“AI换脸拟声”乱象频发,代表委员建议——

从技术源头多维度协同治理 成立专班精准打击违规行为

AI技术作为新一轮科技革命和产业变革的重要驱动力量,正以前所未有的速度向经济社会各领域加速渗透,赋能新质生产力发展。然而,AI技术在快速发展的同时,深度伪造、算法歧视、数据滥用、自动决策等技术滥用乱象开始凸显。其中“AI换脸拟声”等不当滥用已成为违法侵权行为的重灾区。在今年广州市两会期间,如何为人工智能培育创新沃土、擘画可持续的发展蓝图,也是许多代表委员们的共同关切。针对滥用“AI换脸”,代表委员们建议通过技术源头把控、压实平台责任、强化监管等多措并举,实现系统治理。

■采写:新快报记者 黄闻禹 ■摄影:新快报记者 毕志毅

现象

AI换脸侵权案例频发

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。

但伴随该技术在影视和娱乐行业、社交媒体内容创作、教育培训、身份验证等各领域中的加速应用,其潜在风险也逐渐展露。如使用AI换脸拟声技术给视频中的人物换脸换声,甚至直接生成完全虚构的视频,给不法分子带来可乘之机。

过去一年,众多人大代表和政协委员关注AI换脸拟声乱象,呼吁尽快立法、重点整治,相关讨论引发舆论热议。去年11月,媒体报道了演员温峥嵘遭AI盗播事件,多个直播间出现她的带货形象,温峥嵘曾在直播间质问却遭对方拉黑,再次掀起相关话题讨论。此外,还有不法商家利用公众人物为自家“贴金”,仿冒奥运冠军的声音带货,获取流量。

随着AI深度合成技术的快速发展,市场上还出现了大量“换脸”“换妆”等应用软件,此类AI技术在应用中发生的侵犯自然人人格权益的违法行为日益增多。此外,商家在利用AI技术开展经营生产过程中,若罔顾权利保护意识、法律风险意识淡薄,对包含他人肖像的网上素材盲目利用,也有可能涉嫌违法侵权。

聚焦

AI生成内容须强制亮明“数字身份证”

去年9月1日起,国家网信办、工业和信息化部、公安部、国家广播电视总局联合制定的《人工智能生成合成内容标识办法》(下称《标识办法》)正式生效,《标识办法》提出强制添加显式和隐式标识等规范要求,即用AI生成的每一段文字、每一张图片、每一条音频、视频,都必须强制亮明“数字身份证”。

对此,汉盛律师事务所高级合伙人李旻表示,该规定衔接了此前相关法规的标识要求,进一步细化了实施规范,既未对AI技术创新设置过度约束,又明确了服务提供者、传播平台、用户等各方的责任义务,填补了AI内容的监管空白。

“在AI乱象治理中,其意义尤为关键。”李旻指出,通过全链条标识要求,可以帮助公众快速区分AI生成内容与真实内容,减少虚假信息传播风险;同时为监管执法和权益维护提供了清晰依据,能有效遏制AI换脸、AI谣言等乱象。

代表委员建言

●广州市政协常委、广州市新联会副会长黄丽丽

探索“技术源”“责任源”“认知源”系统治理

广州市政协常委、广州市新联会副会长黄丽丽表示,“AI换脸拟声”被不当滥用问题日益凸显,已成为危害个人权益、公共安全乃至社会秩序的重要隐患。“AI换脸拟声”导致“眼见未必为实”,进一步加剧公众对网络信息的信任焦虑,动摇了社会互信的基础。

加强AI乱象治理,她建议探索“技术源”,支持开发更加高效的识别算法与溯源技术,实现AI生成内容可追溯,同时加强对换脸内容的真实性验证工具研发,为监管部门和平台提供强大的技术支撑;压实“责任源”,明确技术平台和内容传播平台的审查义务,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,并在出现违法或高风险内容时及时预警和处置;提升“认知源”,通过普法宣传、公益科普等形式,普及AI换脸滥用的法律风险与辨别方法,引导公众自觉抵制违规内容,主动举报滥用行为。

她还建议坚持“以人为本、智能向善”理念,健全AI行业自律体系。同时,进一步出台、完善针对人工智能技术研发、应用和监管的专项法律法规,明确责任边界,细化监管标准,厘清AI系统中的算法透明性、可解释性和责任归属问题。加强对数据安全与隐私保护的监管,确保数据获取和处理的合规性,避免技术成果滥用和侵权。

●广州市人大代表、广州市创新社区治理发展研究院管委会主任张茜

组建AI深度合成治理专班,精准打击违法违规行为

“AI换脸拟声本是能赋能影视、文创等领域的好技术,但当下被滥用制造虚假广告、实施电信诈骗、侵犯他人肖像权的情况频发,不仅损害公众的合法权益,也扰乱了社会发展秩序。”张茜认为,技术发展的初衷是造福民生,而非制造伤害,这种滥用行为必须被严厉遏制。

针对AI换脸技术应用门槛较低、隐蔽性强等特点,她建议从源头、成体系地治理乱象。首先,压实技术提供方责任,督促相关平台、企业落实《人工智能生成合成内容标识办法》,给换脸拟声内容强制加数字水印和显式标识,从技术端做到可溯源。其次,强化平台审核义务,推动短视频、社交平台升级AI鉴别技术,建立快速下架、举报响应机制。最后,联动执法,由网信、公安、工信等部门组建AI深度合成治理专班,破解监管分散难题,精准打击那些隐蔽的违法违规行为。

敲重点

AI换脸不是想换就换,当心“换来”法律责任

李旻表示,随着生成式人工智能用户规模的持续扩大,其不当滥用已成为违法侵权重灾区。此类滥用行为呈现侵权门槛低、危害范围广、侵权形式多元隐蔽性强、违法成本与维权成本失衡的显著特征,不仅直接侵害公民肖像权、声音权、名誉权等人格权益,还可能滋生诈骗造谣等违法犯罪行为,扰乱网络空间秩序与社会公共利益。

未经授权AI换脸,构成个人信息权益侵害

“AI换脸技术的核心是对人脸生物信息的处理与合成,其滥用直接触碰个人信息权益及人格权保护底线。”李旻指出,根据《人脸识别技术应用安全管理办法》规定,处理人脸信息需以显著方式完整告知处理目的、保存期限等事项,并取得个人“单独同意”;《互联网信息服务深度合成管理规定》第十四条进一步明确,提供人脸编辑功能的,需提示使用者告知被编辑个人并取得其单独同意。违反上述规定,未经同意擅自使用他人人脸信息进行换脸,或未履行告知义务、超范围处理人脸信息的,构成对个人信息权益的侵害,需承担停止侵害、删除信息、赔礼道歉等责任。

禁止采用技术手段删除或篡改AI换脸内容标识

针对AI换脸使用者,《互联网信息服务深度合成管理规定》第六条明确禁止利用深度合成服务制作、传播法律禁止的信息,或从事侵害公共利益、扰乱社会秩序的活动,使用者若通过AI换脸制作虚假信息、淫秽内容等,构成违反治安管理行为的,由公安机关给予治安管理处罚;该规定第十八条同时禁止采用技术手段删除、篡改AI换脸内容的标识,违反此规定的,将由相关主管部门依法处罚,情节严重的可限制或禁止使用相关服务。

滥用AI换脸技术,情节严重者或承担刑事责任

此外,当AI换脸滥用行为达到“危害国家安全、损害公共利益或情节严重侵害他人权益”的程度,将依据法规中的刑事衔接条款追究刑事责任。李旻指出,《互联网信息服务深度合成管理规定》第二十二条、《人脸识别技术应用安全管理办法》第十八条均明确,违反规定“构成犯罪的,依法追究刑事责任”,为刑事追责提供直接依据。