

深度伪造技术在危机传播中对演艺明星名誉的影响与反思

刘睿

河北师范大学新闻传播学院, 河北 石家庄

收稿日期: 2023年12月19日; 录用日期: 2024年1月29日; 发布日期: 2024年2月7日

摘要

深度伪造是基于人工智能深度学习算法, 对音频或文本图像等内容进行制作和修改所生成的目标样品。这在我国公众人物领域也有所涉及, 这给明星名誉带来风险的同时也创造了际遇。本文通过对深度伪造技术在危机传播中对演艺明星的影响进行深入探究, 并提出了几点可行性建议, 旨在为解决深度伪造技术涉及侵犯名誉权、影响演艺明星声誉等问题提供有效的思路。

关键词

深度伪造, 危机传播, 明星名誉

The Impact and Reflection of Deepfake Technology on the Reputation of Entertainment Stars in Crisis Communication

Rui Liu

School of Journalism and Communication, Hebei Normal University, Shijiazhuang Hebei

Received: Dec. 19th, 2023; accepted: Jan. 29th, 2024; published: Feb. 7th, 2024

Abstract

Deepfake is a target sample generated by producing and modifying content such as audio or text images based on artificial intelligence deep learning algorithms. This is also involved in the field of

entertainment stars in our country, which not only brings risks to the reputation of stars, but also creates opportunities. This article conducts an in-depth study of the impact of Deepfake technology on entertainment stars in crisis communication, and puts forward several feasible suggestions, aiming to provide effective ideas for solving problems such as Deepfake technology involving infringement of reputation rights and affecting the reputation of entertainment stars.

Keywords

Deepfake, Crisis Communication, Reputation of Stars

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人工智能深度伪造技术是利用人工智能算法程序，通过合并、替换、叠加图像和视频的方式，创造深度假视频，即“利用机器学习算法，将人脸和声音插入真人的视频和音频，并能够利用整个数字空间创造出逼真的模仿”[1]。Deepfake(深度伪造)一词由“deep learning”(深度学习)和“fake”(伪造)组合而成。Deepfake一词起源于2017年。一名叫做“deepfakes”的Reddit平台用户在该平台发布AI合成的名人换脸视频。此后这种技术被称为deepfake。深度伪造主要基于生成对抗网络(Generative Adversarial Network, GAN)，这是一种人工智能深度学习模型，其工作原理是由两个神经网络相互对抗：第一个用于生成图像，第二个用于判别该输出是否真实。深度伪造的核心思想是利用深度神经网络模型，特别是生成对抗网络(GANs)，来学习和模仿现实中的人脸表情、语音、动作等特征，从而以一种几乎无法辨别的方式将虚构的内容嵌入到真实媒体中[2]。通过大量的训练数据和复杂的算法，深度伪造技术能够生成具有惊人逼真度的虚构内容，使观众难以分辨真实与虚构之间的差异。

美国女学者费恩·班克斯(Kathleen Fearn-Banks, 1941~)认为，危机传播是“在危机事件发生之前、之中以及之后，介于组织和其公众之间的传播”[3]。这类事件可能包括但不限于事故、灾难、丑闻、法律诉讼、产品缺陷等，对组织或个体的声誉、信誉和利益造成潜在威胁。危机传播的关键目标是在面临危机时，通过有效的沟通和信息管理，最大程度地减轻负面影响，维护组织或个体的声誉。在危机传播中，信息的传递和处理变得至关重要，因为公众、媒体和利益相关方对事件的关注度往往很高。

名誉权是个体或组织在社会中所拥有的声望和信誉，是一种宝贵的社会资本。名誉权的形成和维护基于公众对个体或组织的信任度、认可度和评价[4]。这种信任和认可可以建立在多个方面，包括过去的表现、道德操守、专业能力、社会责任感等。演艺明星声誉指娱乐行业工作的个体，如演员、歌手、艺人等，所积累的社会声望和公众对其的评价。这种声誉是通过其在公共场合的表现、作品的质量、道德操守、专业素养以及个人形象等方面建立起来的。演艺明星的声誉对其职业生涯和市场价值至关重要。

随着深度伪造技术的不断发展，出于利益驱使或是追求作品的完整度，自媒体、技术流或是视频制作团队开始利用这种技术并逐步走向成熟。然而在危机传播的背景下，一些不良分子也开始使用这种技术来制造假新闻、不实传闻，一定程度上影响了演艺明星的形象和声誉。

2. 深度伪造技术对演艺明星的名誉影响分析

(一) 深度伪造技术对演艺明星的形象带来的潜在威胁

人工智能深度伪造技术最初的应用场景集中在名人色情虚假视频，俗称为 Deepfake 色情。这一技术被一些不法分子利用，通过 AI 明星换脸的手法，制作淫秽视频并在网络上传播不实新闻和虚假信息以获取流量。这种恶劣行为不仅涉及到侵犯名人的名誉权，也对社会道德和公众认知造成负面影响。更令人担忧的是，这些虚假视频可以根据客户的个人要求进行私人定制，并在网络上以高价售卖。这种商业化的非法活动不仅危害了个体的权益，扰乱了新闻的真实性，也在一定程度上破坏了网络空间的健康环境。

在这一恶性循环中，深度伪造技术的非法应用还涉及到影视剧领域，推动了一股“云交流”的热潮。以强大的粉丝效应为基础，一些技术流派将影视剧的男女主角替换为自己的面孔或粉丝所喜爱的明星，形成一种定制化的观影体验。这引发了对“云交流”趋势的广泛讨论，成为一门备受追捧的生意，同时也探索了人工智能在娱乐产业中新的商业模式。深度伪造技术对形象扭曲和信息误导的能力凸显了数字时代的信息传播脆弱性。逼真的虚假内容可能将演艺明星置于虚构的背景中，使得公众难以辨别真实与虚假。这样的形象扭曲可能深刻地影响公众对明星的认知，甚至在一些情况下，导致明星遭受不公平的批评与指责。在这一背景下，演艺明星的声誉受到了严重的威胁与侵害。

（二）深度伪造技术对演艺明星名誉带来的有益拓展

深度伪造技术的涌现虽带了了不可忽视的负面影响，同时也为演艺明星领域带来了一系列积极的革新。该技术的创新应用促进了表演艺术的进步，在获得演艺明星的授权前提下，深度伪造技术与演艺明星的形象相结合，使演员得以在作品中完成高难度的场景应用，展现更为丰富多彩的表演，将自身巧妙融入多样角色和场景，从而提升了作品的艺术深度与表现力，同时对演员个体的口碑和形象起到了一定的提升作用。值得注意的是，在连续系列的影视作品中，为保持作品的流畅度和剧情的连贯性，制作组会采用深度伪造技术对因不可抗力因素未能参加拍摄的演员进行人工 AI 换脸。这种技术应用曾出现在《速度与激情》第六部中，男主演员保罗·沃克在拍摄期间因车祸不幸身亡，为最大程度挽救损失，电影首次引入 AI 换脸技术。尽管保罗·沃克已经不在人世，但通过 AI 换脸技术，他的形象仍能被继续拍摄，既能够安抚粉丝的内心，也提升了演员本身形象在公众视野中的印象。

同时，深度伪造技术为影视剧的个性化体验和粉丝互动开辟了新的可能性。粉丝通过将自身形象或喜爱的明星融入影视剧情，创造独特而个性化的作品，进一步加强了明星与粉丝之间的互动关系。这种个性化互动不仅加强了粉丝对演艺明星的关注，也提高了粉丝对偶像的粘性，对演艺明星的口碑与价值产生积极影响。深度伪造技术的运用也推动了创新商业模式的勘探，为娱乐产业引入了新颖的商业元素，在这一势头下，创作者发现通过创新方式吸引观众的机会，从而推动了整个产业的发展和革新。深度伪造技术的应用突破了传统娱乐体验的边界，为观众提供了更为沉浸和引人入胜的虚拟现实体验。演艺明星及其作品在数字时代因此更为引人注目。尽管深度伪造技术带来了积极的变革，但在使用上仍需审慎谨慎，平衡技术创新与社会伦理之间的关系，以确保演艺明星的形象和艺术创作能够得到健康持续的发展。

3. 深度伪造技术对演艺明星的名誉保护策略反思

（一）技术层面：加强辨伪技术的检测能力

深度伪造作为一种科技现象，需要科技手段进行有效应对。所以技术层面要引入先进的实时监测技术，例如数字水印技术，允许在媒体内容中嵌入独特标识，使得一旦内容被修改或伪造，监测系统能够迅速识别并警示。或者通过实时监测，演艺明星及其管理团队可以更快地察觉潜在的深度伪造风险，从而及时采取行动。同时，建立系统和定期扫描互联网平台，特别是社交媒体，以检测可能存在的虚假内容，系统采用自动化工具，定期检查各大平台上与演艺明星相关的内容，并通过算法分析内容的真实性，一旦发现可疑或虚假内容，系统能够发出警报，提醒相关团队及时介入。有效监测社交媒体是至关重要的，演艺明星可以通过建立专门的社交媒体监测系统，通过关键词检索、图像识别等技术手段，实时追踪与其有关的

内容,这种监测不仅有助于发现深度伪造,还能提前感知公众舆论动向,为制定应对策略提供参考。在技术层面加强辨伪技术的检测能力,是演艺明星应对深度伪造技术带来声誉威胁的重要一环,通过引入先进技术和建立有效的监测系统,演艺明星及其团队能够更加主动、迅速地应对潜在的形象风险。

(二) 公关层面:提高危机应对的洞察能力

在应对深度伪造技术引发的危机时,演艺明星需要在公关层面展现高度的洞察能力。首先,制定详尽的危机公关计划至关重要。这计划应涵盖紧急声明、争议解释以及舆论风向引导,以确保在危机爆发时能够迅速而有效地做出回应。其中紧急声明是向公众传达真实信息、表达明星个人立场的重要工具,其内容需真实、可信,以迅速平息危机。争议解释需通过各种媒体平台对争议进行澄清和解释,以增强透明度,减少公众对明星形象的误解。舆论引导通过积极的公关手段引导舆论朝着积极正面的方向发展,例如演艺明星可以与媒体合作,通过真实且正面的报道修复声誉,创造更为积极的形象。此外,及时发布真实信息也是关键一环,有助于回应负面信息,及时澄清虚假情报,有效遏制危机扩大。通过这些公关策略,演艺明星能够在深度伪造技术带来的声誉危机中更好地保护个人形象,增强公众信任度。

(三) 法律层面:通过法律手段合理维权

今年8月15日,《生成式人工智能服务管理暂行办法》正式施行。《办法》规定,提供和使用生成式人工智能服务,应当尊重他人合法权益,不得危害他人身心健康,不得侵害他人肖像权、名誉权、荣誉权、隐私权和个人信息权益[5]。其中名誉权根据《民法典》第1024条规定,民事主体享有名誉权,任何组织或者个人不得以侮辱、诽谤等方式侵害他人的名誉权[6]。名誉是对民事主体的品德、声望、才能、信用等社会评价。然而,基于AI换脸等AI合成技术所自带的“伪造”与“欺骗”特性,通过这些技术合成的视频可能导致他人对行为人评价的降低。尤其是当视频内容涉及色情裸露、辱骂、斗殴、吸毒等违反法律法规及大众道德认知的元素时,其构成名誉权侵权几乎是“板上钉钉”的事实。

演艺明星的个人形象与商业价值密不可分,因此发生此类侵权行为时,其维权胜诉的概率相对较大。当演艺明星遭遇“换脸”事件时,可以通过合理的法律手段进行权益维护。其中,寻求法律援助是一项关键举措,以制止和打击侵犯其形象权的行为。这可能包括起诉制作者、传播者或平台,主张侵权行为的追究责任。通过法院的裁决,迫使侵权者承担法律责任,从而达到保护声誉的目的。在这一过程中,法律手段成为演艺明星维护名誉权的有效工具,确保其个人形象和商业价值免受侵害。

4. 结语

生成式AI的发展使得深度伪造模糊了真与假的界限,演艺明星面临着前所未有的名誉保护挑战,而法律体系也在不断迭代中面对新的考验。然而,我们不能一概而论地否定深度伪造技术的价值,因为在某些场景下,它确实为影视制作和广告等领域带来积极的创新。在这一风险与机遇并存的时代,演艺明星需要采取多层次的手段来有效地维护个人名誉权。通过技术手段引入辨伪技术,结合公关策略和法律手段,形成有机的综合保护体系,是应对深度伪造技术带来声誉风险的有效途径[7]。名誉保护不仅仅是个体明星的当务之急,更是维护数字社会信息传播健康发展的不可或缺的一环。因此,我们应在充分认识深度伪造技术潜在风险的同时,促进技术的健康发展,制定更加明确的法规和伦理规范。通过这一过程,我们可以建立一个更加公正、透明、可持续的数字社会,让技术的创新为社会进步和人类福祉提供更多积极的力量。在名誉保护的路上,演艺明星将成为维护数字社会健康发展的先锋,引领着我们迈向更加充满智慧和责任的未来。

参考文献

- [1] Chesneyr, C. (2019) 21st Century-Style Truth Decay: Deep Fakes and the Challenge for Privacy, Free Expression, and

National Security. *Maryland Law Review*, **78**, 884.

- [2] 张灿. 人工智能深度伪造技术的伦理风险与虚无困境[J]. 山东科技大学学报(社会科学版), 2023, 25(2): 11-19.
<https://doi.org/10.16452/j.cnki.sdkjzk.2023.02.006>
- [3] Coombs, W.T. (1995) Choosing the Right Words: The Development of Guidelines for the Selection of the “Appropriate” Crisis Response Strategies. *Management Communication Quarterly*, **8**, 447-476.
<https://doi.org/10.1177/0893318995008004003>
- [4] 曹宋琪. 新媒体环境下公众人物名誉权的限制与保护[D]: [硕士学位论文]. 武汉: 华中科技大学, 2020.
- [5] 国家互联网信息办公室. 生成式人工智能服务管理暂行办法[EB/OL].
https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm, 2023-12-03.
- [6] 全国人民代表大会. 中华人民共和国民法典[EB/OL].
http://www.npc.gov.cn/c2/c30834/202006/t20200602_306457.html, 2023-12-03.
- [7] 彭志红. 深度伪造技术的法律风险及规制研究[J]. 华章, 2022(3): 153-155.