

智能语音合成技术背景下声音权益侵犯与保护

孙梓益 徐海婷

上海政法学院, 上海

摘要 | 人工智能技术在近年经历着重大突破, 人工智能程序和产品使用方式逐渐向简单化、日常化方向发展。随着AI技术浪潮来袭, 出现了大量的AI生成物, 衍生出了大量的法律问题。例如使用智能语音合成技术合成明星歌手的声音进行“AI翻唱”, 这种行为使用被模仿者的声音, 利用被模仿者的价值和影响力获取个人利益。针对自然人声音权益的保护, 我国《民法典》参照适用肖像权保护的有关规定, 确立了声音权益的保护地位。但是, 在科技飞速发展的今天, 对声音权益侵犯的方式越来越多, 情况越来越复杂。基于此现象, 应当明确声音权益保护的细则, 增强法律实施的系统性和规范性, 加速声音权益系统化保护建设进程。

关键词 | 语音合成技术; 人格权益; 法律保护

Copyright © 2025 by authorx (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

<https://creativecommons.org/licenses/by-nc/4.0/>



2024年4月23日上午, 北京互联网法院对全国首例“AI声音侵权案”进行一审宣判。原告殷某是一名从事配音工作多年的配音师, 2023年5月, 殷某发现自己的声音被一款名为“魔音工坊”的App通过AI化以后, 以“魔小璇”的名义对外出售, 并且这款声音产品已经在多个短视频平台被用户使用。经审理, 北京互联网法院认为, 原告声音权益及于案涉AI声音, 最终法院认定作为配音师的原告, 其声音权益及于案涉AI声音, 被告方使用原告声音、开发案涉AI文本转语音产品未获得合法授权, 构成侵权。2024年4月2日, 国家互联网信息办公室官方网站发布关于《生成式人工智能服务已备案信息》的公告。随着生成式AI产业的迅速崛起, 对相关的服务提供商将实施更为严格的监管要求。这些要求主要涵盖了算法设计、训练数据的选择、从模型到内容产出的全过程, 以及用户实名认证和个人隐私、商业秘密的保护

等方面。生成式AI产业即将迎来更加规范和有序的监管环境。

1 AI技术的现实应用与研究问题缘起

人工智能的发展下, AI翻唱迅速产生并被广泛使用。2023年4月, 一条“AI孙燕姿《发如雪》翻唱”突然出现在媒体平台上并获得了大量关注。截至2024年4月, 其在哔哩哔哩上的翻唱歌曲播放量达到了325万次, 并且在多个平台都有视频或音频, 收获大量播放量, 给账号主体带来一定的经济收益。目前大多数AI翻唱制作人是粉丝群体, 为了听到他们喜欢的歌手演唱自己喜欢的音乐而制作, 但是伴随着可观的流量收益, 势必会对歌手、词曲著作权人、录音录像制作者等主体造成侵权^[1]。

通讯作者: 孙梓益, 上海政法学院学生, 研究方向: 民商法学。

文章引用: 孙梓益, 徐海婷. 智能语音合成技术背景下声音权益侵犯与保护 [J]. 社会科学进展, 2025, 7(8): 664-668.

<https://doi.org/10.35534/pss.0708112>

1.1 AI 语音合成技术

智能语音技术的核心构成包括自然语言处理技术、AI语音合成技术以及语音识别技术。其中，AI语音合成技术对于声音的拟真能力，构成了对声音权最主要的侵权风险来源。一般的语音合成流程包含三个关键模块：文本前端，负责将原始文本解析为字符或音素序列；声学模型，将字符和音素序列映射为对应的声学特征；声码器，依据声学特征合成出最终可听的声音波形。

由于声音权的人格权属性，其保护范围应包含声音本身的物理特性，包括但不限于音调、音色、振幅、音高等，以及由这些物理特性构成的视听载体。声音所传递的具体表达内容及其表达形式则不属于声音权保护的直接客体。当前AI语音合成技术的核心挑战在于，它能够精确模仿目标说话人的关键声学特征，如音调、音色、振幅甚至情感韵律，生成与目标声音高度相似乃至难以辨别的合成语音。这种能力极易导致第三人身份的混淆，对声音权保护构成了严重的威胁。

1.2 AI 语音合成技术在现实生活中的应用

声音不仅是我们进行表达的媒介，也是接收信息的重要途径。信息化时代，人们接受信息的方式大部分为数字媒体，阅读和收听成为信息获取的主要途径。随着生活节奏的加快和对效率的追求，语音智能产品被更广泛地使用。

(1) 影视娱乐

当前社会娱乐产业发达，名人明星具有极高的影响力，所以与普通入相比，名人明星的声音也具有更高的经济价值，将明星的声音合成语音用于商业活动，成为一种新型的智能语音使用大势。例如，利用AI语音合成的明星声音进行广告和动画、影视片段的配音与旁白，提高了作品制作效率，降低了时间和经济成本。在游戏领域，智能语音合成技术也进行广泛应用，互动游戏中利用AI语音合成，合成角色CV的声音应用到更广泛自由的游戏剧情中，能更好地提高游戏对玩家的吸引力，提高玩家的喜爱度和满意度。

(2) 智能语音播报

智能音响可以说是最被广泛使用的居家智能产品，产品分类甚至可以精细到不同年龄阶段，“小度”“小爱”等智能音响几乎成为有些家庭的固定成员，深度融入人们的日常生活场景。人们使用语言对智能系统下达“命令”，通过对人类自然语言的识别和处理，了解指令的意图并且找到答案作出动作或者回答。百度地图推出各种“导航语音包”，提取名人明星甚至动画和游戏角色的声音，合成各种导航指令进行语音播报，这种将用户喜爱的声音IP与实用功能深度绑定的策略，极大地增强了产品的趣味性和情感连接，有效刺激了用户的下载欲望，显著提升了软件的下载量和使用活跃度，成为声音经济变现的一个成功范例。

(3) 阅读教育

智能语音技术在教育领域的渗透正日益深入，展现出强大的变革力量。在信息爆炸的碎片化阅读时代，用户对信息获取的便捷性要求空前提高，人们甚至不愿分出精力“使用眼睛进行阅读”，喜马拉雅、小宇宙、懒人听书等各种播客软件 and 有声阅读App数量激增，下载量和活跃度屡创新高。敏锐捕捉到这一需求转变，传统阅读类App如微信读书、起点读书等也纷纷战略性地嵌入“听书”或“AI朗读”功能，将海量文字内容转化为可听的语音流，无缝适配通勤、家务、休息等多场景学习与娱乐，极大拓展了知识的可及性和接收效率。某英语听力练习网站成功“克隆”了多位当红明星的声音特征，生成了高度拟真的明星语音包，并将其应用于英语听力材料的念白。虽然网站课程还在不断丰富优化中，但是已经吸引大量粉丝进行下载学习。

2 自然人的声音权益

2.1 声音的特质

声音就是由物体振动发出的，通过一定介质进行传播，并且能够被人或动物听觉器官察觉的波。人声是由气流使声带振动产生机械波，波的长短等特性决定了声音在音量、音调、音色等方面的不同。科学研究表明，声音具有独特性，因为每个人发声器官的不同，每个人的声音都是独特的，具有强烈的个人属性。即使有些情况下同个体发出的声音会“听起来一样”，但是其声波是各不相同的。在此前提下，声音成为个人的一种独特的标记，是识别不同个体的重要方式，所以声音具有人格属性。从法律层面来看，声音是每个自然人人格的组成要素，彰显了个人的人格尊严，是每个人作为人格主体的重要特征，并可以用来识别每个自然人的身份^[3]。

2.2 声音权法律属性

声音权是指自然人独有的、与生俱来的权利，即自由支配其声音并通过这种方式实现与声音相关的各种利益。声音权的客体是声音利益，即声音所承载的人格利益，兼具精神与财产双重属性。其精神属性源于声音与人身不可分离的专属性，以及使得声音成为人格识别的重要标识的个体独特的声纹特征。同时，声音亦具财产属性，权利人可通过声音的使用获取经济利益。

当前，我国并没有明确声音权的概念，在我国现行法律中，通过适用肖像权规定来实现对声音权益的保护，目前的保护并不足以形成具体的声音权。虽然《民法典》将声音参照肖像权保护，但是声音与肖像完全是两个完全不同的概念，声音权与肖像权的物理特征和客体都是不同的。《民法典》第一千零二十三条只关注到声音利益和肖像权的相似之处，忽视了二者之间的差异，简单地给声音适用肖像权有关条款，是难以有效发挥对声音权益的保护的。机械适用肖像权规则，将难以

有效明确声音权的法律边界,从而损害司法实践的稳定性与法律权威性。当前的主流学说倾向于将声音识别为一种具有特殊性的人格利益,并主张通过法定人格利益保护机制对其进行规制。

3 人工智能时代背景下声音权面临的威胁

人工智能时代的演进,显著提升了语音合成技术的应用水平,不仅有力驱动了声音元素商业价值的开发,也促进了智能语音技术的纵深发展。然而,该技术的广泛应用,潜藏着对自然人声音权益构成侵害的风险,并可能导致个人隐私信息的不当泄露,损害名誉权等问题的出现。因此,在开发和使用AI语音合成技术时,应当及时防范和预防可能出现的风险和问题,有针对性地提出解决问题的措施,在享受科技发展带来的利益的同时,减少和避免科技带来的道德和法律问题。

3.1 AI 语音合成技术出现之前

在AI语音技术出现之前,主要通过剪辑、录音、未经允许使用等方式行为方式侵犯声音权。由于传统媒体传播速度慢、范围窄的特点,以及声音本身“无形”的特性,声音权的侵犯现象发生较少,并且较少产生严重的损害后果,主要以侵犯著作权、商标权等侵犯知识产权的案例为主,未能引起对声音保护的重视。这一时期的司法实践和学术讨论中,声音权(声音利益)往往未能作为一个独立、显性的法律权利受到充分重视。相关的纠纷和诉讼,常常被“包裹”在著作权法、商标法、不正当竞争法甚至人格权中的名誉权框架下进行处理。随着人工智能的出现和深度发展,语音合成技术日趋成熟,传播效率大幅提升,合成成本显著下降,使语音合成技术应用场景显著扩大。伴随智能语音产业的蓬勃发展,声音侵权案件呈现高发态势。

3.2 AI 语音合成技术出现之后

(1) 窃取他人身份

声音作为一种独特的生物识别特征,具有高度专属性和个体唯一性,与特定自然人身份紧密关联。通过声音识别,能够有效确定个体身份。AI语音合成技术能合成出与模仿对象相似,并且具有相似的物理特征的语音,完美地再现他人的声音。这种技术的滥用,使不法分子可利用合成的语音冒充他人身份,进而实施侵害他人财产或损害其人格尊严的违法行为。

(2) 损害他人的名誉

利用人工智能语音合成技术,模仿特定自然人声音并以其身份发表不当言论,会使该个体形象受损、社会评价降低,侵犯其名誉权。AI语音合成技术合成的音频会冒充他人发表不正当的言论,不仅破坏他人社会形象,还会影响社会环境。例如,合成政治家声音,利用政治家身份发表各种极端政治言论,不仅破坏政治家形

象,也会通过对选民的影响扰乱政治活动,不利于社会稳定;或者冒充企业家发表不正当言论,影响公司的声誉和股东对公司的信任,威胁公司的生存。

(3) 侵犯他人的财产权

①利用语音合成进行诈骗

利用合成语音进行财产诈骗最常见的方法是,利用合成的声音向模仿对象的亲朋好友发送借钱的语音信息,利用他人对模仿对象的信任,骗取钱财,进行财产诈骗。2019年发生一起著名的诈骗案,犯罪分子使用AI语音生成软件冒充英国能源公司德国母公司的CEO,成功欺骗同事和合作伙伴,一天内诈骗并转移资金,导致公司损失220,000欧元^[4]。该英国CEO被诱导向匈牙利供应商转账,认为来电者是真正的德国老板。反欺诈专家Rüdiger Kirsch表示,英国CEO因来电者口音和声音旋律相似而轻信。当“德国老板”再次要求第二次付款时,由于转账未到账且电话来自奥地利,公司行政部门开始怀疑并拒绝支付。

②未经允许合成他人声音进行商业活动

人格权要素可纳入商业开发范畴,权利人通过订立合约,授权他人使用自己声音的使用权,据此获取经济利益。智能语音合成技术的商业化已出现在多个领域,例如地图导航语音包,许多导航地图软件与公众人物进行合作,通过录制和合成他们的声音,推出特色导航语音服务,吸引用户下载使用。同时,一些App甚至能够让用户自己录入语音,合成语音包。

同时,大量“AI翻唱视频”在各大娱乐平台出现。去年,某个技术团队开发的模仿孙燕姿的AI技术模型,基本完全复制了孙燕姿的音色,AI孙燕姿生成的歌曲铺天盖地,借由孙燕姿的音色特点,“AI孙燕姿”演唱了《爱在西元前》《半岛铁盒》《水星记》、Melody等不同风格的歌曲,引来网友的听歌热,并且引来孙燕姿本人的关注。在国内某视频播放平台搜索“AI翻唱”“AIcover”等词条,筛选过后发现使用智能语音合成技术制作的翻唱视频最高的播放量已经达到七百万,该视频发布于2023年12月18日,还不到半年时间。

智能语音合成技术在艺术创作领域和商业领域发挥着越来越重要的作用,负责任地使用这一技术,AI具有的巨大潜力会增进人类创造力,推动行业发展。但不幸的是,随着人工智能技术的飞速发展和平民化普及,一些平台和开发者正在使用AI去损害创造力并削弱艺术家、词曲作者、音乐家和版权所有者的利益。在声音权益受到侵害的情形下,自然人依法享有请求损害赔偿的权利。相较于普通个体,公众人物的声音因其蕴含更高的商业价值,更易成为未经授权商业利用的对象,从而引发诸多声音侵权纠纷。此类侵权行为实质性地阻碍了权利人通过声音实现财产价值,损害了声音权的财产性利益,故有必要通过法律途径对此类行为予以规制。

4 国内外声音权保护现状

4.1 国外声音权保护现状

第一,以美国为例的公开权保护模式。公开权的主体是自然人,是与隐私权相对应的权利,是公开自然人的私人信息,对此进行保护和商业利用的权利。姓名、肖像、声音都是公开权的客体^[5]。在声音权益保护中,美国的公开权主要是对声音的财产属性进行保护,利用声音取得经济收益。

第二,以德国为例的一般人格权保护模式。《德国民法典》强调财产权、合同自由、经济权利和财产损害赔偿,财产法的显著特征强调了人们对财产的合法权利,而人格权制度在《德国民法典》起步很晚。德国民法关于人格权制度的规定远远滞后于财产权制度,只在侵权责任法第823条中,对部分人格利益作了明确的规定。此外民法典第12条规定了姓名权;第824条规定了信用权;第825条规定了贞操权。由于德国民法典对人格权规定的滞后,德国联邦最高法院在德国基本法第1条、第2条的基础上,发展确立了一般人格权相关制度,并将一般人格权视为联邦基本法所保护的一项基本权利,具有法律的约束力。同时联邦最高法院把基本法意义上对于一般人格权的解释当作德国民法典第823条第1款的“其他权利”对待,并利用大量的司法判例,将名誉、隐私、肖像和荣誉权等,作为一般人格权的一部分加以保护,从而确立了德国民法典一般人格权制度。在德国除了成文法已经规定的人格权,还包括法院在过去或未来能够通过判例发展出的新权利类型,包括肖像、隐私和声音等。德国的声音权得到了广泛的保护,其中包括通过判例将声音权视为一般人格权进行维护。

第三,加拿大在保护自然人的声音权方面,主要通过隐私权的方式进行,特别是在魁北克省,没有对声音权独立立法。魁北克省民法典第36条规定了侵犯他人隐私的行为,其中包括“盗用或者使用他人的肖像或者声音”,这表明在魁北克省,声音被视为个人隐私的一部分,受到法律的保护。在加拿大,尤其是魁北克省,隐私权的外延相对较广,包括姓名、肖像、声音等多种个人特征。这与一些其他国家或地区将隐私权、肖像权、姓名权等作为并列的具体人格权有所不同。这种保护模式将声音视为个人隐私的一部分,禁止未经许可的使用。然而,如果声音的使用是出于向大众合理公开信息的需要,那么这种使用是被允许的。加拿大的这种保护模式体现了对自然人声音权的重视,并为其提供了法律上的保障。

4.2 国内声音权保护现状

我国《民法典》首次将声音权益纳入民法保护,通过“参照适用肖像权保护的有关规定”对自然人的声音权益加以保护,但是未将声音权作为独立的具體人格权形式在法律上加以确认。《民法典》实际上是将声音

列为一种独立的新型人格要素,把声音作为一种人格利益,参照肖像权的保护模式对声音进行法律保护。既没有将声音作为一种明确的具体人格权加以保护,也没有将声音仅仅作为一般人格权,而是对声音这种新型的人格利益采取独立的人格利益保护模式。

5 声音权作为具体人格权保护的正当性

5.1 人工智能技术发展下声音利用方式不断扩大

随着科学技术的发展、人工智能技术的飞跃,人们对于声音的储存、编辑与使用进入了一个新的阶段。自然人声音的录制更加真实,音频编辑更加流畅自然,对自然人声音的深度模仿的训练更加简单快速。这些都是科技社会发展的自然结果,扩大了声音使用的范围,给人们的生活带来方便,增加了自然人声音所附加的经济价值。在声音的商业价值不断被挖掘的同时,其被侵害的可能性也在大幅度增加。近年来,未经授权利用知名人物声音进行广告宣传,合成自然人语音进行电话诈骗等事件已屡见不鲜;通过现代声音处理技术、智能语音合成技术和深度学习技术对自然人声音的模仿以及利用其进行创作的侵权行为更是不断发生。因此,为了尽量减少科技发展带来的“双刃剑”影响,将声音利益确立为声音权进行保护,是法律制度下一步的完善路径。

5.2 我国声音权益立法保护存在进一步发展空间

我国《民法典》中人格权独立成编是一个创新之处,是对人格权立法的完善和对自然人人格尊严的保护。《民法典》对于肖像权的保护进行了较为具体的规定,这是基于自然人肖像独特的人格属性决定的。声音与肖像相比,虽然不如肖像对于不同个体的独特性展现得直观,但是经过人工智能技术的“放大”后,已经取得甚至超越肖像带来的特殊符号感。我国《民法典》将声音参照肖像权进行保护虽然合理,但是在人工智能技术不断发展优化和声音商业价值的不断开发下,此条款对自然人声音权益的保护明显不够充分。

在《民法典》正式颁布之前,我国对于声音权益的保护主要侧重于其财产利益方面,这种保护主要通过《商标法》和《著作权法》来实现。然而,值得注意的是,商标法中所提及的声音商标与人格权中的声音权在性质上是有显著区别的。在《商标法》的框架下,声音商标所涵盖的声音既可以是自然人的声音,也可以是一段旋律。它允许声音以商标的形式存在,作为商业标识的一种。只要声音具有足够的独特性和区分性,它就可以被注册为商标。这种声音商标的保护,实际上更多的是关注声音作为财产权的商业价值,它属于知识产权的范畴,主要用于保护声音在商业活动中的经济利益。然而,与声音商标不同,人格权中的声音权更多地关注的是声音的精神利益。它强调的是声音与特定自然人的身

份、形象和尊严之间的紧密联系。在《民法典》颁布之前，由于法律体系中对于声音权的规定相对不足，声音的精神利益往往没有得到充分的保护和赔偿。因此，尽管《商标法》和《著作权法》在一定程度上为声音提供了法律保护，但它们在保护侧重点和范围上存在差异。声音商标主要保护的是声音的财产权益，而人格权中的声音权则更侧重于声音的精神利益。随着《民法典》的颁布，我国对于声音权益的保护应当更加全面和深入，进一步加强对声音精神利益的保护和赔偿。

在《著作权法》的框架下，对于声音利益的保护主要通过表演者权等间接方式进行，这种方式允许将非自然人的声音纳入保护范围，甚至扩展到了通过人工智能技术模仿产生的声音。然而，尽管这种保护机制覆盖了一定范围内的声音，但它并不是全面的。具体而言，表演者权等间接保护措施确实为声音的权益提供了一定程度的保障，但它们的保护范围往往受限于特定条件和情境。这意味着，并非所有声音都能得到《著作权法》的充分保护，只有符合特定要求的声音，如与表演活动直接相关的声音，或者通过特定技术手段（如人工智能技术）创造的声音，才能受到该法的保护。因此，虽然《著作权法》在声音保护方面取得了一定的进展，但仍存在局限性。它无法对所有声音进行全面保护，尤其是在保护自然人声音的精神利益方面，仍需要进一步完善和加强。随着技术的不断发展和社会的不断进步，我们期待未来能够建立起更加全面、有效的声音保护机制，以更好地维护声音权益的合法性和正当性。

6 结语

《民法典》将声音权益纳入立法保护，不仅是我国立法完善过程中的重要一步，更是对社会发展产生极大正面影响得举措。通过立法完善声音权益保护的相关制度，将声音权益保护与时代发展背景相结合，更好地发挥对自然人人格权益保护的作用。科技与法律是现代文明的双翼，面对不断进步的科技世界，实现高水平科技自立自强，离不开法治保障。人们更应该审慎思考如何制定和完善相关法律法规，以保障人们的安全和权益，同时推动技术的创新和应用。以高质量法治供给护航科技创新，巩固和加强科技创新取得的重大成就，降低科技带来的不良影响。

参考文献

- [1] 曾梦, 陈娜. AI翻唱的法律风险研究[J]. 江南论坛, 2024(1): 67-72.
- [2] 尚增强, 张鹏远, 王丽. 融合跨说话人韵律迁移的多语种文本到波形生成[J]. 声学学报, 2024, 49(1): 171-180.
- [3] 王利明. 论声音权益的法律保护模式[J]. 财经法学, 2024(1): 3-20.
- [4] 张卓群, 李娜. 人工智能生成内容技术在内容安全治理领域的风险和对策[J]. 通信管理与技术, 2024(2): 53-54.
- [5] 杨立新, 袁雪石. 论声音权的独立及其民法保护[J]. 法商研究, 2005(4): 103-109.

Voice Rights Infringement and Protection with Advances in Intelligent Voice Synthesis Technology

Sun Ziyi Xu Haiting

Shanghai University of Political Science and Law, Shanghai

Abstract: In recent years, artificial intelligence technology has witnessed groundbreaking breakthroughs, with AI programs and products becoming increasingly simplified and integrated into daily use. Along with the surge of AI technology, a vast amount of AI-generated content has emerged, giving rise to numerous legal issues. For instance, the use of intelligent voice synthesis technology to replicate the voices of celebrity singers for “AI cover songs” appropriates the voices of the imitated individuals, leveraging their value and influence for personal gain. In response to the protection of natural persons’ voice rights, China’s Civil Code extends the provisions applicable to portrait rights to establish the legal protection of voice rights. However, in the context of rapidly advancing technology, methods of infringing upon voice rights are becoming more diverse and complex. Given this phenomenon, it is imperative to clarify detailed rules for the protection of voice rights, enhance the systematic and standardized implementation of relevant laws, and accelerate the development of a comprehensive framework for the systematic protection of voice rights.

Key words: Speech synthesis technology; Personality rights; Legal protection