Advances in Social Sciences 社会科学前沿, 2025, 14(12), 666-673
Published Online December 2025 in Hans. https://www.hanspub.org/journal/ass
https://doi.org/10.12677/ass.2025.14121136

Hans汉斯

人工智能深度伪造技术的风险与治理

徐玉钦

长春理工大学法学院，吉林 长春

收稿日期：2025年11月4日；录用日期：2025年12月12日；发布日期：2025年12月23日

摘　要

文章聚焦于人工智能领域深度伪造技术的治理研究，系统梳理其技术原理、应用场景，揭示了深度伪造技术伴随其快速发展而带来的个人权利受损、社会秩序混乱及国家安全风险等问题。文章以风险社会理论、技术规制理论、平台治理理论为分析框架，通过对典型国家及地区现有法律法规比较研究，发现我国在当前规制体系中适用法律、责任分配、投诉机制，监管体系等方面存在明显不足。基于此，文章针对性的提出构建多维法律衔接体系、建立分级责任分配框架，设计投诉过滤机制，创新协同监管模式等治理意见，从而为保障人工智能技术的健康发展与社会秩序的稳定提供参考。

关键词

深度伪造技术，应用场景，风险，治理

Research on the Risks and Governance of AI Deepfake Technology

Yuqin Xu

School of Law, Changchun University of Science and Technology, Changchun Jilin

Received: November 4, 2025; accepted: December 12, 2025; published: December 23, 2025

Abstract

This paper focuses on governance research concerning deepfake technology within the field of artificial intelligence. It systematically examines the technical principles and application scenarios of this technology, revealing the problems arising from its rapid development, including infringement of personal rights, disruption of social order, and threats to national security. Through a comparative study of existing laws and regulations in representative countries and regions, significant deficiencies are identified in China's current regulatory framework, particularly concerning applicable laws, responsibility allocation, complaint mechanisms, and the oversight system. Based on this analysis,

文章引用: 徐玉钦. 人工智能深度伪造技术的风险与治理[J]. 社会科学前沿, 2025, 14(12): 666-673.
DOI: 10.12677/ass.2025.14121136

the paper proposes targeted governance recommendations, such as refining legal applicability, defining clear responsibility boundaries, establishing appropriate complaint thresholds, and innovating regulatory models. These measures aim to provide a reference for ensuring the healthy development of artificial intelligence technology and the stability of social order.

Keywords

Deepfake Technology, Application Scenarios, Risks, Governance

1. 引言

从手机语音助手到自动驾驶汽车，从高效能的算法分析到医疗领域的数字影像，人工智能正以前所未有的方式改变世界，带来便利的同时，其双刃剑的特性也渐渐显露，其中深度伪造技术就带来不小的隐患，在不久前就有人利用企业家雷军的声音和头像，制作了众多的恶搞视频，引起不小热议。本文从深度伪造技术的原理出发，系统分析其应用场景，总结出现有深度伪造技术可能带来的风险，而后分析现有国内外治理模式，吸取有益措施，最后就我国深度伪造技术出现的问题，提出针对性的建议。

2. 深度伪造技术风险的理论分析与应用场景下的风险审视

想要了解深度伪造技术引发的系统性风险，必须要理解其底层技术原理，并掌握关键应用场景，这也是能否有效防控深度伪造技术风险的逻辑起点。

2.1. 深度伪造技术的风险特征与理论分析

深度伪造(deepfake)是由深度学习(deep learning)与伪造(fake)相结合，也就是采用人工智能方法(深度学习)生成逼近真实效果的虚假图像、音频、视频等[1]。深度伪造技术主要依赖深度学习中的生成对抗网络(GAN)模型，这也就是深度伪造技术与以往伪造技术的不同之处。这一模型赋予了深度伪造技术相较于传统伪造手段的颠覆性特征，其一，高度逼真性，深度伪造技术现在能做到单张照片加简短音频即可生成肉眼难以辨认的伪造内容。其二，难以检测性，前文提及深度伪造内容具有高度逼真性，通过光线，动作捕捉等传统检测方法，已经无法完全区分深度伪造内容。其三，低成本高效率性，在深度伪造技术出现以前对于合成视频技术都需要专业设备与专业人员操作，但现在只需要打开小程序或者 APP 输入照片、音频就能自动生成。其四，快速发展性，从《阿甘正传》中将肯尼迪的影像补充到电影中，再到刚刚发生的韩国 Deepfake 事件，预示着人工智能深度伪造技术迭代周期大大缩短。最初对抗生成网络主要局限于人工智能的研究群体，并未大规模在社会上运用。然而，这种局限却被深度伪造技术的出现所打破。不理解原理，就无法深刻理解风险为何如此独特与严重。

风险并不是抽象的，它必然发生在具体应用过程中，不同的应用场景决定了风险的具体表现形式、危害对象等，了解不同应用场景，才能更好的定位风险。当然深度伪造技术并不都是消极效应，其中也不乏应用广泛的积极效应场景。深度伪造技术的应用场景呈现双刃剑特征：在科研、教育、文娱等领域可创造积极价值，但在刑事犯罪、虚假营销、网络侵权等场景中则会引发严重风险。

从风险社会理论视角看，深度伪造技术正推动社会进入一个新的时代，其风险具有三重特征：其一，隐蔽性，技术黑箱与算法的复杂性使得虚假内容难以被普通用户识别。其二，系统性，深度伪造风险不

再局限于单一的技术问题，而是渗透到个人权利、社会秩序与国家安全的系统性挑战。其三，全球性，深度伪造内容通过互联网快速传播，形成全球性的信息扩散。平台治理理论进一步揭示，在深度伪造技术发展的新模式下，平台不仅是技术应用场所，更是成为风险分配的关键节点，其审核能力，责任分配与治理效能直接决定风险防控的效果。从技术社会建构理论，也即技术的影响由社会场景塑造，本文通过场景分类与风险等级评估，构建深度伪造技术的风险映射体系，明确不同场景下的风险边界与防控重点，如表 1 所示。

Table 1. Application scenarios of deepfake technology
表 1. 深度伪造技术应用场景

积极效应场景	科研领域	辅助自动驾驶仿真系统模拟完成测试
	教育领域	创建虚拟教师，模拟历史人物进行讲解让历史活起来
	文娱领域	换脸技术完成影视作品，创造逼真特效或虚拟角色
	新闻领域	创建虚拟主播进行新闻播报或重现历史事件
	商业领域	虚拟偶像产业或虚拟网红直播带货
	医疗领域	创建虚拟病人进行医学培训，通过深度伪造技术进行疾病诊断
消极效应场景	刑事犯罪	韩国 Deepfake 事件：不法分子伪造普通女性不雅视频，实施胁迫、牟利等犯罪行为，形成规模化网络犯罪产业链。
	虚假新闻与政治宣传	伪造基层公共政策虚假视频(如虚构“取消惠民补贴”“提高民生收费标准”)，引发公众误解与社会秩序混乱；伪造局部地区公共安全事件未通报的虚假信息，制造社会恐慌。
	网络恶搞与社交媒体	伪造普通商家产品质量问题虚假视频(如虚构食品加工卫生不达标)，用于恶意竞争或网络恶搞。

通过对原理及应用场景的理解可以发现深度伪造技术凭借其独有特征，在各类应用场景中，具备了制造以假乱真内容的强大能力。在自媒体蓬勃发展的时代背景下，信息传播模式正快速从中心化向去中心化转变，信息流通速度显著提升[2]，这无疑加剧了深度伪造技术下信息真伪辨识的难度。下面将从个人、社会、国家三个层面分析深度伪造技术的应用风险。

2.2. 深度伪造技术在应用场景下的风险审视

2.2.1. 侵犯公民人身权利与财产权利

深度伪造技术很可能侵犯公民的人身与财产权利。深度伪造技术最早为众人熟知是由于美国网站 Naughty America 利用深度伪造技术，在社交媒体上疯狂传播[3]，这些生成的内容往往是未经他人同意或授权而使用他人肖像或声音的，这就直接侵犯了个人的肖像，隐私，名誉权等权利。除了可能侵犯公民的人身权利外还可能涉及公民的财产权利，引发一系列犯罪，还有犯罪分子利用深度伪造技术合成照片或视频，进行敲诈勒索。还有甚者利用深度伪造技术绑定盗用他人身份，实施诈骗，盗窃等犯罪活动。由此可见，深度伪造技术的应用一定程度上加剧了公民人身、财产权利遭受侵犯的风险，为犯罪提供了空间[4]。

2.2.2. 不利于社会稳定与市场有序运行

深度伪造技术的滥用，对社会信任基础与市场有序运行带来了深层次的冲击。深度伪造技术的发展给各种犯罪活动提供便利，长此以往会引发社会信任危机，加深公民对政府的不信任感，会对真实视频照片产生怀疑，迷失在真假信息之中，对政府的辟谣与澄清也会持怀疑态度。犯罪分子就利用深度伪造技术制造了不少社会恐慌事件[5]。深度伪造技术已经产生扰乱政治辩论，诋毁国家领导人等危机，本在现实生活中不可能发生，但在深度伪造技术的加持下，一切都变得可能，通过深度伪造技术生成图像视

频，让他们“说出”自己原本并没有说过的话，一些不明所以的民众对此深信不疑，由此产生巨大的社会问题。深度伪造技术不仅会对社会稳定有所影响，对于市场经济有序发展也会带来一定冲击。其很可能会使消费者产生信任危机，例如，前文所提及张文宏医生的“带货视频”。此类视频不仅侵犯了他人的肖像权和名誉权，构成对消费者的欺诈，更是破坏市场的诚信经营，消磨消费者的信任。

Table 2. A comparative analysis of deepfake technology regulations across countries and regions
表 2. 不同国家(地区)对深度伪造技术规制对比

国家	类型	具体措施	缺陷	优点
美国	自下而上型	立法方面：① 各州进行立法。得克萨市州《关于制作欺骗性视频意图影响选举结果的刑事犯罪法案》、田纳西州《确保肖像、声音和图像安全法》；② 联邦立法：《保护内容来源和完整性，防止编辑和深度伪造媒体法案》(COPIED ACT)、《2024 年精准伪造图像和未经同意编辑法案》 技术检测方面：《2019 年深度伪造报告法案》中将国土安全部设为深度伪造技术的主管部门；美国国家标准与技术研究院(NIST)被要求制定内容出处信息，水印和合成内容检测的指南和标准 美国以利益为导向，奉行先发展后治理的路径，侧重点也放在深度伪造技术对国家安全的威胁。	第一，各州都可立法，导致没有统一的治理标准。各州与联邦之间，各州与各州之间立法不够统一，在有些州认为是犯罪的行为可能到了其他州就不认为是犯罪了，与此同时也为跨区域治理增加了难度。 第二，治理理念导致监管过于宽松。美国遵循以市场为主导，政府治理为辅的理念，从而形成了自下而上的治理模式，多家公司也承诺采取自愿监管的措施防范技术开发带来的风险，但这种方式太过依赖企业的自觉性，可能会出现很多问题。	有利于新兴科技发展
欧盟	自上而下型	立法方面：《人工智能法案》《通用数据保护条例》《反虚假信息行为准则》(不按照要求删除深度伪造和其他虚假信息，则会处以高达公司全球营收 6%的罚款)、《数字服务法》 欧盟并没有对深度伪造技术进行专门立法，其主要规制方向为深度伪造技术滥用个人生物识别信息行为，更加注重保护个人权益。	第一，没有对深度伪造技术专门立法，只在《人工智能法案》中对深度伪造技术进行相关规定。 第二，较严格的法律阻碍了科技创新发展。 第三，深度伪造技术的监管需要欧盟内部各成员国投入人力，物力，资金等。但各个国家资源配置有所不同，一些小国可能缺乏技术支持或资金投入。	保护个人权益，注重保障人权
韩国	专项治理型	立法方面：2024 年 10 月，韩国政府表决通过三部涉及深度伪造性犯罪的相关法律，分别是《关于处罚性暴力犯罪的特殊法》《关于防止性暴力和保护受害者的法律》修正案、《关于儿童和青少年防性侵的法律》修正案。 治理手段：开展打击深度伪造犯罪的专项行动；重点清查整治制作并传播非法影像的行为；计划开展国际合作；加强对青少年的防范教育。 韩国由于其最近发生的各种社会事件，对于深度伪造犯罪重点打击的方向是性犯罪，对此韩国女性民友会发布《群体性暴力参与者达到 22 万人，社会崩溃要被放任到何时》中提及，韩国是全球在深度伪造淫秽物品领域“最脆弱”的国家。	第一，应用场景局限，韩国专项治理，主要是对深度伪造淫秽影像的规定，对于虚假新闻报道、虚假带货等问题并没有规定。 第二，对于规制内容，保护对象，处罚结果，规定在三部法律中，可能会造成法律使用混乱。	对弱势群体保护，针对性强且符合韩国国情，缓解社会矛盾，同时也填补了法律空白。

2.2.3. 破坏国家安全与国际秩序

深度伪造技术可能会成为信息战的强大武器，破坏国家间的正常关系[4]。深度伪造技术很可能通过伪造国家领导人讲话或发表相关言论等，造成社会混乱国家动荡，也可能制造各种假新闻，迷惑普通民众。除对国家内部安全构成威胁外，深度伪造技术还可能被用于破坏国际秩序[6]。此外，深度伪造技术还可能被用于跨国犯罪活动中(表 2)，深度伪造技术在未来很有可能会影响国家安全与国际秩序。

综上所述，深度伪造技术凭借其强大的伪造功能，已对公民的个人权利、社会与市场秩序的稳定、国家安全与国际关系产生威胁，对于风险的认知，是构建有效治理手段的前提。

3. 深度伪造技术现有治理模式分析

面对深度伪造技术带来的危害，全球各个国家都在积极的治理，各地区因为政治，历史，文化，政策的不同，所采取的措施也有所区别。下面将对域内外治理路径进行对比，分析现有域内政策有何不足之处，域外有何规制措施可以借鉴。

中国目前对深度伪造技术治理更加类似于多方参与型，采用多方参与治理能够最广泛地吸收集体智慧，综合各方意见以提高决策的民主性和科学性，制定更加符合民意的人工智能安全治理制度[7]。早在 2020 年中共中央印发的《法治社会建设实施纲要》中就有提及深度伪造技术的应用问题，紧接着在 2021 年往后国家互联网信息办公室、公安部等部门接连出台相关管理规定，与此同时各类企业、委员会也在积极发布相关报告操作手册。2025 年 3 月更是出台《人工智能生成合成内容标识办法》，想要逐步形成“内容可标识，来源可追溯，责任可追究”的网络空间治理机制。

可以看出我国对深度伪造技术的风险治理一直保持积极态度，出台了不少规范性文件，但其中还是存在一些问题。例如，现有规定的法律规范层次过低，多为倡导性的暂行规定，缺乏约束性的规定；制定文件的主体过多，造成整体的协调性规范性有待提高；现有规定对生成内容的责任分配及投诉机制、监管制度等都有可以完善的空间。其一，关于现有法律规定层次过低，就算是最新颁布的《人工智能生成合成内容标识办法》也只是部门规章，是对《互联网信息服务深度合成管理规定》中标识要求的细化。具体实践中，仍是以《民法典》为裁判的主要依据，《人工智能生成合成内容标识办法》则是判断平台是否承担责任的补充依据。例如，北京互联网法院审理的一起网络服务合同纠纷中，平台用户和网络服务提供者因人工智能生成合成内容标识起了争端。法院依据《民法典》合同编判决平台违约，同时援引《人工智能生成合成内容标识办法》第十条，强调平台需对算法判定结果承担“适度解释义务”，否则担责[8]。目前深度伪造技术中应用较为广泛的还是 AI 换脸技术，其可能涉及到《民法典》中对人格权、肖像权、名誉权、隐私权的保护；《个人信息保护法》中的第 28，29 条；刑法中的侵犯公民个人信息罪，还有相关的行政法规。对于现有法律来说，刑法、民法、行政法缺乏有效衔接，没有进行一体化规制。并且现有的法律也很难做到完全保护公民权益不被侵犯，还是有很多细节没有进行规定。例如，对于肖像权的保护，能不能保护到视频原型的肖像权，原视频原型的外部特征能否得到肖像权的保护；刑法第 253 条侵犯公民个人信息罪中只规定了出售，非法提供或者窃取，以其他方式获取，并不能完全规制制作并出售 AI 换脸视频或者淫秽视频的行为。对于侵犯公民个人信息的犯罪行为的规制类型主要还是以“非法获取、非法提供”为主，如果是合法获取的原视频或者信息呢？是否还能得到保护？其二，责任分配不合理，加重平台负担，《人工智能生成合成内容标识办法》要求内容传播平台对隐式标识进行核验，并对未标识内容二次检测，这样平台就需要处理海量的内容，但现有平台技术检测能力不足，中小平台更是很难做到，并且这种依赖于技术手段的治理方式不可避免的具有技术局限性。此外，该《办法》明确服务提供者，传播平台，用户三方责任，但并没有规定生成端责任，开源模型开发者未被纳入责任主体，形成监管盲区。其三，对于投诉机制而言，《生成式人工智能管理办法》设置了对平台更为严格的投诉

处理要求，平台要采取针对性的阻止措施，但对投诉人并没有严格限制，使投诉的门槛变得极低，这样的投诉机制很可能会造成很多的恶意投诉，甚至会影响健康的网络发展。其四、对深度伪造技术应用的监管层面也有很多问题，监管的主体不统一，责任规定也不够明确[9]，对于此类技术的监管目前涉及多个部门，例如网信部、公安等，很可能造成责任分工不明确，互相推诿的现象。在深度伪造技术应用方面不仅涉及公权力对其的监管，同时也应该注意到网络平台的责任，对于以前的通知－删除义务可能已经不能适应深度伪造技术带来的危害。

4. 中国深度伪造技术的治理建议

目前对于深度伪造技术的规制各个国家(地区)都有其独特制度方法，对此我国也应该找到一条适合本国国情的规制路径，对于前文所提及的现有对深度伪造技术规制的不足，下文将提出改进意见。

4.1. 构建多维法律衔接体系

目前对于深度伪造技术规制主要集中于民法、刑法、行政法规领域，并没有进行专门立法，也没有形成相关的法律法规体系，所以还需要对现有法律法规进行补充与调整。就拿深度伪造技术应用最为广泛的 AI 换脸来说，第一，对于肖像权的保护范围有限，《民法典》对于肖像权的定义为“特定自然人可以被识别的外部形象”，但对“外部形象，可识别”，并没有给予判定标准与解释。对于原视频人的肖像权是否应该保护？AI 换脸虽然只被替换了脸，但原视频可能不止脸具有独特特征，该原视频人的外部形象是否能够通过肖像权得以保护？随着科技进一步的发展，除了人脸信息作为侵犯肖像权的表现外，也应当细化外部形象的具体内涵，出具相关解释，从而更好的保障公民的权益。第二，面对现有侵犯公民个人信息罪采取的规制手段来说，已经不能完全保护个人信息安全，应当在规制“非法获取、非法提供”行为的基础上，增加“非法利用”行为[10]。这是因为对于 AI 换脸淫秽视频来说，公民个人信息的保护力度远远不够，大多只能通过民法得以保护，对非法利用行为进行规制，有助于减轻 AI 换脸淫秽视频的发生。也有学者提出了以“信息保护 + 应用治理 + 平台监管”模式构建刑法规制路径，以强化个人信息保护，增设罪名规制犯罪行为，强化网络平台责任[11]。但在考虑将深度伪造行为纳入刑法规制时，需审慎评估其必要性，避免过度犯罪化。目前还是民法，刑法进行分别规制，希望未来可以做到刑法、民法、行政法等相互配合，通过民刑衔接，刑行衔接来实现对个人信息的保护。因此，可从完善民事保护，优化刑事规制，推动专门立法等方向构建多维法律衔接体系。

4.2. 建立分级责任分配框架

基于平台理论，可构建与平台规模、风险等级相匹配的责任体系，大型平台要加强对隐式标识核验系统，承担主体责任；中小平台对低分险领域进行抽样检测，豁免全面检查的义务；生成端则要设置不可删除的标识模版，从源头控制风险。虽然《深度合成管理规定》中规定了服务提供者与技术支持者的责任，但在《生成式人工智能服务管理办法》中只规定了服务提供者责任，对此应当统一规定责任主体，《人工智能生成合成内容标识办法》的颁布理清了这一规定的混乱之处，但对《人工智能生成合成内容标识办法》中提出的平台核验责任，应当予以限缩，对核验责任采取分层治理，按照平台规模与风险大小进行分级处理。对大型平台强制部署隐式标识核验系统，对中小平台低风险领域的未标识内容进行抽样检测，豁免其全量核验的义务。如果一味追求标识，不与内容风险大小、受众识别能力与传播范围相匹配，很可能会出现适得其反的结果。分层治理对于《人工智能生成合成内容标识办法》的落地实施更为有利，平衡了技术创新与风险防控。其实想要适度减轻平台核验责任，还可以从生成深度伪造内容的源头入手，要求开源模型预设不可删除的标识模版，当用户使用此模版进行创作后，想要消除此标识上

传到传播平台时，内容生成端需对未设置标识内容进行风险提示，否则就要与用户承担连带责任。

4.3. 设计投诉过滤机制

《生成式人工智能服务管理办法》对于投诉–针对性阻止的规定，想要真正适用，还需要加以细化，首先就是要提高投诉–针对性阻止的门槛，如果不能提高门槛，很可能会出现权利滥用的情况，反而不利于人工智能的长远发展[12]。在欧盟的《数字服务法》中也有设立投诉机制，其也是有门槛的，包括对非法内容进行说明，明确投诉人姓名，确认投诉真实性的承诺书等等，这也是增加服务提供者的责任，负担审查投诉内容是否真实的义务。中国可以借鉴欧盟相关规定结合基本国情，设定合理的投诉–针对性阻止条件。第一，在投诉主体方面，要求投诉者提供较详细的个人信息，例如姓名、联系方式等。也可以适当引入身份验证，大大降低恶意投诉的可能。欧盟也提出了相类似的概念“可信标记者”，使用身份证号码与人脸识别进行身份验证，极大地增加投诉的可信度，滥用投诉权利的可能性也会减少。第二，对投诉–针对性阻止措施实施设置等级，面对不同等级实施不同的措施，对投诉进行分级。对于一些没有造成社会恐慌的、没有迷惑性的深度伪造视频，图片，可以设置内部解决机制，直接进行删除等操作；如果是一些造成社会恐慌，涉及犯罪的深度伪造视频，照片的，应当进入严格的投诉处理流程，并同步通知有关部门，如公安，国安等。除了设置等级外，对于一段时间内频繁投诉者，可以限制投诉次数，如真有深度伪造的视频或照片需要投诉的，可以要求提供更多证据。第三，规定投诉者对投诉内容进行说明，对于深度伪造的视频，照片，真的具有违法性才可以。欧盟也有相关规定，要求投诉必须能够定位到具体内容。对于投诉内容提供相关证据，有条件的情况下可以使用区块链技术保证投诉内容不可更改与真实性。

4.4. 创新协同监管模式

深度伪造技术仅仅依靠政府部门的力量很难实现全面监管，就目前中国具体情况来看，监管部门需要统一明确规定，除了依靠政府的力量还需要引入其他积极有效的监管方式。建立政府–平台–用户三方协同机制，强化平台内容审核主体责任，同时提升公众进入网络世界的整体素养。一方面，政府可以要求专业部门制定有关标准，规制深度伪造有害内容，就如美国参议院就提出要求美国国家标准与技术研究院(NIST)制定内容出处信息、水印与合成内容检测指南标准等，政府就从源头进行了监管。对已发生的深度伪造技术犯罪事件，执法时应当区分不同场景，如发生重大深度伪造犯罪事件，对国家，社会可能造成重大危害的，可以借鉴韩国成立专门特别工作组，对专项犯罪展开专项行动；涉及影响国际正常秩序、干扰国家治理、散播不实信息造成社会重大恐慌、伪造淫秽视频等，应当设立更加严格的处罚，就如英国明确将创建，分享深度伪造不雅图像的行为纳入刑事犯罪的范畴，最高可判处两年监禁。如只是恶搞视频或图片，对被害人侵害较小，没有造成对国家，社会的不良影响时，适当监管，同时运用民事，行政法律规定进行规制。科技是把双刃剑，当正向的使用深度伪造技术时，对此的监管可以适度放宽，有利于新兴科技的发展。

另一方面，还可以借助其他方式对深度伪造内容进行监管，传统视听资料的鉴定人员和鉴定方法难以发挥作用，此时鉴定部门拥有精准的检测技术显得尤为重要[13]。首要就是使用算法检测深度伪造技术，利用其生成对抗网络训练一个鉴别器去识别真实的视频，图片，从而区分深度伪造的内容；紧接着在深度伪造内容制造出来时，按照法律规定加上标识，对于未主动标识深度伪造内容的视频，图片的作者应当予以警告，如多次警告仍不标识的应当予以一定处罚，例如禁止关注，封锁账号等。对于删除–通知规则也可以进行一定改变，如前文所说，设置投诉–针对性阻止措施，使得平台监管更加完善与健康。

针对目前深度伪造技术出现的一些问题提出建议的同时还应当以人为本，重视使用深度伪造技术的

人与被深度伪造内容欺骗的人，提高公民的信息素养，认识到散播深度伪造内容造成社会恐慌是违法行为，对于被骗的公民应当提高辨别虚假信息的能力，面对网络信息应当加以甄别，检查其来源，发布者，日期等信息的真实性。法律固然是强有力的规制工具，但要充分降低智能技术的应用风险，唯有技术，法律，伦理规制协同[14]。

5. 结语

深度伪造技术犹如一把双刃剑，带来新的娱乐体验与生活便利的同时，使用不当也会对国家，社会，个人权益产生一定威胁。现阶段对于深度伪造技术治理产生的有关问题，未来治理应摒弃单一依赖法律规制的思路，转向技术、法律、伦理协同的综合治理范式，既防范技术滥用风险，又为创新保留适度空间，最终实现人工智能技术的健康发展与社会秩序的稳健维护。

参考文献

[1] 张桦. 网络空间“深度伪造”的威胁及治理[N]. 网络空间安全, 2020, 11(5): 45-51.

[2] 沈国麟, 易若彤. 从网络社会到平台社会——传播结构的去中心化到再中心化[J]. 探索与争鸣, 2024(3): 156-165.

[3] 邬林桦. 当“眼见不再为实”, 全新风险如何应对? [N]. 解放日报, 2024-02-21(005).

[4] 龙俊, 王天禹. 人工智能深度伪造技术的法律风险防控[J]. 行政管理改革, 2024(3): 69-79.

[5] 龙坤, 马钺, 朱启超. 深度伪造对国家安全的挑战及应对[J]. 信息安全与通信保密, 2019(10): 21-34.

[6] 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[J]. 政法论坛, 2020, 38(4): 144-154.

[7] 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.

[8] 桑雪骐. 平台须对 AI 生成判定结果进行适度解释说明[N]. 中国消费者报, 2025-06-26(003).

[9] 文铭, 孙圆圆. 深度伪造技术应用风险及法律规制研究[J]. 中国科技论坛, 2023(4): 158-167.

[10] 黄陈辰. 大数据时代侵犯公民个人信息罪行为规制模式的应然转向——以“AI 换脸”类淫秽视频为切入[J]. 华中科技大学学报(社会科学版), 2020, 34(2): 105-113.

[11] 李腾. “深度伪造”技术的刑法规制体系构建[J]. 中州学刊, 2020(10): 53-62.

[12] 姚志伟, 李卓霖. 生成式人工智能内容风险的法律规制[J]. 西安交通大学学报(社会科学版), 2023, 43(5): 147-160.

[13] 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.

[14] 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.