

人工智能语音合成有声书的著作权保护研究

徐英浩

澳门科技大学法学院, 澳门

收稿日期: 2025年4月10日; 录用日期: 2025年5月14日; 发布日期: 2025年5月23日

摘要

随着以人工智能技术为代表的数字技术的蓬勃发展, 融合了既有表现形式与新兴语音合成技术的人工智能语音合成有声书应运而生, 并对现有的著作权环境造成了一定破坏。本文以著作权保护为核心, 揭示语音合成技术在有声书领域造成的核心困境, 寻求可行的解决方法, 以期构建新旧技术和谐繁荣发展之前景。

关键词

人工智能, 语音合成, 有声书, 著作权, 网络服务提供者

Research on Copyright Protection for AI Speech Synthesized Audiobooks

Yinghao Xu

Faculty of Law, Macau University of Science and Technology, Macau

Received: Apr. 10th, 2025; accepted: May 14th, 2025; published: May 23rd, 2025

Abstract

With the vigorous development of digital technologies represented by artificial intelligence, AI-generated audiobooks that integrate traditional forms of expression with emerging speech synthesis technology have emerged. These have caused certain disruptions to the existing copyright environment. Centering on copyright protection, this article aims to reveal the core challenges posed by speech synthesis technology in the audiobook field, seek feasible solutions, and strive to build a future where old and new technologies can develop in harmony and prosperity.

Keywords

Artificial Intelligence, Speech Synthesis, Audiobook, Copyright, Internet Service Provider



1. 引言

随着互联网时代的蓬勃发展,人们进行娱乐消费的方式呈现出丰富化、科技化的趋势,随着文本阅读发展日渐繁荣,人们就提出了更进一步的视听享受要求,于是有声书(Audio books)形式应运而生并蓬勃发展。随着人工智能技术的不断发展,语音合成技术也逐渐进入到大众的生活,并为有声书带来了更进一步的上升空间,根据上海图书馆与樊登读书 App 发布的《2022 年阅读趋势研究报告》可知,有声书市场在 2016 年至 2010 年的 5 年间的增长率达到了 161%之多[1]。与此同时,著作权方面的纠纷也随之而来,2024 年世界范围内出现了多起针对人工智能合成技术的诉讼案件[2]。面对这种现实,我们可以通过研究人工智能语音合成技术的运行机理来深入探讨人工智能语音合成有声书的著作权所面临的核心困境以及针对其的保护路径。

2. AI 语音合成有声书及其产业链

在有声书的概念塑造方面,目前传播行业普遍比较认同的是美国音频出版协会提出的定义,即有声书是指文字内容占比不少于 51%、经由复制和包装后以成盒式磁带、高密度光盘以及单纯数字文件等储存方式进行销售的录音制品[3]。另有学者提出,有声书可以认定为以文字、图形等作品内容为基础的,对原内容不做更改或者稍加改编的,以音频放送为最终模式的一种网络出版物[4]。不论如何定义,有声书均具有三个组成部分,即文本内容、制作以及传播。

根据方式的不同,有声书的制作可以被分为两个部分,其一是提供声音的主体,其二是有声书的制作工艺。当提供声音的主体为自然人时,有声书采用的是人声朗读模式。当提供声音的主体为非自然人时,可以根据所使用技术的不同将有声书划分为机械音频转化模式和人工智能(Artificial Intelligence, AI)语音生成模式。音频转化模式是指通过 TTS (Text to Speech) 技术等将原文本的内容逐字逐句地进行朗读,从而使得作品内容从视觉层面进入到听觉层面,例如喜马拉雅、番茄免费小说等 App 推出的自动朗读、听书等功能[5]。AI 语音生成技术可以视为机械音频转化技术的进阶版本,提取自然人的声音,通过语音合成技术将所提取声音的元素打碎并合成全新的拟自然人的合成声音,或者直接提取出主要元素以模拟声音提供者的声音,再通过文语转化技术制作有声书。前者例如 2017 年新华社与北京搜狗科技发展有限公司联手推出的全球首个智能 AI 主持人“新小浩”,后者如利用语音合成技术对已故配音演员李易的声音进行复原并据此为纪录片《创新中国》进行旁白配音。与机械式文语转化相较,AI 生成的有声书更有感情、更富有温度,且效率更高,时间与金钱成本更低,可以为人们的阅读提供更为丰富多彩的选择。

语音合成技术诞生于多个学科的交叉,它通过分析人类语音的底层参数,如音高、音长、音强和音质等,结合语言学知识,将这些参数按照一定的规则进行组合,从而生成与人类语音相似的合成语音。在实际操作上,语音合成通常包括三个主要步骤:文本预处理、声学建模和波形合成。文本预处理阶段主要涉及对输入文本的分词、词性标注等任务;声学建模阶段则依赖于大量的语音数据训练;最后在波形合成阶段将声学特征转化为实际的语音波形,生成最终的合成语音。近些年来,随着深度神经网络研究的深入,合成语音的真实度得到了进一步提升。但正是由于语音合成技术的运行机理,使得数据大模型在研发过程中无法保证输入训练的文本以及声音的来源合法性,从而引发人工智能生成有声书独特的著作权困境。

3. AI 语音合成有声书的著作权保护的核心困境

AI 语音合成在世界范围内属于较为新型的技术门类，其应用边界不断扩张，技术手段不断革新，由此带来因规制之滞后性而约束不能所造成的种种困境，著作权保护不充分的问题也日趋严重。这其中虽有属于有声书的共性问题，例如授权侵权、平台侵权、取证困难等。但更多的是因技术本身的特点而产生的特性问题，即训练数据与生成内容的版权合法性的边界模糊。

3.1. 训练数据的版权授权争议

人工智能模型的训练依赖于海量文本数据，而这些数据所涉及的作品往往都处于版权保护的范畴。根据《中华人民共和国著作权法》(以下简称《著作权法》)第五十三条的规定，未经著作权人许可，复制、发行、表演、放映、广播、汇编、通过信息网络向公众传播其作品的，除本法另有规定的，应当根据情况，承担本法第五十二条规定的民事责任。在未经原作者授权的情况下，使用其作品进行 AI 模型训练可能被视为“非法复制”。自 2023 年《纽约时报》等多家媒体向纽约南区联邦法院提起诉讼，指控微软及 OpenAI 侵犯其版权之后，国外就出现了多起针对 AI 大模型训练数据的诉讼。我国目前没有针对大模型训练的有效法律手段，但民间的声音始终存在。2024 年中国文字著作权协会发出倡议，呼吁生成式人工智能语料库建设等应确保数据来源合法。随着人工智能等新技术的迅猛发展，版权界限的模糊态势随着作品范围的扩大而日渐扩展，传统的版权保护方式越来越难以应对新技术、新业态、新模式的挑战[6]。

3.2. 生成行为对原作品的潜在侵权风险

AI 合成的有声制品需要经过声学建模步骤，这一训练过程需要向数据库中输入足够数量的声音作品，以期形成逼真度较高的自然人声音，合成语调自然、表达流畅的声音内容，例如智能语音导航应用。这个过程可能涉及到对著作权中“表演权”的侵害[7]。根据《著作权法》第十条之规定，表演权即公开表演作品，以及用各种手段公开播送作品的表演的权利，有声书将文字内容用特定的情感表达，佐以声音形式向大众传播，笔者认为其构成表演，而非简单的复制。故而，在生成有声书的过程中，如果未经声音作品著作权人之认可，使用了相应的声音作品为合成行为，即造成对著作权人表演权的侵犯[8]。

此外，训练模型的过程还可能侵犯自然人的人格权。《民法典》第 1023 条第 2 款规定，对自然人声音的保护参照适用肖像权保护的规定[9]。2023 年，北京互联网法院审理了一起 AI 声音侵权案。从事声音工作的殷女士发现，一款名为“魔音工坊”的 App 擅自将她的声音进行 AI 化处理后供用户使用，并以此牟利，遂以侵害其声音权为由将有关若干家公司诉于法庭。本案的焦点在于 AI 声音是否应受到声音权益保护。法庭经审理认为，自然人以声纹、音色等特点可以与个人的行为与身份高度绑定，而该案中 AI 化后的声音仍可以关联到原告本人，最终判定原告胜诉。若 AI 合成的声音高度模仿特定自然人，则合成行为可视为对其声音权构成侵权[10]。

3.3. 生成内容的著作权归属不明确

AI 生成内容的著作权归属问题，本质上是一场关于“人类创造力”与“技术工具性”的认知博弈[11]。我国著作权以独创性作为认定之核心标准，在这种判断标准的指引下，作品的可版权性与自然人的智力贡献高度相关。然而，随着人工智能的技术迭代，AI 创作正朝着人机协作的方向进行发展，传统的独创性判断标准遭遇严峻挑战。2024 年常熟市法院判处了一起涉及人工智能生成内容的著作权侵权纠纷，案件的核心争议点在于，对于原告通过输入提示词等方式使用人工智能软件生成的图片，原告是否享有著作权。法院最终判决认为，案涉图体现了作者独特的选择与安排，具有独创性，构成著作权法意义上的美术作品，应受到著作权法保护。

以此案为例，法官通过“提示词调整”“参数筛选”等具体操作捕捉人类的创造性痕迹，但这种“行为主义”判定路径在 AI 有声书领域却面临双重困境：一方面，语音合成的技术黑箱导致用户操作与生成内容之间的因果关系难以追溯，独创性投入存在举证真空，这使得受害者在寻求损害赔偿时面临巨大困难；另一方面，法律尚未对“最低限度智力劳动”设定量化标准，个案裁量易引发“同案不同判”的司法乱象。这切实展现出工业时代的版权逻辑与数字时代技术范式的深层冲突——当 AI 从辅助工具演变为“创作伙伴”，人类在指令输入、数据喂养、结果遴选中的角色是否足以构成法律意义上的“作者”？

4. AI 语音合成有声书著作权保护路径探索

针对 AI 语音合成有声书面临的特性困境，笔者认为有必要从多种角度出发探究其著作权方向的保护路径。

4.1. 构建数据授权与合理使用制度

1) 明确数据训练的法律边界

面对 AI 训练数据的版权争议，我们可以建立灵活的法律框架来保证数据尽可能多的得到授权。首先，需划分训练数据的“合法获取”与“合理使用”的应用场景。例如，日本《著作权法》第三十条规定，允许非商业研究或小规模训练在支付合理补偿费后使用不超过原作 30% 的版权数据，但需确保生成内容不实质性替代原作市场价值；其次，需建立分层补偿机制，不同目的的主体适用不同的收费计算方式。促进教育事业发展的学术机构等公益组织可以适用最低费率，而企业研发则需按照标准费率支付版权费用，尽可能明确数据训练的法律边界，在技术发展与版权保护之间寻找平衡。

2) 推行数据授权集体管理机制

为了简化 AI 开发者获取海量数据的时间成本，我们可以考察并探索设立著作权集中许可制度的可行性与方式，由音著协、文著协等著作权集体管理组织统一代理版权授权以便提高作品的市场利用率，降低点对点对接授权的分散性所带来的资源消耗[12]。

4.2. 完善生成内容的版权认定与责任划分

1) 独创性标准的司法细化

AI 生成有声书内容是否属于“作品”呢？根据近期 AIGC 有关案件的汇总，我们不难看出，法律实务中判断 AIGC 是否属于作品主要从两方面进行判断，一是生成物是否满足独创性要求，二是生成物是否属于人类智慧成果。在“提示词 - 算法模型 - 生成结果 - 修改应用”这一行为链条中，如果人类的智慧与思考经由提示词与多次修改呈现于生成物中，则可视此生成物为作品。同时对 AI 生成内容实施“强制署名制度”，要求标注“AI 生成”及包括提示词设计者、参数调整者在内的主要贡献者，避免权利归属混乱。

2) 侵权责任的动态划分

若生成物存在著作权侵权风险，那么我们需要根据 AIGC 具体产生过程对侵权责任进行合理分配。

平台责任。若 AI 模型本身使用的训练数据即存在非法使用的侵害著作权人权益的情况，那么 AI 平台需要承担连带责任；同时，平台有义务设置内容过滤系统对侵权高风险的内容予以删除，从源头降低生成物侵权的风险概率[13]。

用户责任。如果用户有通过修改参数等操作诱导 AI 模型生产存在著作权侵权风险的内容，例如要求生成的有声书刻意模仿某位著名作家的文风或某位知名配音演员的声音，则用户需要承担直接的侵权责任[14]。

4.3. 技术手段与司法实践相结合

技术手段。《生成式人工智能服务管理暂行办法》第十二条规定：“提供者应当按照《互联网信息服务深度合成管理规定》对图片、视频等生成内容进行标识。”数字水印技术被认为是一种在图片、音频、视频等内容上进行标识的有效手段，在生成式内容的标识效果方面比较理想。在生成物产生时，人工智能会将其产生过程中所涉及到的各种信息汇总转化成标识信息，并以水印的方式嵌入到生成物的合适位置。生成物的使用者可以通过明水印确认生成物的性质，并可根据提取暗水印的方式进一步了解该生成物的具体内容。暗水印较为隐蔽，稳定性强，在生成物发生著作权纠纷时可以通过内容提取对生成物进行精确溯源^[15]。该技术应用于音频内容中，不仅可以对生成物的质量起到一定的规范作用，也可以有效保护生成物的知识产权，为侵权行为受害方提供积极救济。

司法层面。在 AI 生成内容侵权判定中，“实质性相似 + 接触”原则的适用需结合技术特性进行动态调整。该原则要求权利人证明侵权方存在接触原作品的可能性，且生成内容与原作构成实质性相似。然而，AI 技术的复杂性对传统判定标准提出了新的挑战。其一，“接触”的认定需突破物理载体限制，使用推定规则。若 AI 模型训练数据包含受版权保护的作品即可推定侵权人与原作品产生了技术性接触；其二，“实质性相似”的比对需分层处理。对风格模仿模式与元素复制模式应区别认定侵权程度。未来需通过司法解释明确“接触”认定的推定规则，并引入算法辅助进行相似性分析，同时要求平台自证训练数据的合法性，否则即要求其承担举证不利后果，以此达到平衡版权保护与技术创新的目的。

5. 结语

技术发展是一把双刃剑，在促进经济全面发展的同时也会带来难以预料的消极影响。在合成语音使用愈发广泛的当下，我们应当注意到对 AI 合成语音无约束的滥用所造成的著作权危机，以及由此引发的对市场秩序的破坏和对创作氛围的打击。《著作权法》第一条即规定保护著作权的目的是促进社会主义文化和科学事业的发展与繁荣^[16]。有声书作为精神文明的一种传播载体，同样受到人工智能技术的冲击。我们应该积极寻求既有形式与新技术的融合，在分析研究 AI 语音合成有声书著作权保护核心困境的基础上对漏洞加以填补，在保护著作权人法益的基础上激发创作者使用 AI 工具进行创新创作的积极性，为知识产权法律制度之完善和人类创作精神之繁荣延续添砖加瓦。

参考文献

- [1] 李秀丽. 移动互联时代有声书的开拓与走向[J]. 编辑之友, 2017(6): 19-22.
- [2] 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(1): 55-59.
- [3] 郭雪, 罗晓薇, 王壮. 近五年我国有声书研究文献综述——基于对 2017-2021 年八种出版核心期刊相关文献的分析[J]. 新闻文化建设, 2022(12): 60-62.
- [4] 王娟娟. 我国有声书发展的现状、困境与破局[J]. 科技与出版, 2021(12): 63-67.
- [5] 刘茜芸. 数字有声读物产业中的版权保护风险与应对研究[J]. 科技与出版, 2021(1): 123-129.
- [6] 焦和平, 梁龙坤. 人工智能合成音乐的著作权风险及其化解[J]. 知识产权, 2023(11): 103-125.
- [7] 王利明. 论声音权益的法律保护模式[J]. 财经法学, 2024(1): 3-20.
- [8] 陈杰. AI 表演的知识产权问题研究[J]. 知识产权, 2023(7): 56-75.
- [9] 李雅箐, 刘宇星. AIGC 技术赋能数字音频内容生产: 应用场景、存在问题与应对策略[J]. 数字出版研究, 2023, 2(3): 13-20.
- [10] 姜晓华. 声音的法律属性论争与证成——我国《民法典》第 1023 条第 2 款的法教义学分析[J]. 北方法学, 2022, 16(5): 70-80.
- [11] 廖斯. 论人工智能创作物的独创性构成与权利归属[J]. 西北民族大学学报(哲学社会科学版), 2020(2): 79-85.

- [12] 熊琦. 著作权集中许可机制的正当性与立法完善[J]. 法学, 2011(8): 101-110.
- [13] 祝建军. 传播有声小说侵害著作权的判定[J]. 人民司法, 2017(20): 90-92.
- [14] 吴汉东. 侵权责任法视野下的网络侵权责任解析[J]. 法商研究, 2010, 27(6): 28-31.
- [15] 朱红儒, 孙巍巍, 孙勇, 彭骏涛, 赵霁飞, 杨锐, 李世奇. 基于数字水印技术的生成式人工智能标识研究[J]. 中国信息安全, 2023(11): 67-70.
- [16] 朱景文. 人工智能时代有声书的版权保护研究[J]. 传播与版权, 2023(22): 112-116.