刑事司法科学与治理
Criminal Justice Science & Governance
2024 年 第 5 卷 第 2 期

“深度伪造”技术滥用的风险与治理研究

钟一畅

中南财经政法大学，武汉

摘　要 | 随着大数据和人工智能的发展，“深度伪造”技术在现实生活中越来越普遍，滥用“深度伪造”技术的现象也越来越突出。“深度伪造”技术通过对个人生物识别信息进行篡改，深度合成虚假信息。在“深度伪造”当中，人脸图像和声音是使用最多的数据信息。“深度伪造”技术会给个人和国家利益带来极大的危害，必须对滥用技术的行为进行规制。对“深度伪造”行为的规制需要个人、平台、法律制度等多方配合，形成治理合力。

关键词 | 人工智能；伪造；合力；规制；个人信息

2019年，“ZAO”智能换脸软件在网络上迅速“走红”。该软件利用“深度伪造”技术，实现人脸互换。用户只需将自己的照片上传到平台，就可以将视频中的人脸替换成自己的人脸。人们纷纷模仿，将自己的照片上传到平台，体验科技带来的乐趣。然而，娱乐的背后隐藏着危机。不法分子通过收集用户的人脸信息并利用“深度伪造”技术进行人脸替换，进而从事违法犯罪活动。实践中，诈骗、敲诈勒索等犯罪行为也越来越多。面对技术带来的生活变革，我们不得不思考，“深度伪造”技术的边界在哪里，“深度伪造”技术的滥用又该如何预防和治理？

一、“深度伪造”概述

伪造并非一个新鲜的话题，但“深度伪造”却带来一个全新的格局。在“深度伪造”当中，个人生物识别信息成为了原料和素材，而声音和视频则是“深度伪造”的产品。[1]“深度伪造”技术的滥用，导致了信息的虚假。

（一）“深度伪造”概念

“深度伪造”是指，使用AI技术修改数据，最终达到“仿真”目的。[2]从词语的构成来看，“深度伪造”是由深度学习和伪造两个词合成的。“深度伪造”首先在国外掀起了热潮。在2017年，Reddit论坛流传了一段关于好莱坞明星的色情换脸

[1] 李怀胜．滥用个人生物识别信息的刑事制裁路径--以人工智能“深度伪造”为例[J]．政法论坛，2020（4）：150.

[2] 刘霞．机器智能生产：媒介智能融合的溯源、特征与伦理挑战[J]．中国广播电视学刊，2021（5）：19.

视频，这一视频的点击量迅速攀升。“深度伪造”也伴随着换脸视频而逐渐进入人们的视野当中。

与国外发展相似，国内“深度伪造”技术也在一些网站上蔓延。而且，“深度伪造”的对象也往往针对当前知名度较高的明星。2019年，Bilibili视频网站上流传一段关于杨幂的换脸视频。近年来，关于“深度伪造”的热度居高不下，每一次的伪造基本上都会掀起轩然大波，引发舆论关注。国家有关部门也加强了对互联网行业的监管，确保网络秩序的有序进行。

深度合成的虚假信息并非凭空产生，其与个体生物识别信息紧密相连。个人生物识别信息与个人信息的概念不同，个人信息涵盖的范围更加广泛，个人生物识别信息属于个人信息的下位概念。个人生物识别信息具有生物属性和社会属性的特征。个人生物识别信息是对个体生物的记载，具有生物属性。个人生物识别信息的生物属性是其本质属性，能够具体表达生物个体的特点。依赖于个人生物识别信息，我们能够准确区分不同生物。个体作为社会的产物，也让个体生物识别信息具备了社会的属性。当前，人的面部信息、声音、指纹等个人生物识别信息被运用于生活和工作的各个方面。所有的这一切，也赋予了个人生物识别信息以社会属性。

（二）“深度伪造”特征

“深度伪造”作为信息技术的产物，具有高度真实性、简单易操作性和中立性三个方面的特征。对“深度伪造”行为特征的把握有利于因地制宜，准确形成对策。

首先，“深度伪造”作品具有高度真实性。“深度伪造”最为鲜明的特征就是伪造具有高度真实性。图像、音频等的修饰和篡改技术早已有之。例如，通过软件PS就可以将图像进行编辑、修改。但是，这种技术伪造出来的图像大都还是可以识别出来的。随着时代的发展，技术取得了翻天覆地的变化。与之前的伪造相比，高度伪造所依赖的技术更为先进，伪造出的图片、音频、视频等很难用肉眼区分开来。更大的数据、更优的算法，也让合成内容无限接近于真实。[1]而且，“深度伪造”还表现出快速成长的趋势。技术的变革给“深度伪造”带来了适宜的土壤，“深度伪造”也快速发展起来。从当前来看，“深度伪造”呈现出加速发展的态势。[2]

其次，“深度伪造”技术具有简单易操作性。除了高度真实性之外，“深度伪造”最为重要的一个特征就是简单操作性。对图片、音频或者视频的编辑、篡改经历了一个由复杂到简单的过程。早在上个世纪80年代，虽然已经出现了“换脸”特效技术，但是毕竟只有少数人能够掌握。而且，“换脸”的技术操作非常复杂。“深度伪造”技术发展到现如今，已经进入一个相对成熟的阶段。对于一般民众而言，很容易对一张图片或者一段视频进行“深度伪造”。如ZAO软件，只需要用户输入图片，就可以制作出相应的一段视频。可以说，“深度伪造”技术进入了“傻瓜相机”时代，每个人仅需要简单的学习，就可以轻松掌握。这种操作的简便性也推动了“深度伪造”迅速传播。

最后，技术的中立性。人工智能的发展是为了弥补人类信息处理能力方面的不足。[3]“深度伪造”作为一种技术，具有中立性的特征。一方面，“深度伪造”技术能够为个人娱乐、商业开发等提供技术支撑，“深度伪造”在一定程度上具有正向价值。通过“深度伪造”制作的视频，可以让用户享受到独特的体验，增加生活的乐趣。对于商业主体而言，“深度伪造”也可以发挥重要作用。例如，天气预报中的虚拟主播、电商平台中的虚拟客服。通过“深度伪造”，给用户带来更加便捷的服务。此外，“深度伪造”在社会公共服务方面也可以发挥作用。另一方面，如果“深度伪造”技术被滥用，将会导致负面的效果。不仅可能会对个人的人身与财产安全造成威胁，还可能危及国家安全和社会秩序。因此，也有人会以“深度合成”这一更加中性的词来进行替代。事实上，两者在内涵上并没有多大区别。本文采用“深度伪造”这一用语进行表达。

---

［1］熊波．“深度伪造”的扩张化刑事治理风险及其限度［J］．安徽大学学报（哲学社会科学版），2020（6）：107.

［2］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）：60.

［3］赵国宁．智能时代“深度合成”的技术逻辑与传播生态变革［J］．新闻界，2021（6）：68.

二、“深度伪造”技术应用的风险

“深度伪造”具有中立性。如果恰当利用这一技术，会给生活和生产带来很多好处。但是，如果滥用这一技术，则会产生相反的效果。具体来讲，“深度伪造”技术的滥用将带来以下几个方面的风险。

（一）侵犯个体权利

对于个体公民而言，深度伪造可能对其人身权利和财产权利造成侵犯。深度伪造滥觞于虚假色情，其又助推了虚假色情的泛滥。随着深度伪造技术的发展，这一技术的应用也日趋操作简易化和平民化。除了虚假色情内容的传播，深度伪造还可能被用于实施侵犯公民人身权利或者财产权利的其他行为，如诈骗、敲诈勒索等。实践中，通过深度伪造他人的生物信息，实施诈骗他人钱财的行为也逐渐增多，对公民人身权利的侵犯也屡见不鲜。例如，利用“深度伪造”技术制作关于明星的色情视频，从而侵犯明星的名誉权。不管是公民的人身权还是财产权，都是公民不容侵犯的权利，应当受到合理的保护。

（二）破坏社会秩序

“深度伪造”技术不仅对公民的人身和财产利益造成侵犯，还可能扰乱和破坏社会公共秩序。“深度伪造”通过对个体生物识别信息的篡改，制造虚假信息，从而可能破坏社会秩序。当“深度伪造”进入金融领域，可能导致金融领域的动荡。[1]当“深度伪造”的材料被用作证据时，对司法秩序也构成了威胁。[2]再如，制作虚假新闻。在当代，网络已成为人们生活中不可或缺的一部分，生活中的大多信息都是通过网络这一途径获取的。与传统媒介相比，网络的传播速度极快，而且范围广泛。通过“深度伪造”的新闻一旦在网上传播，很快能够形成大规模覆盖，产生不良影响。公众的辨别能力是有限的，特别是面对“深度伪造”的作品。如果虚假的信息占据舆论的主导，社会秩序必然会因此受到影响。此外，“深度伪造”将伪造的虚假信息传递给大众，混淆公众认知，也会产生信任上的危机。[3]

（三）危害国家安全

当前，人们正处于一个信息爆炸的时代。互联网上的信息有时候眼花缭乱，让人难辨真假。公众对信息的筛选产生了疲惫感，对信息的真实性也产生了焦虑感。与以往的信息伪造不同，“深度伪造”的技术水平较高，不仅能够伪造高度真实的图片，还能够伪造高度真实的视频。如果“深度伪造”在某一方面形成爆发点，将会对社会产生难以控制的影响。在美国，“深度伪造”在政治角逐当中的负面作用也表现得十分明显。部分不法分子利用伪造的视频攻击另一政党，从而干预国家选举。“深度伪造”让眼见不一定为实。[4]

正是由于“深度伪造”的滥用会引发公民个人人身权利和财产权利的侵犯，导致社会公共秩序的破坏，甚至在某种程度上影响国家安全。因此，对于“深度伪造”行为的治理迫在眉睫，必须在治理深度和治理强度上加大力度。

三、“深度伪造”治理的困境

当前，“深度伪造”的滥用越来越突出。随之而来的是如何进行规制。在“深度伪造”的规制当中，存在涉及犯罪面广、侦查难度大、前置法难以形成周延保护等多方面的问题。

（一）涉及犯罪面比较广泛

“深度伪造”技术的滥用不仅可能损害个人利益，还可能损害社会利益和国家利益。因此，从法益侵害的角度来看，“深度伪造”行为侵害的社会法益多样，所涉及的犯罪种类较多。从司法实践上来看，“深度伪造”行为所侵害的法益主要包括财产权、人格权、国家安全以及社会公共秩序等几个类别。可以看出，“深度伪造”行为侵犯的法益具有广泛性，跨越多种类型。每一种法益类型下面包括多种犯罪名称。例如，在侵犯财产类的法益当中，可能涉及诈骗罪、敲诈勒索罪等。此外，一般

---

［1］李腾．“深度伪造”技术的刑罚规制体系构建［J］．中州学刊，2020（10）：55.

［2］范玉吉，于雅洁．网络传播中“深度伪造”技术及其产物的刑法规制［J］．犯罪研究，2022（1）：56.

［3］尚海涛．“深度伪造”法律规制的新范式和新体系［J］．河北法学，2023（1）：30.

［4］赵国宁．智能时代“深度合成”的技术逻辑与传播生态变革［J］．新闻界，2021（6）：70.

而言，“深度伪造”犯罪所涉及的被害人众多，既包括社会知名度较高的人物，也包括普通的公众人物。[1]从性别上来看，女性被害人的数量也远远大于男性。“深度伪造”侵害的法益类型多样，被害人涉及面广，这也给社会治理带来一定程度的挑战。

（二）伪造技术水平高，侦查难度加大

如前所述，“深度伪造”具有高度仿真性，伪造技术先进，这也导致在实际侦查过程中难度加大。“深度伪造”技术具有自我优化的功能。随着机器学习能力的提升，一般的检测系统无法对图片、视频的真伪进行识别。[2]在一定程度上，这也加大了对违法犯罪行为的侦查难度。此外，“深度伪造”大多发生在互联网领域，网络的传播速度快，难以及时的跟踪，对于“深度伪造”行为的根源也存在着侦查上的困难。此外，“深度伪造”作品的制作者利用加密技术，可以实现匿名化。[3]事实上，针对“深度伪造”技术的滥用，我国也在逐渐探索破解的方法。实践中，也研发了针对“AI换脸”的视频检测系统。然而，“深度伪造”技术也随之逐步更新，伪造检测系统往往比“深度伪造”慢一个节奏。“深度伪造”技术给司法实践中的侦查带来一定程度的挑战，需要采取合理的方式予以应对。

（三）前置法难以形成周延保护

针对“深度伪造”这一行为，立法上也予以了积极的关注和回应。从部门规章到全国人大制定的法律，均对“深度伪造”行为进行了规定。近年来，网信办、广电总局针对网络音频、视频等信息内容制定了多部规章，明确“深度伪造”行为的行政违法类型。民法典第1019条明确规定，不得通过信息技术手段侵犯他人的肖像权。“深度伪造”技术属于信息技术的一种，民法典的这一规定也是对“深度伪造”行为予以回应。《网络音视频信息服务管理规定》还确立了“标识义务”。“标识义务”主要目的在于从源头上区别信息的真实性。从目前法律对“深度伪造”的规制来看，前置的民事与行政规范对“深度伪造”的调整略显无力。而目前刑罚对“深度伪造”的调整往往也侧重于后果的惩罚，刑罚的谦抑性使得对“深度伪造”技术难以构成完善的规制。在网络时代，特别是针对“深度伪造”的问题，刑法应当从过度谦抑走向有限预防。

四、对“深度伪造”规制的路径展开

对于“深度伪造”行为的治理和规制，应当从多个角度展开。从技术对抗、平台监管、法律规制以及个人预防等多个方面综合发力，确保“深度伪造”技术合理适用。

（一）技术内部应对

“深度伪造”技术作为一种人工智能的应用并非不可突破。从发展上来看，针对任何一种技术的滥用，基本上都可以形成与之相对抗的技术。“深度伪造”技术也不例外。面对“深度伪造”技术滥用的危害，启动技术对抗是一种有效的手段。

传统上，依赖一些基本软件的伪造或者合成技术，通过认真观察，就可以识别出伪造的漏洞。面对“深度伪造”技术，依赖肉眼识别几乎成为不可能。技术对抗成为弥补人体器官不足的重要方法。虽然，“深度伪造”技术也在不断地优化和调整，但是，对抗技术也可以加速灵活调整。实施技术对抗的前提是前技术的存在，因此，技术对抗具有一定的滞后性。然而，这并不能否定技术对抗的作用。目前，要想实现对“深度伪造”作品的准确识别，创建检测系统必不可少。[4]在某种程度上来看，技术对抗具有一定程度的可靠性和准确性。这也是技术对抗存在的最大价值。面对“深度伪造”的加速崛起，内部对抗技术必须加紧脚步，紧随发展。目前，检测和追踪技术逐渐发展起来，与“深度伪造”技术形成博弈和对抗。

（二）确立平台的管理义务

在网络时代，“深度伪造”作品主要借助社交

[1] 邱雅娴．深度伪造型网络犯罪及其治理研究[J]．浙江警察学院学报，2023（6）：88.

[2] 郭泽恩，李蓉．“深度伪造”新型犯罪的危害及治理对策[J]．新疆警察学院学报，2022（2）：53.

[3] 张远婷．人工智能时代“深度伪造”滥用行为的法律规制[J]．理论月刊，2022（9）：121.

[4] 李天琦，刘鑫．深度伪造技术的证据风险和规制路径[J]．证据科学，2022（1）：78.

媒体平台的力量进行传播，对于公民权益的侵犯也产生在传播过程中。因此，对于平台而言，负有相应的审查义务，以保护公民的合法权益。平台作为运营者和利益的获得者，赋予其合理的审查义务就有正当性。平台也应当积极履行这一义务。

一方面，审查“深度伪造”作品的正当性。对于未经授权的“深度伪造”作品，禁止其在相关平台上进行传播。对于个人形象“深度伪造”的视频，应当表明是否为合成作品。有些平台则通过在创作作品上设定标签的形式，防止作品被盗用。例如，在视频或者照片上生成创作的信息或者作者的签名。有些平台则明确要求用户上传视频作品的属性，是原创还是转载。以上各种方式、方法均为平台的基本义务，对作品的审查能够在一定程度上从源头规制“深度伪造”，让伪造的行为自我进行取舍。

另一方面，平台也应当积极引进智能识别技术，用技术对抗“深度伪造”技术，以实现力量上的平衡。例如，运用区块链技术追溯视频来源，实现对视频的拦截和筛选。此外，研发“深度伪造”检测软件，对图片和视频的可信度进行分析。所有的技术研发均以对抗“深度伪造”技术的滥用为目的。通过技术的对抗，从而最大化规避“深度伪造”的技术风险。

（三）刑法的积极规制

在实际生活当中，由于“深度伪造”是以个体生物识别信息作为基础处理的数据，这也导致在特殊情况下出现了身份冒用的情况。因此，有观点认为，应当在刑法中确立身份盗窃的罪名。[1]这种观点的主要理由在于，公民的个人身份是一种新型的法益，刑法有必要予以介入并加以保护。“深度伪造”的本质还是对个人生物识别信息的滥用。在“深度伪造”技术出现之前，公民的个人生物识别信息也存在被侵害的现象，刑法已予以保护。然而，刑罚对公民个人生物识别信息的保护侧重于后端的保护，是法益侵害后的惩罚措施。这种模式的弊端在于没有对前端行为手段进行有效打击。但是在网络时代，“深度伪造”所产生的后果要求刑罚必须从被动走向主动，积极预防“深度伪造”的滥用。“科技向善”的希望之下蕴含着对法律积极规制的需求。[2]因此，笔者认为，有必要增设身份盗窃罪。

（四）自我道德规范与媒介素养提升

在“深度伪造”的治理当中，作为公民的个体也可以有所作为。一般而言，个体可能是“深度伪造”作品的参与者，也可能是“深度伪造”作品的受害者。因此，公民应当双线并行，加强自我道德规范和媒介素养。

一方面，公民应当加强自我道德修养。“深度伪造”技术水平越来越高，公民操作的简易性也随之增强。这意味着，“深度伪造”的门槛也越来越低。现如今，国内外应用程序发行的数量相比之前有很大程度的增加。相应地，“深度伪造”滥用带来的危害呈现出随机、破坏性强的特点。对于个体公民而言，遵守法律规定是最基本的要求，不得滥用“深度伪造”技术。此外，道德规范的约束力也是十分必要的。道德是内心的“法律”，通过不断加强自身的道德修养，能够让公民更加服从法律规范，不侵犯他人合法权益。

另一方面，公民作为社会的重要主体，也是“深度伪造”作品的受众对象。因此，公民在遵守道德规范的基础之上，还需要提高自身的媒介素养。现代社会是信息社会，对信息的识别能力也是必备的修养。公民可以不断学习相关的知识，培养对媒介素材的敏锐力和观察辨别能力。对于网络上的信息保持谨慎的态度，对于有疑问的作品可以通过各种途径进行验证。

五、结论

犯罪的本质在于侵害了法律所保护的法益。“深度伪造”技术具有中立性的特征。如果运用得好，可以为社会带来好处；如果滥用这种技术，则可能给社会带来危害。“深度伪造”的滥用不仅涉及个人利益，还关系到社会公共利益和国家利益。依据我国民法典第1019条规定，任何组织和个人不得利用信息技术手段等方式侵害他人的肖像权。这

---

［1］李怀胜．滥用个人生物识别信息的制裁路径--以人工智能“深度伪造”为例［J］．政法论坛，2020（4）：152.

［2］章琦．深度伪造的刑罚规制路径［J］．中州大学学报，2022（1）：66.

是民法典对“深度伪造”这类技术侵权的直接规定。这也表明我国对“深度伪造”行为的规制提高到了一个新的程度。“深度伪造”行为的危害性要求必须对其进行综合治理。平台、个人以及法律制度应综合发挥作用，形成社会治理的合力，对“深度伪造”行为进行全方位的打击。

（责任编辑：李秀玲）

Research on the Risks and Governance of “Deepfake” Technology Abuse

Zhong Yichang

Zhongnan University of Economics and Law, Wuhan

Abstract: With the development of big data and artificial intelligence, “Deepfake” technology has become increasingly common in daily life, accompanied by a prominent rise in its abusive applications. “Deepfake” technology alters personal biometric information to synthesize deceptive information. Facial images and voices are the most frequently used data in such fabrications. “Deepfake” technology poses substantial threats to both individual and national interests, and the abuse of technology must be regulated. The regulation of “Deepfake” activities requires the joint efforts from individuals, platforms and legal frameworks, forming a comprehensive governance synergy.

Key words: Artificial intelligence; Forgery; Synergy; Regulation; Personal Information