

智能语音合成技术创作物蕴含的法律风险与未来进路

朱影

(青岛科技大学传媒学院, 山东 青岛 266061)

摘要 根据技术生成方式的不同, 智能语音合成技术创作物可分为文本驱动型、语音克隆型、参数定制型、多模态融合型四类合成物, 不同类型的合成物创新了各行业的生态, 但在语音合成过程中, 对原声、内容素材和个人信息的采集引发了人格权、著作权与数据安全方面的风险, 为此可以通过“明确声音权法律地位、完善侵权救助机制”“界定著作权侵权行为、强化分发与传播管理”“筑牢数据安全防线、健全平台和政府治理体系”等措施进行治理, 为智能语音合成技术创作物的发展保驾护航。

关键词 智能语音合成技术; 有声阅读; 法律风险

智能语音合成技术已成为驱动相关行业生产模式变革的核心“引擎”, 但技术红利的快速释放也往往伴随着潜在的法律风险。2023年, 全国首例智能生成声音侵权案已经为智能语音合成技术的发展敲响警钟, 如何在技术变革与法律滞后之间实现平衡, 让智能语音合成技术既能为相关行业带来智能化的新质生产力, 又能避免陷入潜在法律风险的困境, 已成为当前亟待解决的问题。本文从智能语音合成技术创作物的概念与分类出发, 深入探讨其引发的法律风险并提出相应的治理方案。

1 智能语音合成技术创作物的概念与分类

智能语音合成技术是将文本信息转化为自然、流畅的语音信号的技术, 主要通过文本解析、声学建模、声码器生成及情感建模四个核心模块实现语言符号向自然语音的高效、精准的转换。当前智能语音合成技术有两种应用形式, 一是机械语音合成, 即基于文语转化 (Text to Speech, TTS) 技术直接将文字转换为语音; 二是模拟语音合成, 即利用自然语言处理技术学习特定人的音色、语调、语速甚至情感表达方式, 生成高度相似的合成语音。

智能语音合成技术创作物是人类在智能语音合成技术的协助下, 主导或参与生成的具有一定独创性的智力成果。智能语音合成技术创作物的本质上是技术与人类创意结合的产物, 所以判断某一语音内容是否为智能语音合成物需满足三个核心要素: 其一, 依托智能语音合成技术生成, 包括基于深度学习的文本转语音、语音克隆、情感语音生成等。其二, 人类对生成过程的创造性干预, 即在智能语音合成的过程中需由人类发挥主导或辅助作用, 而非完全由 AI 自主生成。其三, 最终成果需以音频形式存在且核心内容涵盖语言、语调、情感表达等语音要素。

根据技术生成方式的不同, 可将智能语音技术创作物分为四种类型: 一是文本驱动型合成物, 即通过文本转语音技术直接将文本生成对应语音。比如微信读书的听书功能可将电子书转化为音频并朗读。二是

作者简介: 朱影, 硕士研究生, 主要研究方向为数字出版、新媒体出版。E-mail: zhuying1609@163.com。

语音克隆型合成物，即以少量语音样本为训练模型，生成模仿特定人音色、语气、说话习惯的语音，也称为“声音克隆”。比如网络平台中走红的“让郭德纲狂飙英文”“用林俊杰的声音唱周杰伦的歌”等音频。三是参数定制型合成物，即通过调整音色、情感、语速、韵律等参数生成符合特定风格的语音，此种合成物不依赖固定文本或样本。比如喜马拉雅为恐怖故事定制的低沉、颤抖语音，为儿童早教内容定制活泼、夸张的卡通角色语音。四是多模态融合型合成物，即结合文本、音乐、环境音等多元素串联形成的复合音频作品。比如番茄畅听中 AI 语音演绎的广播剧，将角色对话、背景音乐、特殊音效都融入其中。

2 智能语音合成技术创作物蕴含的法律风险

智能语音合成技术为许多领域带来了创新性的探索实践，但其生成的创作物也引发了人格权、著作权等方面的法律争议，若缺乏系统性、可操作性的制度安排，势必对技术红利的持续释放形成掣肘，进而引发更普遍的法律难题与治理挑战。

2.1 原声采集引发人格权侵权风险

利用智能语音合成技术生成语音克隆型合成物进行录制时仅需一段“原声”，经模型学习与训练便可高度还原^[1]。《中华人民共和国民法典》第四编明确规定了生命权、身体权、肖像权、隐私权等一系列具体人格权，除具体人格权外，自然人也享有基于人身自由、人格尊严产生的其他人格权益。典型的比如声音权益，但现行法律框架中并未就“声音权”进行细化立法或作具体司法解释，这导致智能语音合成技术在有声阅读“仿声”录制中的应用处于灰色地带。北京互联网法院在全国首例人工智能生成声音人格权侵权案中，认定平台擅自克隆配音演员声音构成侵权，从图 1 可以看出，本案涉及多方主体，完整展现了 AI 语音合成从采集到市场端用户使用的上下游全链条。但网络公司与科技公司在本案中被判定为免责，这就给相关主体以“技术中立”或“用户默认许可”为由规避责任提供了参考，从而加剧了用户的维权难度^[2]。

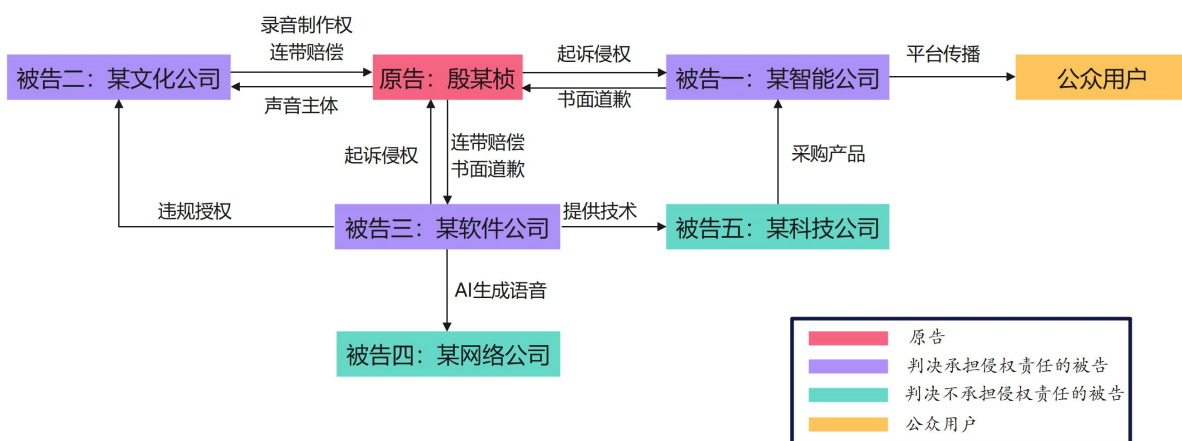


图 1 中国大陆首例人工智能生成声音人格权侵权案详解图

2.2 素材采集引发著作权侵权风险

复杂的制作与传播环节加上相关行业规范的缺失，使得智能语音合成创作物的著作权侵权问题较为严峻。

智能语音的合成离不开海量文字作品、录音制品等素材，在大规模商业化应用的背景下，常常出现平台或技术方未经权利人许可，便将受著作权保护的内容纳入模型训练库的情况。这种未经授权的不当使用行为直接触及《中华人民共和国著作权法》中有关复制权的规定，复制权包括以不同载体和表现形式再现作品，平台若对原文字作品进行了实质性复制，形成了可固定、可再现的复制件，且能基本“再现”原件内容，便侵犯了原作者的复制权^[3]。阅文集团 2025 年批量起诉喜马拉雅平台未经权利人许可便擅自将 19 部网文作品进行语音合成并供用户进行播放、下载，剥夺了著作权人的权益。此外，若平台将未经著作权人授权的智能语音合成物播放供用户收听，则侵犯了著作权人的信息网络传播权；若利用智能语音合成技术对原作品进行改编、翻译、注释、整理等演绎行为，生成新的语音作品而未获得原著作权人许可，会侵犯原作者的作品演绎权；在收集、整理语音数据时，如果将受保护的多个语音作品或片段进行汇编，形成新的语音集合体，且未获得相关著作权人许可，会侵犯著作权人的汇编权。中国裁判文书网显示，2024 年至 2025 年头部的有声阅读平台在智能语音合成领域面临的著作权案件数量并不少，具体如表 1 所示，可见平台侵犯作者著作权的现象已经屡见不鲜。

表 1 部分有声阅读平台 2024-2025 著作权类案件诉讼数据统计表

平台	著作权类案件数量	调解成功率
喜马拉雅 FM	355 则	35%
蜻蜓 FM	126 则	58%
荔枝 FM	89 则	62%

(注：数据来源：作者整理)

2.3 个人信息采集引发数据安全风险

语音数据是生成智能语音合成技术创作物的“燃料”，但数据的采集对用户的信息安全构成了多维度威胁与挑战，涉及个人隐私泄露、数据滥用、身份伪造等风险。

数据采集阶段，部分平台的隐私条款表述模糊，未明确告知用户语音数据去向，或强制用户授权语音信息用于“功能优化”，不同意就不能使用语音控制功能，涉嫌过度收集数据，违反了《中华人民共和国个人信息保护法》中“最小必要原则”和“明示告知—充分同意”要求，使得用户对于自身声音数据可能被二次利用的情况缺乏应有的知情权^[4]。一旦上述语音数据被用于深度伪造或诈骗冒用，权利人往往难以及时行使撤回或限制处理等权利。可以说，若数据采集之初缺乏明确边界和规范，后续更难以在监管或技术层面实现有效回溯与补救。更值得警惕的是，合成语音的“深度伪造”特性，进一步加剧了用户的数据安全风险。不法分子可伪造语音实施转账诈骗，近年来频发的“智能换声”诈骗事件表明，技术门槛因合成算法的进步而不断降低，使公民个人防范的难度急剧增大。与此同时，技术迭代速度远超监管规则的更新步伐，当前数据监管框架尚未完善、权责界定仍不清晰，相关平台在合成物传播中的审核责任也缺乏明确标准，从而导致大量违规合成内容仍在网络空间肆意流通。

3 智能语音合成技术创作物的未来进路

根据智能语音合成技术创作物可能蕴含的法律风险，可从明确声音权法律地位、界定著作权侵权行为、筑牢数据安全防线等方面探索其未来进路。

3.1 明确声音权法律地位，完善侵权救助机制

《中华人民共和国民法典》在第 1023 条第 2 款设立了法定类推规则，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”。据此，声音权保护可直接援引肖像权的相关规范，具体包括：有第 1019 条禁止他人擅自制作、使用、公开肖像权人肖像的规定，第 1020 条关于肖像权的保护规则，第 1021 条、第 1022 条关于肖像许可使用合同的规定。但智能语音合成技术对该条款以及《中华人民共和国反不正当竞争法》《中华人民共和国著作权法》《中华人民共和国个人信息保护法》等涉及声音人格权益法律法规的解释和使用都提出了挑战，诸如声音的权力主体、声音权益的可识别性判断等，所以明确声音权的独立法律地位尤为重要。2024 年 3 月，美国田纳西州修改《个人权利保护法》，将 AI 声音纳入个人形象权的保护范围。同时该法律也明确，声音权利属于财产性权利，可以转让、继承，不会因为自然人的死亡而终止^[5]。将“声音权”独立不仅能够更加精准、高效地处理相关案件，平台和技术提供方也可据此明确责任边界，减少侵权行为的发生。

此外，鉴于声音权侵权行为普遍存在隐蔽性特征且涉及较高技术壁垒，为了规避权利人在面临侵权行为时的发现滞后、举证能力不足等实践困境，相关政府机关应当完善声音权侵权的救济手段与维权机制。为提升权利救济的实效性，司法机关可依据《中华人民共和国民法典》第 997 条规定，构建更为灵活的禁令制度，在侵权风险被初步确认时，及时裁定中止侵权行为，以防止损害结果的进一步扩大。在损害赔偿机制层面，应建立多元考量的计算体系：一方面，综合评估被侵权声音的市场价值与权利人的社会知名度等因素，确保赔偿数额与侵权行为造成的实际损害相匹配；另一方面，针对损失难以量化的情形，应明确法定赔偿的适用标准与裁量边界，避免赔偿数额的随意性。此外，对于因侵权行为导致权利人遭受严重精神损害或声誉贬损的情形，应适度提高精神损害赔偿的额度，通过强化法律责任的惩戒功能，增强对潜在侵权行为的威慑力，从而为声音权提供更为周全的法律保障^[6]。

3.2 界定著作权侵权行为，强化分发与传播管控

合理运用智能语音合成技术，在强化人格权保护的同时，更应当落脚在著作权方面，为此，可以从立法与执法两个层面进行治理，从而化解著作权确权与侵权认定方面的争议。

首先，细化法律条文，清晰界定侵权行为。现行著作权法在面对智能语音合成技术带来的新问题时，存在一定的模糊性，所以应完善著作权的法律条文，明确未经授权将受著作权保护的作品用于智能语音合成技术的训练数据收集、模型训练以及产品应用等各个环节，均构成侵权。对于以“合理使用”为由进行抗辩的情形，应制定更具操作性的判断标准，充分考量使用行为的目的、性质、对作品市场价值的影响等因素。若使用行为具有商业营利性，且对原作品的正常使用或著作权人的合法利益造成实质性损害，则不应认定为合理使用。某些有声阅读平台在商业推广中，未经授权使用了大量知名作家的作品片段进行语音

合成演示，虽声称是为了展示技术功能，但因其商业目的明显且可能影响相关作品的市场销售，应被判定为侵权行为。

其次，摆脱执法困境，管控二次分发与碎片化传播。与纸质载体相比，智能语音合成物的碎片化传播与二次剪辑具有更低的技术门槛和更高的操作便捷性，这使得传统版权水印与登记机制的适用效能大幅衰减。对此，一方面可引入语音指纹等技术构建智能化溯源体系，通过对音频特征的精准提取与比对实现对侵权内容的快速识别与实时拦截，从而遏制规模化盗用及恶意的二次传播行为^[7]；另一方面，针对跨平台、跨区域侵权行为的蔓延态势，执法机关与行业协会应推动建立跨平台的协同治理机制，通过标准化的联动处置流程与证据共享机制提升维权响应效率，形成对二次及多次侵权行为的全链条防控。

3.3 筑牢数据安全防线，健全平台和政府治理体系

为保护用户数据安全，使得智能语音合成技术创作朝着更专业、智能的方向发展，应当构建用户数据安全治理体系，推动相关平台与政府成为语音数据保护与合成内容监管的核心，让智能语音合成技术在合法合规的前提下真正服务于社会进步。

首先，平台责任层面，相关平台在数据采集阶段应当强化合法授权与“最小必要原则”的落地执行，以清晰易懂的语言制定隐私条款，禁止以“功能优化”为由强制捆绑非必要授权。例如，将“语音数据用于算法训练”与“基础语音控制功能”拆分为独立授权项，用户可单独选择是否允许数据用于模型优化，且拒绝后不影响核心功能使用。同时建立数据采集备案制度，平台需向监管部门报备采集计划及安全保障措施，从源头防范过度收集。此外，为减少原始数据泄漏风险，应强制要求合成语音嵌入包含生成平台、时间及授权信息的数字水印，从而便于溯源追责。

其次，政府监管层面，需双管齐下筑牢语音技术应用的安全防线。一方面要加快完善监管体系，在现有生物信息保护框架中细化语音数据的特殊保护条款，明确声纹特征、完整语音片段等核心数据的采集边界，要求企业以“最小必要”为原则获取用户授权，严禁通过“默认勾选”“功能捆绑”等方式过度采集。同时需量化非法滥用的处罚标准，对倒卖语音数据、伪造身份实施诈骗等行为，按涉案金额和影响范围设定阶梯式罚款，并将情节严重者纳入行业黑名单，追究刑事责任。另一方面要牵头制定行业规范，压实平台对合成内容的审核责任，建立全流程管控机制，对模仿公众人物、政务人员等敏感场景的合成语音，必须经过人工复核。同时强制推行违规内容快速下架机制，明确从举报受理到内容移除的最长时限，对传播虚假信息、侵权诽谤的合成语音实行快速响应，并同步溯源创作者信息，在鼓励技术创新的同时守住安全底线。

四、结 语

智能语音合成技术正以突破性速度、深度、广度融入各行各业，其在为内容生产提供规模化、智能化革新路径的同时，亦引发了人格权保护、著作权侵权认定、数据安全等深层次法律挑战。展望未来，随着该技术多维度的持续拓展，现行法律制度需构建动态调整机制，建立更具预见性的规则体系。唯有在明确的法律框架与伦理边界内实现技术创新与法律规制的良性耦合，智能语音合成技术与相关行业方能在智能化转型浪潮中实现可持续发展与业态繁荣。

参考文献

- [1] 郝明英.人工智能语音合成有声书著作权保护研究[J].中国出版, 2023, (01): 55-59.
- [2] 北京互联网法院课题组.智能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.
- [3] 胡开忠, 江璐迪.生成式人工智能机器学习的版权争议及应对(上)[J].版权理论与实务, 2024 (9) : 44-53.
- [4] 林爱珺, 马瑞萍.人工智能时代声音权立法的前瞻性思考[J].青年记者, 2019 (34) : 72-73.
- [5] 王绍喜.《民法典》时代声音保护的解释与适用[J].法律适用, 2023 (6) : 35-44.
- [6] 任安麒.数字出版领域智能语言模型的应用、风险与治理: 基于 ChatGPT 技术特征的分析[J].出版科学, 2023, 31 (3) : 94-102.
- [7] 李秀娟.人工智能语音合成技术在有声读物制作中的应用[J].数字技术与应用, 2025, 43(04): 46-48.

The legal risks and future paths inherent in the creations of intelligent speech synthesis technology

ZHU Ying

(College of Communication, Qingdao University of Science and Technology, Qingdao Shandong 266061, China)

Abstract According to the different methods of technology generation, the creations of intelligent speech synthesis technology can be divided into four types: text driven, speech cloning, parameter customization, and multimodal fusion. Different types of creations have innovated the ecology of various industries. However, in the process of speech synthesis, the collection of original sound, content materials, and personal information has caused risks in terms of personality rights, copyright, and data security. Therefore, measures such as "clarifying the legal status of sound rights, improving infringement relief mechanisms", "defining copyright infringement behavior, strengthening distribution and dissemination management", "building a strong defense line for data security, and improving platforms and government governance systems" can be taken to govern the development of intelligent speech synthesis technology creations and safeguard them.

Keywords Intelligent speech synthesis technology; Audio reading; Legal risk