddit 平台用户“deepfake”通过面部替换技术制作并传播名人色情视频，引发全球范围内的滥用模仿热潮；2019 年我国“ZAO-逢脸造戏”APP 因用户照片滥用风险引发公众对个人信息安全的担忧，此类事件既对传统法律体系、社会治理模式提出全新挑战。基于此，本文借助比较研究法对比美国联邦与州两级分散立法、欧盟《一般数据保护条例》（GDPR）统一监管等模式，结合规范分析法以我国现有法律框架为基础，系统剖析深度伪造技术的原理与风险本质，旨在提出技术规制与法律规制的协同治理路径，为构建适配技术发展的法律防控体系提供理论支撑与实践参考。

2 深度伪造技术的原理与应用

2.1 深度伪造技术的原理

深度伪造的实现方式是把个人照片、视频等数据导入智能算法，经反复训练后算法会自主完成“换脸”动作[5]。它的核心突破在于运用对抗生成网络模型，该模型区别于传统深度学习技术的单向输出特性，新增了“对抗”机制[6]。其运作逻辑是借助生成器与判定器两个神经网络高仿图像：先从互联网等渠道搜集足量目标人物的视频或图像信息，搭建生成器算法训练所需的数据库；生成器经神经网络训练后生成相关视频、图像，再由判定器对原始真实图像数据与伪造数据展开比对博弈。

若二者偏差极小，达成“以假乱真”的效果，就留存伪造数据作为“换脸”素材；若偏差较大，系统会把数据退回生成器开展新一轮计算优化，直到生成图像通过判定器核验方可启用。最终经反复训练筛选，生成器生成的数据与原始真实数据拥有一致的分布特征，让判定器难以精准区分差异，二者形成动态平衡，进而让伪造视频骗过大多数人的视觉识别。

2.2 深度伪造技术的应用

深度伪造技术目前已广泛应用，核心应用场景包括人脸替换、人脸再现、人脸合成和语音合成四大类[1]。具体来看，人脸替换即大众熟知的“换脸 功能，比如曾走红的“ZAO”APP，可根据用户上传的照片自动替换影视角色面部，生成用户“出演”的短视频。人脸再现通过操控调整视频中人物的面部神态与语言内容，使其表现出非本人真实意愿的言行。人脸合成依靠人工智能调整人物面部特征，或生成全新人脸影像。语音合成可实现文字转语音、翻译等功能，还可构建新型语音体系，如虚拟歌手初音未来等。此外，该技术正朝着综合应用、全身合成、数字虚拟人等方向发展，彰显出其巨大的发展潜力与广阔前景。

深度伪造技术作为人工智能领域的突破性成果，拥有高仿真性、泛应用性及快速进化性等特性[3]。这些技术优势在实际应用中潜藏风险，它不仅是一把双刃剑，更可能是一把失衡的双刃剑[4]。若使用不当或被滥用，其负面效应将愈发突出。

3 深度伪造技术应用的法律风险

3.1 技术异化风险

技术异化指技术脱离预设价值目标，成为危害人类利益的工具，有学者指出，其本质是工具理性凌驾于价值理性之上[4]。深度伪造技术的异化风险体现在两方面：一是技术中立性被打破，低门槛性与高隐蔽性降低违法成本，普通公众借助简易工具即可制作虚假内容实施诽谤、诈骗，且高仿真性增加识别与追责难度；二是“算法霸权”问题，算法黑箱特性导致技术应用缺乏透明度，开发者、数据掌控者可能利用技术优势操控信息传播、干预公众决策，甚至形成技术垄断，损害市场公平与公共利益。人工智能时代的技术异化还可能威胁社会治理体系与国家主权[9]。

3.2 信息失真风险

信息失真风险是技术最直接的风险表现，指伪造信息传播导致社会信息环境混乱，破坏信息真实性与可信度。与传统剪辑、拼接的虚假信息不同，深度伪造内容高度仿真，普通人难以通过感官甄别，加剧“后真相时代”困境。法律层面，该风险引发三重问题：一是侵害公众知情权，虚假商品宣传、医疗信息误导决策，危害消费者权益与公众健康；二是破坏社会信任体系，虚假信息累积导致公众对官方媒体、政府机构产生怀疑，弱化公信力并引发信任危机[6]；三是扰乱公共秩序与国家安全，伪造政治人物言论、外交信息可能引发社会动荡、加剧国家间矛盾。

3.3 信息泄露风险

深度伪造技术以大量数据采集为基础，应用过程伴随个人信息泄露风险。个人信息已成为人工智能时代的核心资源，而技术对信息的利用常突破传统边界[8]。该风险体现在三维度：数据采集环节，部分主体为提升伪造效果，通过爬虫、偷拍等非法方式过度搜集个人信息，甚至获取脸部、视网膜等生物识别信息，而此类信息具有唯一性、不可更改性，敏感度远高于普通信息，泄露后对个人安全构成长期威胁[7]；数据存储环节，训练数据库包含大量敏感信息，若防护措施不足，易遭遇黑客攻击导致泄露；数据使用环节，部分主体将信息用于非法目的或非法交易，使信息在黑灰产业链流转，引发连环侵害。我国《民法典》《个人信息保护法》虽对个人信息保护作出规定，但技术特殊性导致法律适用面临采集边界模糊、生物信息保护不完善等挑战[8]。

4 深度伪造技术应用的协同规制

4.1 以技术规制技术

技术规制是风险防控的前置防线，通过发展溯源防伪、反向破解技术降低滥用风险，实现“以技术对抗技术”。网络可信身份管理技术是应对风险的重要支撑。具体路径包括：一是溯源防伪技术应用，以区块链技术为载体，将原始内容的创作时间、创作者身份、哈希值等信息上链存储，形成不可篡改的溯源链条；同时在内容中嵌入“数字水印”，即便被伪造，水印仍可提醒受众内容存疑并辅助溯源。二是反向破解技术研发，如微软推出的视频认证工具，通过检测合成内容的混合边缘、细微褪色等特征识别伪造视频[8]；此外，建立伪造内容数据库训练识别模型，推动技术开源共享，形成多元化防护体系。三是技术标准制定，行业组织牵头明确数据采集、模型训练、内容生成的技术规范，为合规应用与执法认定提供依据。

4.2 以法律规范技术

法律规制是风险治理的核心保障，需构建多层次体系，结合国内外经验优化规制路径。国外经验方面，美国采用联邦与州两级分散化立法，截至 2019 年已出台 12 项专项法案，如弗吉尼亚州将未经授权传播伪造内容定为刑事犯罪，加利福尼亚州禁止技术用于色情传播与政治助选[10]；欧盟通过 GDPR、《欧盟反虚假信息行为守则》构建统一监管体系，GDPR 第 9 条禁止使用生物性数据，守则将伪造内容纳入虚假信息管理[10]。

我国法律规制体系的完善需聚焦四方面：一是明确技术使用边界，细化《民法典》中肖像权、名誉权的保护标准，将“未经授权使用他人肖像进行深度伪造”直接认定为侵权；在《刑法》中增设相关罪名，对利用技术实施诈骗、危害国家安全的行为明确定罪量刑标准。二是完善个人信息保护规则，《个人信息保护法》需细化数据采集的合法性要求，采集信息用于技术训练需取得明示同意并告知范围与期限；强化生物识别信息的特殊保护，严格存储安全义务；建立数据泄露通知制度，要求处理者及时告知信息主体与监管部门并采取补救措施[8]。三是强化平台责任与行业自律，明确网络平台对伪造内容的审核、删除、屏蔽义务，要求建立专门审核机制；鼓励行业组织制定自律公约，开展合规培训提升行业合规意识。四是构建动态法律更新机制，采用框架性立法与配套法规结合的模式，通过定期评估调整规范内容，避免法律滞后性带来的规制漏洞。

4.3 以商业伦理规制技术

商业伦理规制是风险治理的“中间纽带”，聚焦技术应用主体的伦理自觉，通过规范企业行为填补技术与法律之间的规制空白。作为深度伪造技术的主要研发者与推广者，商业主体的伦理底线直接决定技术应用方向，具体可通过四方面构建约束机制：一是行业伦理准则的制定，由互联网协会、人工智能产业联盟等组织牵头，出台《深度伪造技术商业伦理指南》，明确“禁止利用技术制作色情、暴力内容”“不得未经授权使用他人肖像”等核心准则，将“知情同意”“最小必要”“损害预防”确立为三大伦理原则[5]；二是企业内部伦理审查机制的建立，要求研发或应用深度伪造技术的企业设立“伦理审查委员会”，对技术项目进行前置审查，如所有深度伪造相关项目需提交伦理审查申请，说明技术用途、数据来源及风险防控措施，未通过审查的项目不得推进；审查委员会需包含法律专家、伦理学者及公众代表，避免“企业自审自定”的漏洞。三是伦理责任的追溯与问责，明确企业在技术全生命周期中的伦理责任，如技术推广时需在用户协议中明示“合成内容识别方法”，发现技术被滥用时需及时采取“内容下架、账号封禁”等补救措施。四是行业伦理协作与监督，建立“深度伪造技术伦理共享平台”，企业可上传合规经验与风险案例，平台对违规企业进行“伦理警示”并向社会公示；同时开展“伦理培训”，提升企业员工的伦理意识。商业伦理规制并非替代技术与法律，而是通过“软约束”引导企业主动规避风险，形成自我规制与行业监督的良性循环。

5 结语

深度伪造技术的发展是人工智能时代技术进步的必然结果，其应用机遇与风险挑战并存。应对技术风险不能简单采取禁止发展的保守策略，而需构建“技术规制筑牢防线、商业伦理引导自律、法律规制守住底线”的三维协同体系：技术规制通过溯源、识别技术降低滥用可能性，商业伦理通过企业自觉与行业监督填补规制空白，法律规制通过明确责任划定合法边界。三者并非孤立存在，而是相互补充。技术为伦理与法律提供可操作工具，伦理为技术与法律提供价值指引，法律为技术与伦理提供强制保障。

未来，随着深度伪造技术向“全身合成”“实时交互”等方向发展，风险形态将进一步演变，需加强法学、计算机科学、伦理学的跨学科合作，深入研究“AI 生成内容的法律定性”“生物识别信息的特殊保护”等前沿问题；同时推动国际规制协作，应对跨境深度伪造虚假信息传播等全球性挑战，最终实现技术创新与风险防控的平衡，让深度伪造技术真正服务于人类福祉与社会进步。

参考文献

[1]曹建峰.AI 生成内容发展报告2020——“深度合成”(deep synthesis)商业化元年[R/OL].(2020-05-11)[2020-06-20].

[2]王禄生:《论“深度伪造”智能技术的一体化规制》,载《东方法学》2019 年第 6 期.

[3]Goodfellow I J,Pouget-Abadie J,Mirza M,etal.Generative Adversarial Networks[J].Advances in Neural Information Processing Systems,2014,3:2672-2680.

[4]陈仕伟.大数据技术异化的伦理治理[J].自然辩证法研究,2016(1):46-50.

[5]牛静、侯京南:《基于人工智能的换脸视频伦理问题探讨》,载《青年记者》2019 年第 15 期.

[6]蔡士林.深度伪造”的技术逻辑与法律变革[J].政法论丛,2020(6):131-140.

[7]孟雪,刘宗媛,李倩.深度伪造技术给网络可信身份管理带来的挑战与对策[J].网络空间安全,2020(6):75-79.

[8]张新宝.从隐私到个人信息:利益再衡量的理论与制度安排[J].中国法学,2015(3):38-59.

[9]吴汉东.人工智能时代的制度安排与法律规制[J].西北政法大学学报,2017(5):128-136.

[10]Deepfake legislation:A nationwide survey—State and federal lawmakers consider legislation to regulate manipulated media[EB/OL].(2019-09-25)[2019-1 1-19].https://www.jdsupra.com/legalnews/deepfakelegislation-a-nationwide-86809/.

[11]丁晓东译.《一般数据保护条例》[EB/OL].(201 8-05-23)[2020-06-02]https://www.tisi.org/5029.

作者简介：陶海婷（2001 年-），女，汉族，籍贯：浙江绍兴，学位：学士，主要研究方向：民法。