

人工智能合成语音二创短视频的权利冲突与法律规制

张怀远

宁波大学马克思主义学院, 浙江 宁波

收稿日期: 2026年8月6日; 录用日期: 2026年8月17日; 发布日期: 2026年9月8日

摘要

随着生成式人工智能与短视频二次创作深度融合, AI语音二创在降低创作门槛的同时, 也引发声音人格权益、著作权及个人信息权益等复合冲突。本文结合实际案例与司法裁判, 围绕训练、生成、传播三个环节分析法律边界, 并以人格识别利益、数据利用利益和商业开发利益构建“三阶法律结构”。在此基础上, 文章以“可识别性”为声音人格利益判断的重要边界, 结合著作权合理使用、训练数据合法利用、平台责任与风险分级提出规制路径, 以期兼顾人格保护、技术创新与二创产业发展。

关键词

著作权法, 合成语音, 短视频, 知识产权, 人工智能

Rights Conflicts and Legal Regulation of AI-Synthesized Voice-Based Derivative Short Videos

Huaiyuan Zhang

School of Marxism, Ningbo University, Ningbo Zhejiang

Received: August 6, 2026; accepted: August 17, 2026; published: September 8, 2026

Abstract

With the deep integration of generative artificial intelligence and short-video secondary creation, AI voice-based secondary creation has lowered the threshold for creation while also giving rise to complex conflicts involving voice-related personality rights and interests, copyright, and personal information rights and interests. Drawing on real-world cases and judicial decisions, this paper analyzes legal boundaries across the three stages of training, generation, and dissemination, and

文章引用: 张怀远. 人工智能合成语音二创短视频的权利冲突与法律规制[J]. 法学, 2026, 14(9): 115-126.

DOI: 10.12677/ojls.2026.149263

constructs a “three-tier legal structure” comprising personality identification interests, data utilization interests, and commercial exploitation interests. On this basis, “identifiability” is taken as an important boundary for determining voice-related personality interests. In conjunction with copyright fair use, lawful use of training data, platform responsibility, and graded risk governance, the paper proposes regulatory pathways aimed at balancing personality protection, technological innovation, and the development of the secondary-creation industry.

Keywords

Copyright Law, AI-Synthesized Voice, Short-Form Videos, Intellectual Property, Artificial Intelligence

Copyright © 2026 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

人工智能语音合成技术能够通过声音克隆、文本转语音等方式学习并再现特定个体的音色、语调和发音风格。随着短视频平台发展，AI 翻唱、AI 配音等二创内容大量出现，在丰富数字文化供给的同时，也引发声音权益、著作权及平台责任等法律争议。

人工智能合成语音二创所涉及的法律关系具有明显的复合性。现行法对声音相关利益的保护并非集中于单一制度，而是分散于人格权、著作权及邻接权、个人信息保护等规范体系。语音生成也并非对既有录音的机械复制，而是通过算法提取声音特征并重新合成新的语音内容，使声音在不同环节分别呈现人格识别、数据处理与商业开发属性。《中华人民共和国民法典》¹(以下简称《民法典》)第 1023 条第 2 款明确规定对自然人声音的保护参照适用肖像权保护的有关规定；《中华人民共和国著作权法》²(以下简称《著作权法》)则分别规范音乐作品、表演及录音制品等权利客体；声音数据在符合个人信息构成条件时，还可能受到《中华人民共和国个人信息保护法》³(以下简称《个人信息保护法》)调整。与此同时，《互联网信息服务深度合成管理规定》⁴(以下简称《深度合成规定》)、《生成式人工智能服务管理暂行办法》⁵(以下简称《暂行办法》)以及 2025 年 9 月 1 日起施行的《人工智能生成合成内容标识办法》⁶(以下简称《标识办法》)已从训练数据、内容管理和生成合成内容标识等方面形成行政治理规则。由此，人工智能语音二创中的侵权构成、规范竞合、责任主体与救济路径，均有必要在多法域交叉的框架下加以考察。

从国内研究看，围绕人工智能合成语音及其二创风险，现有文献已从协同治理^[1]、域外制度^[2]、AI 表演的著作权定性^[3]与 AI 翻唱风险^[4]等方面展开直接讨论。在声音权益的法律性质上，学界仍存在明显分歧。王利明认为，《民法典》并未创设独立的具体“声音权”，而是采取法定独立人格利益的保护模式^[5]；汪倪杰则主张，声音作为人格要素在人格权法中产生一般人格权益，并可与《著作权法》《个人信息保护法》《中华人民共和国反不正当竞争法》⁷(以下简称《反不正当竞争法》)等特别法结合形成表

¹<https://www.court.gov.cn/zixun/xiangqing/233181.html>

²https://www.gov.cn/guoqing/2021-10/29/content_5647633.htm

³http://www.npc.gov.cn/npc/c2/c30834/202108/t20210820_313088.html

⁴https://www.gov.cn/zhengce/zhengceku/2022-12/12/content_5731431.htm

⁵https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm

⁶https://www.gov.cn/zhengce/zhengceku/202503/content_7014286.htm

⁷https://www.samr.gov.cn/zw/zfxgk/fdzdgknr/fgs/art/2023/art_3737890d856a4e44a8ea07c50c90c116.html

演者权、个人信息权益与竞争法权益，由此需要处理人格权法与特别法之间的规范竞合[6]。在商业利用层面，王利明从人格权商品化角度肯定标表性人格要素的经济利用价值[7]；房绍坤、曹相见则对人格权商品化的一元解释提出批判，主张将人格标识使用利益理解为源于人格促销价值、含有人格利益的特殊财产权[8]；孙雅琪进一步从智能媒体产业发展出发，反思人格权模式对声音商业利用的过度覆盖，主张区分人格尊严保护、人格标识商业利用与普通自然人声音数据利用[9]。由此可见，声音人格利益与财产性利用利益究竟如何划界、衔接，仍是 AI 语音场景下尚未完全解决的基础理论问题。

围绕 AI 语音的具体利用行为，既有研究又形成了不同分析进路。陈杰从 AI 表演及录音产业关系出发，将 AI 翻唱理解为对特定表演者的技术性模仿，强调其与一般机器学习对抽象、非特定对象的模仿存在差异，并据此讨论知识产权规制与创作自由的平衡[10]；北京互联网法院课题组则从人格权司法认定出发，将自然人声音以及 AI 生成声音的“可识别性”置于侵权判断的核心位置[11]；陈笑春、李泽林从声音生产与创作的纵向层级讨论自然人原声、声音数据与 AI 合成声音之间的主体及权益配置[12]。在训练数据著作权问题上，李安主张区分非表达型与表达型机器学习并实行梯度化规则[13]，刘晓春则从训练行为的非特定性、生产过程性等特征出发，提出“非作品性使用”的合法化路径[14]。上述研究对一般生成式人工智能训练具有重要解释力，但针对以特定自然人为直接模仿对象的声音克隆与 AI 翻唱，训练数据、人格识别和后续商业利用往往同时发生，不能当然套用一般模型训练的单一结论。总体而言，既有成果已分别在人格权、知识产权、数据利用和产业治理等方面形成较为充分的讨论，但对 AI 语音二创中多种法益如何在“训练-生成-传播”全过程交织，以及不同创作目的与技术路径如何改变主导性法律风险，仍有进一步整合分析的空间。

基于此，本文仍以“人工智能合成语音应当保护什么、如何认定侵权、如何构建治理体系”为主线，在不改变现有权利体系的前提下，尝试作两方面整合：其一，将人格识别利益、数据利用利益与商业开发利益概括为“三阶法律结构”，用以说明同一声音要素在不同利用阶段可能对应不同规范基础及其竞合关系，该结构属于法益识别与规范适用的分析工具，并非创设三项新的法定权利；其二，在既有“训练-生成-传播”全过程分析基础上，引入娱乐翻唱、新闻播报、商业数字人和恶意冒充等不同使用目的，使侵权判断与规制措施能够进一步对应具体场景。本文的贡献主要在于对既有分散研究进行结构化整合，而非主张以单一部门法重构声音权利体系。受篇幅与研究对象所限，本文主要讨论民事权益保护及与之密切相关的平台治理，对刑事责任、跨境数据流动及具体算法技术标准等问题仅作必要涉及。

2. 人工智能合成语音二创短视频的概念和种类

2.1. “二创”短视频的概念

“二创”短视频是指创作者利用其他创作者的原创作品作为基础视频素材所进行的二次创作短视频[15]。视频平台中多以游戏、体育、影视、原创短视频等作品素材为基础，通过片段剪辑、原著解说等形式呈现。与完全原创的短视频不同，二创短视频通常融入二次创作者本身的理解，在原有作品或素材基础上形成新的内容表达。需要说明的是，本文所称“二创”主要是对短视频平台内容生产方式的事实性描述，并不当然意味着相关成果构成《著作权法》意义上的改编作品。在生成式人工智能技术介入之前，二次创作已成为短视频平台中较为常见的内容创作模式，而人工智能技术的发展进一步降低了语音类二创的门槛，使人工智能合成语音二创短视频成为具有代表性的二创类型。

2.2. 人工智能合成语音的概念

“AI 技术涌入传统音频合成场域，借助数据标注和情感分析为视听产品赋予情感化特征，并通过数

据驱动和算法迭代赋能产品创新升级。”^[1]传统有声出版完全依赖人类的配音,长期以来维持着成本高、效率低、难更改的状态,导致市场供需失衡。而人工智能语音合成技术的出现则突破了该瓶颈,文本可自动转换为高质量的语音,大幅降低了生产成本还提高了效率。技术的革新降低了有声内容生产的准入门槛,使中小出版机构和独立创作者也能够以较低成本参与市场竞争,极大地丰富了有声内容市场的供给,也为不同类型创作者开展内容二次创作提供了便利。而人工智能合成语音相比早期的语音合成技术,更接近自然人发声的语调、情感与风格。相关对比研究表明,人工智能合成语音与自然语音在部分声学表达特征上仍存在差异^[16],但其拟真程度的提升也使其适用的场景从传统有声出版拓展至短视频二次创作等多个大众内容生产领域。

从法律分析角度,还需要区分自然人声音、声音载体、声音数据与 AI 合成声音。关于声音利益究竟应定性为法定人格利益^[5]还是具体人格权^[6],学界并无一致结论。本文采用“声音权益”这一相对中性的概念,主要指自然人围绕具有可识别性的声音所享有的人格利益,以及声音进入表演、数据处理、商业利用等特定场景后由相关特别法确认的合法权益。录音制品等只是声音固定和传播的载体,其相关著作权或邻接权并不当然等同于自然人的声音权益;声音经数字化后,如属于与已识别或者可识别自然人有关的信息,则可能进入《个人信息保护法》的调整范围,其中声纹等生物识别信息属于《个人信息保护法》第 28 条所列敏感个人信息范畴,只有在具有特定目的、充分必要性并采取严格保护措施的情形下方可处理;依该法第 29 条,处理敏感个人信息还应取得个人的单独同意,法律、行政法规规定应当取得书面同意的,从其规定;AI 合成声音则属于算法基于声音样本特征生成的新语音,其是否仍受特定自然人声音权益约束,应重点考察公众能否通过音色、语调、发音风格等特征将其与该自然人建立稳定联系^{[11][12]}。

2.3. 人工智能合成语音“二创”的主要类型和表现形式

“AI 语音合成技术重塑了有声出版行业的生产模式、服务模式和产业结构,带来从效率到生态层面的多重跃迁。”^[2]生成式人工智能语音二创是指利用深度学习、大语言模型与神经网络语音合成技术,对既有的声音样本进行学习与建模,在不直接复制原始录音的前提下生成具有特定主体声音特征的新语音,并进一步结合短视频、数字人或虚拟主播等内容形式进行再创作与传播的数字内容生产方式。传统录音复制表现为对既有录音文件的直接拷贝、剪辑或拼接,而生成式人工智能语音则不同,它首先通过深度神经网络算法去提取声音样本中的多维声学特征,包括音色、音高、韵律、语速、停顿方式及情绪表达等方面,并将这些特征转化为参数化的数学模型。用户输入新文本后,模型依据设定好的参数去重新合成对应语音。因此,AI 语音生成属于以学习声音特征为基础的再合成,而非传统意义上的录音复制。而人工智能合成语音技术的迭代在现实应用领域催生出了全新的业态。

从技术实现方式看,AI 语音二创主要可区分为声音转换与基于模型的文本转语音、声音克隆等路径。声音转换通常保留既有语音中的音高、节奏或演唱信息,再以目标主体的音色特征进行覆盖,AI 翻唱多采用此类路径;文本转语音或声音克隆则先以目标主体的声音样本训练声学模型,再根据新输入文本生成近似其音色、语调和发音风格的语音。陈杰对 AI 翻唱的技术流程进行分析时,将其概括为由“训练素材-声学模型-音色覆盖或合成-新录音”逐步完成的过程^[10]。这说明,AI 语音在技术上虽非对某一原始录音的简单机械复制,但训练与生成环节仍可能分别涉及自然人声音、音乐作品、表演、录音制品以及个人信息,技术上的“重新合成”并不当然排除法律上的权利利用。

按使用目的,AI 语音二创可分为娱乐翻唱与配音、新闻及公共信息传播、商业广告与数字人运营、恶意冒充等场景。技术相近并不意味着风险相同:前者侧重声音与版权冲突,新闻传播强调身份真实性,

商业使用关注授权边界，恶意冒充则可能危及交易与公共秩序。

AI 语音合成在网络视听等领域具有降低成本、提升个性化体验等积极价值^[17]。法律规制应围绕是否指向特定自然人、是否利用受保护作品、是否处理个人信息及具体传播目的进行差异化判断，而非否定技术本身。

3. 人工智能合成语音二创短视频侵权的界定与范围

3.1. 人工智能合成语音二创短视频侵权的界定

相较于一般二次创作短视频的著作权纠纷，AI 合成语音二创的风险具有更明显的权利复合性与流程延伸性。2024 年北京互联网法院审结全国首例 AI 生成声音人格权侵权案，案涉自然人声音被 AI 化处理为文本转语音产品，并进一步被用于短视频内容，争议贯穿声音素材提供、AI 化处理、产品销售与终端调用等环节^[18]。本文所称人工智能语音二创短视频侵权，是指在训练声音模型、生成语音内容以及制作、传播二创短视频的过程中，相关行为违反人格权、著作权及邻接权、个人信息保护等规范并侵害他人合法民事权益的情形。其判断对象并非抽象的“AI 使用行为”，而应具体识别行为人在不同阶段是否未经许可利用具有可识别性的自然人声音，是否使用受《著作权法》保护的作品、表演或录音制品，是否违法处理相关个人信息，以及是否存在法律规定的许可、合理使用或其他免责事由。因此，需要结合人工智能语音合成技术的运行逻辑，区分训练、生成与传播阶段分别评价，既不能将合理的技术使用和创作使用当然认定为侵权，也不能因最终生成内容未直接复制原始录音而忽视前端数据处理或后端传播中的权利侵害。

3.2. 人工智能语音二创短视频侵权的范围

人工智能语音二创的侵权范围首先取决于其所涉及的法益类型与具体利用环节，而不能由一般二次创作短视频的宏观侵权数据直接推定。鉴于一般二次创作短视频的侵权数量、账号规模和点击量等统计并不能直接反映 AI 合成语音这一细分类型的发生规模，本文不以此类宏观数据作为风险程度的直接依据，而将分析重点置于专门司法案例所揭示的权利链条与具体利用环节。就 AI 合成语音自身而言，至少涉及人格识别利益、数据利用利益与商业开发利益。三者具有不同的规范基础，却会在模型训练、内容生成与平台传播过程中相互交织，本文将其概括为 AI 合成语音利益的“三阶法律结构”。需要强调的是，这一结构属于分析意义上的法益分层，并非主张在现行法之外创设三项新的独立权利。

第一阶：人格识别利益。自然人在长期语言实践中形成相对稳定的音色、语调、语速和发音风格，声音由此具有指向具体人格主体的识别功能。从法律属性看，本文倾向于将该利益理解为以《民法典》第 1023 条第 2 款为规范基础的法定人格利益，而不另行创设独立的“声音权”；但应当承认，学界围绕声音权益究竟属于具体人格权、一般人格权益^[5]还是法定人格利益仍存在不同解释^[6]。无论采用何种理论路径，“可识别性”均是 AI 声音进入人格利益保护范围的重要前提。北京互联网法院课题组提出，可从一般社会公众或者相关领域公众的主观识别、声纹与音色语调等客观特征以及具体使用方式等方面综合判断 AI 生成声音是否仍指向特定自然人^[11]。因此，AI 生成语音与原声音并不需要达到完全一致，只要足以使相关公众建立稳定的人格联系，即可能触及声音人格利益。

从法理上看，人格识别利益之所以构成“三阶结构”的基础，并不只是因为声音能够辨认身份，更在于声音与特定自然人的人格同一性和自主决定之间存在稳定联系。王利明将《民法典》第 1023 条第 2 款理解为对声音确立“法定的独立人格利益”保护，强调声音虽不是独立具体人格权，却可以不依附姓名、肖像等其他人格标识而获得单独保护^[5]；汪倪杰则认为，声音作为人格要素首先产生一般人格权益，同时还可能在表演、商业竞争和个人信息处理等场景中形成特别法上的利益^[6]。两种观点虽然在权利性

质上并不相同,但都说明声音保护的第一层逻辑仍然是维持声音与人格主体之间的法律联系。因此,本文在现行法框架下采用“人格识别利益”的表述,既避免把“声音权益”直接等同于一项独立的声音权,也为后续特别法利益的识别保留空间。

第二阶:数据利用利益。声音样本进入模型训练后,声音由人格标识进一步呈现为可以被收集、存储、分析和建模的数据资源。但本文所称“数据利用利益”并不意味着存在抽象的“声音数据所有权”。对能够识别特定自然人的语音数据,应首先依《个人信息保护法》判断其处理活动是否具有合法基础;若涉及声纹等生物识别信息,则应依《个人信息保护法》第28条、第29条适用敏感个人信息的特别保护规则。陈笑春、李泽林指出,声音数字化并被纳入产品开发与流通之后,声音数据来源者、采集者、产品开发者 and 语音库经营者之间会形成多元主体的利益交集[12]。与此同时,若训练素材同时包含受《著作权法》保护的音乐作品、录音制品或其他作品,数据处理合法性与知识产权合法性还应分别审查,不能以其中一种合法性当然替代另一种。

数据利用利益与人格识别利益的区别,主要体现为规范评价对象由“声音作为人格标识”转向“声音数据的处理行为”。声音被数字化并进入算法模型后,其法律问题不在于自然人是否对数据享有抽象的所有权,而在于采集、存储、训练、调用和删除等具体处理活动是否具有合法基础,以及数据处理者、声音来源者和模型开发者之间的利益如何配置。陈笑春、李泽林即从声音生产与创作的纵向层级指出,自然人声音数字化后,会与采集者、产品开发者、语音库经营者等多元主体发生数据权益交集[12]。因此,第二阶并非第一阶人格利益的简单延伸,而是声音进入数据处理关系后产生的独立规范问题;但当训练数据仍可识别特定自然人,尤其涉及用于身份识别的生物识别信息时,数据利用的效率目标仍受人格尊严和个人信息保护规则约束。

第三阶:商业开发利益。具有较高辨识度和社会影响力的声音可以被用于广告、游戏配音、有声读物、直播及数字人运营,因而形成稳定的市场利用价值。围绕这一价值的法律属性,学界存在人格权商品化与人格标识使用权等不同解释。王利明强调标表性人格要素具有可以授权利用的经济价值[7];房绍坤、曹相见则提出,人格标识使用权源于人格的促销价值,可被理解为含有人格利益的特殊财产权[8];孙雅琪进一步主张应防止人格权保护对声音商业利用空间形成过度挤压[9]。基于这些争论,本文将商业开发利益理解为声音人格要素进入市场利用后形成的财产性利益,其保护不能脱离人格主体的授权、具体使用目的和市场利用方式,更不宜简单表述为对声音享有“所有权”。

商业开发利益的法定定位则集中体现了人格权与财产权理论之间的张力。王利明在讨论人格权商品化时区分德国模式与美国公开权模式,前者仍在人格权制度内部处理人格要素的经济价值,后者则将具有财产价值的部分从人格权中分离并赋予独立保护[7];房绍坤、曹相见进一步批判将人格标识的决定、变更与商业使用全部纳入同一人格权结构,主张人格标识使用权属于“含有人格利益的特殊财产权”[8];孙雅琪在AI声音语境下也提出,应当避免人格权保护对声音商业利用空间形成过度挤压,并区分知名人士声音的促销价值与普通自然人声音的数据利用价值[9]。考虑到本文旨在解释现行法下AI语音二创的复合风险,而非重新建构人格权体系,本文不对上述一般理论争议作终局选择,而是将商业开发利益作为第三层分析对象:其产生以声音进入市场利用为条件,在具体适用上仍应分别借助人格权许可规则、合同规则以及著作权、竞争法等既有规范确定权利边界和利益分配。

3.3. 三阶法律结构的相互关系与冲突顺序

三阶法律结构并不是三项彼此割裂的权利清单,而是同一声音要素进入不同社会关系后呈现出的复合法益。人格识别利益是声音与自然人主体保持法律联系的基础;声音被采集、数字化并用于算法训练时,进一步产生个人信息处理与数据利用问题;声音被投入广告、配音、数字人等市场活动后,又可能

形成商业开发利益。三者既具有递进性，也可能在同一行为中同时存在。

发生规范冲突时，不宜机械确立“人格-数据-商业”的绝对等级，而应根据被侵害法益和具体行为为进行分层适用。第一，使用行为直接影响人格尊严、身份真实性与人格自主的，应当优先确保人格利益不因技术利用而被架空。第二，声音利用已经落入《著作权法》《个人信息保护法》《反不正当竞争法》等特别规范调整范围的，应先依其构成要件和责任规则进行评价，在特别法无法覆盖的声音人格利益范围内，再由《民法典》人格权规则提供保护，这也与学界关于一般法与特别法协调的观点相契合^[6]。第三，不同权利对象的许可不能相互替代：取得录音制品或作品的利用许可，并不当然意味着取得将特定自然人声音进行 AI 化处理的许可；反之，取得声音主体的许可，也不当然取得相关音乐作品、表演或录音制品的利用权。全国首例 AI 生成声音人格权侵权案正体现了区分录音制品相关授权与自然人声音 AI 化授权的必要性^[11]。

由此，“优先次序”应理解为规范适用和利益保护的先后逻辑，而不是抽象宣告三类利益在价值上存在固定高低。第一步，应先判断声音利用是否已经触及人格尊严、身份真实性或人格自主等不可被市场交易完全吸收的核心人格利益；一旦触及，应确保人格主体的基本控制不因数据训练或商业合同而被架空。第二步，在人格核心利益得到保障的前提下，再根据具体利用行为进入相应特别法规范：作品、表演和录音制品利用适用《著作权法》，个人信息处理适用《个人信息保护法》，具有市场竞争意义的人格标识利用则依据其具体构成考察竞争法规则。第三步，对于同一行为同时满足多种规范构成要件的情形，应以各规范实际保护的法益和救济功能区分请求权，避免因规范竞合产生重复评价或重复赔偿。汪倪杰关于人格权法与特别法之间一般法、特别法关系的讨论，为这一适用顺序提供了重要理论参照^[6]。

在商业开发与创新利用之间，则应进一步考察授权范围、使用目的、识别程度、市场影响与公共利益等因素。对于以特定自然人为直接模仿对象并产生明显身份关联的商业利用，应提高授权与透明度要求；对于不指向特定自然人、仅进行一般性技术训练或非商业创作的情形，则应保留合理的技术创新与表达空间。由此，三阶结构的功能不在于扩大权利边界，而在于避免以单一部门法处理复合型 AI 语音争议。

三阶结构因此可以概括为“人格联系奠基-数据处理展开-市场利用延伸”的动态关系：人格识别利益回答“声音为什么仍然属于特定的人格主体”，数据利用利益回答“声音被转化为训练资源后如何合法处理”，商业开发利益则回答“声音进入市场后如何进行许可、收益分配与竞争秩序保护”。这种分层并不要求每一个 AI 语音二创案件同时具备三阶利益，而是要求裁判与治理在具体案件中识别实际被激活的层次，再选择相应规则。对仅涉及一般合成音色且无法识别特定自然人的应用，人格识别层可能并不成立；对具有高度可识别性但纯属非商业表达的使用，商业开发层也未必成为主导；只有在特定声音被数据化、模型化并进一步投入商业传播时，三阶利益才可能同时叠加。

4. 人工智能合成语音二创短视频侵权的案例研讨与分析

AI 语音二创的法律风险横跨人格权益、著作权及邻接权、个人信息权益等领域，同时受到生成式人工智能和深度合成治理规则的约束。目前司法实践已经出现具有代表性的声音权益裁判规则，但相关规则仍处于发展过程中。2023 年，“AI 孙燕姿”式翻唱集中进入公众视野，相关创作者通常以既有歌手人声样本训练音色模型，再将生成的目标音色用于其他歌曲翻唱，形成了具有高度可识别性的 AI 声音二创⁸。这一现象表明，对 AI 语音二创的法律判断不能仅以“是否高度相似”作统一结论，而应分别审查生成声音对特定自然人的可识别性、训练和生成阶段是否利用受保护作品或录音制品、传播行为是否取得必要许可，以及相关服务提供者和传播平台是否依法履行内容管理、标识与处置义务⁹。

⁸<https://finance.people.com.cn/n1/2023/0526/c1004-32695583.html>

⁹<https://ent.people.com.cn/n1/2023/0520/c1012-32690711.html>

4.1. 人工智能语音二创短视频创作表现及其法律边界

AI 语音二创的风险具有明显场景差异。为避免对娱乐表达、公共信息传播、商业运营和欺骗性使用采取同一强度的评价,本文结合使用目的与主要技术方式作四类分析。

上述分类以使用目的为主轴,以技术是否定制模仿特定自然人为辅助判断。陈笑春、李泽林将人工智能合成声音区分为基于语音大模型生成的通用声音与针对目标发音人单独训练的定制声音:前者训练样本来源较为庞杂,输出音色通常不指向特定自然人;后者则需要以目标发音人的声音数据进行单独创建,更容易形成稳定的身份关联[12]。因此,即使处于同一种应用场景,技术实现方式不同也会改变主导风险:使用通用合成音色时,争议可能主要集中于作品利用、内容真实性与标识义务;使用可识别的特定自然人音色时,则会进一步触发声音人格、个人信息处理与商业授权等问题。

4.1.1. 娱乐翻唱与一般娱乐二创

娱乐翻唱通常以声音转换或声音克隆为主要技术路径。如果以既有演唱录音作为输入并进行音色覆盖,可能同时涉及目标歌手声音人格利益、词曲著作权、原录音制作者权以及原表演者相关权益;如果由创作者自行演唱或直接依据文本、曲谱生成新录音,则对原录音和原表演的利用程度可能降低,但对目标歌手可识别声音以及音乐作品的利用仍需分别判断。陈杰对 AI 表演的技术分析表明,音色转换与直接从文本生成录音在权利关系上并不完全相同[10]。因此,该场景的核心不是笼统判断“AI 翻唱是否侵权”,而是识别具体技术流程中实际使用了何种权利客体,并结合《民法典》和《著作权法》规定判断许可、法定许可或合理使用等规则是否适用。

4.1.2. 新闻播报与公共信息传播

新闻播报、公共服务提示等场景多通过文本转语音或声音克隆,使新文本以主持人、播音员、专家等特定主体的声音呈现。此时,著作权问题往往不再是唯一或主导风险,公众是否误认为内容系该自然人本人发表、信息是否真实以及 AI 生成属性是否得到充分提示更为关键。尤其当合成声音被用于具有公共影响力的信息传播时,声音所承载的身份识别功能会被显著放大,因而应强化身份真实性、内容审核与显著标识要求。

4.1.3. 商业广告与数字人运营

商业广告和数字人运营往往以声音克隆、文本转语音与数字形象生成相结合,并具有持续调用模型、反复生成内容的特点。该场景的主导风险集中于人格标识的商业化利用及授权边界,包括授权目的、使用场景、使用期限、是否允许模型训练及再许可等。一次性配音或录音制品制作授权不应被当然解释为无限期、全场景的声音模型开发和商业调用授权。对于具有稳定市场价值的声音[11],还需要在保障人格主体控制力[9]的同时处理商业开发利益的许可[8]与收益分配问题[7]。陈笑春、李泽林的研究还指出,部分专业配音人士对文本转语音技术的核心担忧,正集中于个人声音模型的应用领域和使用期限[12]。这说明商业数字人场景中的授权审查不宜停留在一次性录音许可,而应进一步明确模型训练、持续调用、使用期限、适用场景以及再许可等生命周期边界。

4.1.4. 恶意冒充与诈骗

恶意冒充通常利用实时变声、声音克隆等方式,将技术目标直接指向身份欺骗。与普通娱乐二创相比,此类行为不仅可能侵害声音、姓名、名誉及个人信息等民事权益,还可能直接危及交易相对人的财产安全和网络公共秩序,符合相应构成要件时还可能进入行政法、刑法的规制范围。因此,其风险等级和平台处置强度应明显高于一般娱乐性使用,不能以“二次创作”或“技术中立”作为当然的责任减免理由。

由此,四类场景的主导性风险和规制重点并不相同:娱乐翻唱与一般娱乐二创以作品利用和特定声

音的可识别性为中心,宜重点审查权利客体、授权范围、合理使用边界及 AI 标识;新闻播报与公共信息传播以身份真实性、信息误导和公众误认为中心,应提高内容来源审核与显著标识强度;商业广告与数字人运营以授权目的、期限、再许可和持续模型调用为中心,应强化合同边界与商业用途核验;恶意冒充与诈骗则以身份欺骗、财产安全和公共秩序为中心,应采取更严格的技术阻断、快速处置以及行政刑事衔接。场景化区分的目的并非另立新的权利类型,而是据此确定同一组法律规范在不同使用目的下的主导风险和治理强度。

4.2. 人工智能语音二创博主的侵权表现分析

责任认定还应沿“训练-生成-传播”链条展开。不同阶段的对象和规范不同,不能把全过程风险笼统归结为“未经授权使用声音”,也不能因生成物未机械复制原录音而忽视其他环节的违法利用。

4.2.1. 训练阶段:声源数据采集与模型训练

实践中, AI 语音模型的训练素材可能来自公开发布的专辑录音、演唱会音频、采访语音,也可能直接调用第三方共享的音色模型或预训练模型。数据已经公开并不当然意味着可以不受限制地用于模型训练。《个人信息保护法》第 27 条允许在合理范围内处理个人自行公开或者其他已经合法公开的个人信息,但个人明确拒绝的除外;但处理已公开的个人信息对个人权益有重大影响的,应当依照该法规定取得个人同意。若训练涉及声纹等生物识别信息,则还应依第 28 条、第 29 条适用敏感个人信息规则。对于落入《深度合成规定》适用范围且提供人脸、人声等生物识别信息编辑功能的服务提供者和技术支持者,第 14 条第 2 款还要求其提示使用者依法告知被编辑的个人,并取得其单独同意。著作权层面,学界对于训练数据使用是否必须取得许可存在明显争议:李安主张依据非表达型与表达型机器学习实行梯度规则[13],刘晓春则认为不针对特定作品的生产过程性训练可以被理解为“非作品性使用”[14]。但这些观点主要针对作品的大规模、非特定训练,若 AI 语音模型以模仿某一特定自然人为直接目的,则还同时触及声音人格与个人信息问题,不能简单以一般训练数据理论替代人格权益判断;陈杰也特别指出,对特定人的 AI 模仿与一般的文本、数据挖掘存在不同的利益结构[10]。

全国首例 AI 生成声音人格权侵权案进一步表明,权利人对录音制品或音频素材享有、取得相关利用权,并不当然意味着其有权将配音者的声音进一步 AI 化。法院以“可识别性”为关键标准,认定在一般社会公众或者相关领域公众能够根据音色、语调和发音风格关联到特定自然人时, AI 生成声音仍可能落入其声音权益保护范围。因此,训练阶段至少需要分别核查声源主体的声音权益、个人信息处理基础以及训练素材所涉及的作品、表演和录音制品权利¹⁰。

4.2.2. 生成阶段:声音模拟与作品利用

内容生成阶段应根据具体技术实现判断权利利用范围。若 AI 翻唱直接以既有伴奏、原录音或受保护的词曲作品为生成基础,在不符合法定许可、合理使用等条件且未获必要许可时,可能分别涉及复制权、信息网络传播权、录音制作者权或表演者相关权益;若创作者以自行演唱的基础音轨进行音色转换,或者直接通过文本转语音生成新的声音内容,则对原录音和原表演的利用程度可能不同,其主要风险可能转向目标主体的声音人格利益。因而,仅凭“AI 音色替换”本身不能当然推导出对复制权、改编权等全部著作权权能的侵害,仍应对具体使用对象和行为方式逐项判断。对于合理使用,《著作权法》第 24 条采取法定情形列举,并要求相关使用不得影响作品的正常使用、不得合理损害著作权人的合法权益;所谓市场“实质性替代”可以作为评估正常使用和权利人利益影响的重要因素,但不宜被表述为我国著作权法上独立、唯一的法定判断标准。需要特别说明的是,《著作权法》第 42 条第 2 款关于录音的法定

¹⁰<https://www.court.gov.cn/zixun/xiangqing/467511.html>

许可具有特定适用条件：录音制作者使用他人已经合法录制为录音制品的音乐作品制作录音制品时，可以不经著作权人许可，但应当按照规定支付报酬，且著作权人声明不许使用的不得使用；该规则不能被泛化为 AI 翻唱的一般豁免。

4.2.3. 传播阶段：公众误认、商业利用与内容标识

传播阶段的法律风险主要表现为公众误认、商业利用与生成合成内容标识不足。博主标注“仅供娱乐”“非商业用途”并不能当然排除责任，是否侵权仍需结合声音是否具有可识别性、是否属于《民法典》第 1023 条第 2 款参照适用第 1020 条时可以不经同意的合理使用情形、是否使用受保护作品以及实际传播方式综合判断；平台流量分成、打赏、广告、直播带货等商业收益则可作为判断利用性质、主观过错及损害后果的重要因素。以特定自然人姓名或声音标识进行引流，还可能与姓名、名誉等人格权益发生竞合。

在 AI 内容标识方面，现行规则已经较原稿所依据的单一规范更为具体。《深度合成规定》第 17 条第 1 款第 2 项明确规定，深度合成服务提供者提供合成人声、仿声等语音生成或者显著改变个人身份特征的编辑服务，可能导致公众混淆或者误认的，应当进行显著标识。《标识办法》第 4 条第 2 项进一步规定，音频生成合成内容应在起始、末尾或者中间适当位置添加语音提示、音频节奏提示等显式标识，或者在交互场景界面中添加显著提示标识；第 5 条要求服务提供者在文件元数据中添加隐式标识；第 6 条规定网络信息内容传播服务提供者对生成合成内容的核验、提示和标识功能义务；第 10 条则要求用户发布生成合成内容时主动声明并使用服务提供者提供的标识功能。因此，对 AI 语音二创的传播治理不应只关注侵权发生后的删除赔偿，还应通过标识、来源识别、投诉处置等机制降低身份误认和风险扩散¹¹。

北京互联网法院课题组对全国首例 AI 生成声音人格权侵权案的梳理显示，该案从声音数据采集、数据处理、人工智能合成直至终端用户使用，完整呈现了 AI 生成声音的全链条关系[11]。因此，同一娱乐翻唱可能在训练阶段涉及声音与数据处理，在生成阶段涉及作品、表演或录音制品，在传播阶段又涉及商业利用和内容标识；同一种声音克隆技术用于新闻、广告或诈骗时，也会因使用目的不同而呈现不同的风险强度。以“场景 × 流程”的交叉判断作为分析方法，才能避免将技术类型直接等同于侵权类型。

5. 人工智能合成语音二创侵权的法律规制路径完善

5.1. 声音权益的体系化保护

我国未来制度完善的首要任务不在于另行创设一项独立的“声音权”，而在于现有法律体系内明确声音权益的规范定位与适用边界。关于声音应被理解为法定人格利益[5]还是具体人格权[6]，学界仍有不同解释。结合《民法典》第 1023 条第 2 款的“参照适用”规范结构，本文继续采用“声音权益”这一中性表述，并以可识别性作为人格利益保护的重要入口：当 AI 生成声音足以使一般社会公众或者相关领域公众将其与特定自然人建立稳定联系时，应纳入声音人格利益的评价范围；对于无法识别特定自然人的一般合成音色，则不宜仅凭“相似”当然启动声音人格利益保护，但如另行侵害名誉、隐私、个人信息或其他合法权益，仍应依相应规范处理。商业使用方面，应进一步区分声音录制、模型训练、内容生成、广告代言、数字人持续运营等不同授权目的和范围，并可参照《民法典》第 1021 条关于肖像许可使用合同的解释规则：对声音许可条款的理解存在争议时，结合第 1023 条第 2 款的参照适用，应作有利于声音权益主体的解释。救济方式也应根据侵害类型、传播范围和损害后果分别确定停止侵害、消除影响、赔礼道歉或损害赔偿等责任，而不宜将商业性作为声音权益成立的必要条件。

¹¹参见脚注 4 和 6。

5.2. 训练数据授权与合法利用机制的完善

训练数据治理不宜采取“一律授权”或“一律合理使用”的单一模式，而应根据训练对象、训练目的与所涉及法益实行分层规则。著作权维度上，李安主张区分非表达型与表达型机器学习，并按照大众表达、个体模仿与非营利科研等情形实行梯度安排^[13]；刘晓春则提出，对不针对特定作品、仅用于生产过程的生成式人工智能训练，可从“非作品性使用”角度讨论其合法性^[14]。这些观点说明，训练数据使用本身存在较大理论争议，不宜预设所有训练行为均须取得著作权许可。但 AI 声音克隆若直接以特定自然人为建模对象，则除版权问题外，还涉及声音人格利益与个人信息权益，其合法性不能仅依训练数据版权理论判断。

据此，可以建立三层合规机制：对于不指向特定自然人、仅用于一般模型能力训练的数据，应在合法来源、必要性、数据安全和权利保留机制基础上，为技术创新保留适度空间；对于以特定自然人可识别声音为直接训练对象的模型，应重点核查声音主体的许可范围，并依《个人信息保护法》判断个人信息处理基础，涉及敏感个人信息的还应履行更严格义务；对于训练素材中包含受《著作权法》保护的作品、表演或录音制品的，应另行审查相应知识产权许可或限制规则。同时，《暂行办法》第 7 条要求生成式人工智能服务提供者依法开展预训练、优化训练等训练数据处理活动，使用具有合法来源的数据和基础模型；涉及知识产权的，不得侵害他人依法享有的知识产权；涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其他情形。商业化语音模型提供者还应建立训练数据来源和授权记录、权利人查询投诉渠道及必要的退出、删除机制。对规模化使用作品和录音制品的情形，可以探索集体管理、标准化许可等降低交易成本的方式；但涉及特定自然人声音人格利益或依法需要个人同意的个人信息时，不宜以集体授权当然替代本人许可。

5.3. 平台算法治理机制

平台治理重点应前移至风险预防阶段，但责任配置应区分 AI 生成服务提供者与网络信息内容传播平台的不同角色。对于同时提供声音克隆、文本转语音等生成服务的主体，应重点落实训练数据管理、输入与输出审核、生成合成内容标识及投诉处置等义务；对于主要承担内容传播功能的平台，则应侧重识别生成合成属性、补充或提示标识、建立便捷投诉渠道并对明显冒充、诈骗等高风险内容快速处置。具体治理强度还应与前文场景分类相对应：一般娱乐性 AI 配音可侧重显著标识和权利人投诉机制；具有特定自然人可识别声音的翻唱和二创，应加强授权证明、重复侵权处置与权利人通知机制；新闻播报、公共信息传播应提高身份真实性和内容来源审核强度；商业数字人应重点核验声音授权目的、期限与商业使用范围；恶意冒充、诈骗等高风险场景则应设置更灵敏的识别、限制与报告机制。通过“主体角色 + 应用场景”的双重区分，可以避免把所有平台和所有 AI 语音内容适用同一治理标准。

5.4. 规制范式转型：从侵权救济走向风险治理

AI 语音侵权风险具有全过程扩散特征，在数据收集、模型训练、内容生成与平台传播各环节逐步累积。传统侵权救济主要依赖权利人在损害发生后主动维权，不仅举证成本较高，也难以应对 AI 语音内容快速复制和扩散的特点。因此，治理目标应从单纯的事后救济进一步转向分级、分类和全过程风险治理。具体而言，可以以“可识别程度 + 使用目的 + 传播影响”为核心建立风险分层：对于不指向特定自然人的一般合成语音和低风险技术应用，以基础标识、透明度和一般内容治理为主；对于能够识别特定自然人的娱乐性二创，提高授权提示、投诉处理与商业化审查强度；对于新闻播报、商业代言、数字人持续运营等可能强化身份误认或持续利用的场景，增加身份、授权和显著标识核验；对于以冒充、诈骗为目的的高风险使用，则通过功能限制、快速阻断、证据留存和依法报告等机制强化治理。风险分级的目

的不是简单扩大平台责任或压缩二创空间，而是在技术创新、人格保护、版权秩序和公共安全之间形成与风险程度相匹配的制度回应，推动治理模式从“侵权发生－个体维权”转向“风险识别－过程控制－责任追溯”。

6. 结语

生成式人工智能推动数字内容生产方式深刻变革，声音逐渐成为兼具身份识别、数据处理与商业开发属性的重要数字化人格要素。AI 合成语音二创在降低创作门槛、丰富数字文化供给的同时，也使人格权、著作权及邻接权、个人信息保护与平台治理规则在同一技术流程中交叉适用。本文在梳理国内相关研究争议的基础上，将 AI 合成语音所涉及的人格识别利益、数据利用利益与商业开发利益概括为“三阶法律结构”，并强调该结构是一种法益分析框架而非新的权利创设；在侵权认定层面，文章进一步以应用场景和技术流程为双重线索，分别考察娱乐翻唱、新闻播报、商业数字人、恶意冒充以及训练、生成、传播各阶段的主导风险；在治理层面，则主张在现有法律框架内通过声音权益体系化解释、训练数据分层治理、平台角色区分和风险分级机制，推动治理从单纯事后救济向全过程风险控制转型。本文的局限也较为明确：当前我国 AI 声音司法案例仍然有限，文章主要聚焦民事权益保护及与之直接相关的平台治理，对刑事责任、跨境数据流动、算法可解释性与具体技术标准未作系统展开。随着多模态人工智能进一步融合声音、肖像和动作等人格要素，数字人格的整体保护、授权边界与利益分配仍有待通过更多司法实践和跨学科研究继续检验。

参考文献

- [1] 田常清, 屈泳翰, 李采薇, 等. 新质生产力视角下 AI 合成语音二创短视频侵权困境与协同治理研究[J]. 出版科学, 2025, 33(6): 32-42.
- [2] 张炎坤. AI 语音合成技术在有声出版中的法律风险及其治理策略: 域外经验与中国方案[J]. 出版科学, 2025, 33(3): 27-38.
- [3] 王迁, 吴俊儒. 智媒时代 AI 表演的著作权法定性与保护[J]. 中国出版, 2026(4): 26-32.
- [4] 邹哲琦. AI 翻唱的著作权风险及规制路径[J]. 影视产业研究, 2025(3): 105-110.
- [5] 王利明. 论声音权益的法律保护模式[J]. 财经法学, 2024(1): 3-20.
- [6] 汪倪杰. 论声音权益的法律性质与保护路径[J]. 法学, 2024(10): 132-146.
- [7] 王利明. 论人格权商品化[J]. 法律科学(西北政法大学学报), 2013, 31(4): 54-61.
- [8] 房绍坤, 曹相见. 标表型人格权的构造与人格权商品化批判[J]. 中国社会科学, 2018(7): 139-162.
- [9] 孙雅琪. 智媒时代声音权益的法律保护: 困境成因与纾解之策[J]. 出版发行研究, 2026(2): 70, 71-78.
- [10] 陈杰. AI 表演的知识产权问题研究[J]. 知识产权, 2023(7): 56-75.
- [11] 北京互联网法院课题组. AI 生成声音侵害声音权益的法律认定——以殷某某诉北京某智能科技公司等人格权侵权案为例[J]. 法律适用, 2024(9): 123-133.
- [12] 陈笑春, 李泽林. 谁的声音: 生成式人工智能语境下声音的主体与权益[J]. 西南民族大学学报(人文社会科学版), 2025, 46(3): 141-150.
- [13] 李安. 机器学习的版权规则: 历史启示与当代方案[J]. 环球法律评论, 2023(6): 97-113.
- [14] 刘晓春. 生成式人工智能数据训练中的“非作品性使用”及其合法性证成[J]. 法学论坛, 2024, 39(3): 67-78.
- [15] 张学永, 周豪. “二创”短视频侵犯著作权问题实证分析及规制路径研究[J]. 新闻爱好者, 2024(12): 59-62.
- [16] 廖方菱, 陈蔓青, 陈胜湘, 等. 人工智能合成语音与自然语音的对比研究[J]. 中国司法鉴定, 2026(2): 38-45.
- [17] 陈志业, 张智骞, 王兵, 等. AI 语音合成技术的应用与展望[J]. 影视制作, 2023, 29(3): 51-55.
- [18] 筑牢人格权益保护“防火墙”——人民法院贯彻实施民法典强化人格权司法保护[EB/OL]. 2025-05-28. <https://www.court.gov.cn/zixun/xiangqing/466211.html>, 2026-08-22.