• 专题：AI 复活的伦理困境与善治路径 •

编者按：

以生成式人工智能、深度学习等为手段的 AI 复活技术正在刷新人类面对死亡的传统观念。AI 语音、AI 视频等数字手段在重构生者与逝者情感互动方式的同时，也面临文化适应性、情感依赖性、人格尊严与隐私保护等诸多伦理困境。本专题三篇文章从不同角度反思了 AI 复活的伦理困境，并寻求其善治之策。第一篇邱德胜与罗译泓的文章直面 AI 复活的关键技术——AI 语音合成技术，揭示其在促进人机情感互动、实现声音复活方面的作用，进而警示其可能引发的语音欺诈、情感依赖与自我认同混乱等问题，最后提出了涵盖技术生态、协同治理框架与预警机制的全方位治理路径。第二篇黄金泉的文章首先对 AI 复活在情感连接、文化传承与智慧延续等方面的社会需求进行了梳理，由此指出该技术在数据隐私保护、情感依赖、人格权益和生命尊严层面带来的伦理挑战，进而提出了对该技术进行负责任价值指引的应对之策。第三篇王振辉的文章创新性地从儒家伦理的视域审视了 AI 复活技术。文章认为 AI 复活应当融入丧祭仪式，既体现了对儒家传统文化智慧与人类情感表达的尊重，也为数字时代的丧祭礼仪变革提供了具有启发性的理论框架。三篇论文探讨的角度各有侧重，观点方法也不尽相同，但均为我们重新审视生命的意义和价值，反思 AI 复活的伦理与社会挑战提供了有益借鉴。

（专题策划：邱德胜）

AI语音合成技术：问题类型与治理路径

Problem Types and Governance Pathways of AI Voice Synthesis Technology

邱德胜 /QIU Desheng[1,2]　　罗译泓 /LUO Yihong[1]

（1. 西南大学哲学系，重庆，400715；2. 西南大学科技伦理治理研究中心，重庆，400715）
(1. Department of Philosophy, Southwest University, Chongqing, 400715;
2. Research Center of Science and Technology Ethics Governance, Southwest University, Chongqing, 400715)

摘　要：近年来，AI语音合成技术已被广泛应用于有声读物、游戏配音、歌曲制作及人机交互等多个领域，深刻塑造并改变了人的听觉体验。其逼真性、互动性与定制化等技术特征促使个体在身体、社会与自我三个维度保持在场，在优化沉浸体验、促进人机互动与声音美化等方面发挥了巨大作用。然而，这些特征也引发了语音欺诈、过度情感依赖与自我认同混乱等伦理与社会问题。为有效应对挑战，亟须

基金项目：重庆市教委人文社会科学研究项目“人工智能伦理治理研究”（项目编号：24KSGH338）；西南大学研究阐释二十届三中全会精神专项项目“生成式人工智能发展规律与管理机制研究”（项目编号：SWU2509306）；西南大学创新研究2035先导计划（项目编号：SWUPilotPlan018）。
收稿日期：2025年4月9日
作者简介：邱德胜（1975–）男，湖北武汉人，西南大学哲学系教授，西南大学科技伦理治理研究中心研究员，研究方向为科学技术与社会、科技伦理学。Email: dsqiu@swu.edu.cn
罗译泓（1996–）男，重庆人，西南大学哲学系博士研究生，研究方向为科技伦理学。Email: longyihong314@qq.com

对AI语音技术在研发、验证与应用等各阶段进行全方位伦理审视，从发展敏捷治理的AI语音技术生态、构建多元主体共治共享的协同框架、推进AI语音技术的动态预警机制出发，引导其择善而行。

关键词：AI语音合成　在场　逼真性　互动性　定制化

Abstract: In recent years, AI speech synthesis technology has been widely applied in multiple fields such as audio books, game dubbing, song production, and human-computer interaction, profoundly shaping and transforming people's auditory experiences. Its technical features such as realism, interactivity and customization enable individuals to remain present in the physical, social and self dimensions, playing a significant role in optimizing immersive experiences, promoting human-computer interaction, and enhancing sound aesthetics. However, these characteristics have also given rise to ethical and social issues such as voice fraud, excessive emotional dependence and confusion of self-identity. To effectively address the challenges, it is urgent to conduct a comprehensive ethical review of AI voice technology at all stages, including research and development, verification, and application. Efforts should focus on developing an agile governance ecosystem for AI voice technology, building a collaborative framework for multi-stakeholder co-governance and sharing, and establishing a dynamic early-warning mechanism to guide its ethical and beneficial development.

Key Words: AI voice synthesis; Presence; Realism; Interactivity; Customization

中图分类号：B80; TN912.33　DOI: 10.15994/j.1000–0763.2026.01.001　CSTR: 32281.14.jdn.2026.01.001

近年来，基于深度学习的语音合成技术展现出广阔的应用前景，被广泛应用于有声读物、游戏配音、歌曲制作、情感疗愈等场景，引发了越来越多的关注。该技术在显著提升人类听觉体验的同时，其“类真性”风险也使之颇受争议。已有研究揭示了AI语音技术可能导致的金融诈骗、隐私侵犯、信息误导、责任归属不清等风险。[1]-[3]但既有研究多聚焦于风险现象的表层梳理与归纳，尚未深入探讨其背后的生成逻辑与内在动因。声音不仅是信息传递的基本载体，亦是人类情感联结与社会文化建构的重要媒介。在此意义上，AI语音技术对声音生成效果的重构，与“在场”（presence）这一哲学、传播学领域的重要概念密切相关，成为理解该技术社会影响的关键理论视角。本文基于韩国著名学者李宽珉（Kwan Min Lee）对在场概念的三种分类——身体在场（physical presence）、社会在场（social presence）、自我在场（self presence），[4]尝试将AI语音合成技术产生的作用和引发的主要问题区分为三种类型，进而从理论层面厘清其风险的生成逻辑。在此基础上，选择与之相适应的治理路径，稳步推进AI语音合成技术的伦理善治，实现技术理性与人文价值的有效融合。

一、AI语音的逼真性：身体在场与语音欺诈

以高度模仿他人声音为目标的AI语音克隆技术有着难以分辨的逼真性。谷歌公司于2018年发布的第二代Tacotron模型，该模型的语音合成质量在人类的主观评价中已经能够达到“以假乱真”的地步了。微软亚洲研究院于2020年发布的Fastspeech2语音架构又更进一步，在持续时间、音高、能量等指标的细粒度控制以及合成推断速度上都取得了重大突破。此类基于深度学习的合成语音已经逐渐摆脱了“机械感”与“冰冷感”，在语速、发音、节奏等方面已与真人语音无异，普通人已经难以分辨“被克隆的声音”和“真实的声音”。[5]英国伦敦大学举行了一个让数百人来分辨语音片段真实性的实验，在事先明确告知听众他们所听的内容可能是AI伪造的情况下，仍然有30%的人无法分辨出真假。全球最大的安全技术公司迈克菲的一项调查也显示，77%的受害者在面对AI语音克隆的信息时都无法准确辨别真伪。

当AI语音以其难以分辨的逼真性与沉浸感充斥人的声音环境时，将唤起并保持个体的身

体在场。李宽珉指出，身体在场发生于用户未能识别技术中介对象或环境的拟真性质与人工性质时。[4]马修·隆巴德（Matthew Lombard）等更是明确指出“在场”就是一种逼真感（realism）与多感官的沉浸感（immersion）。[6]AI语音技术通过生成高度拟真的声音内容，让用户“声临其境”。在无需原有发声者在场的前提下，构建出足以唤起并维系个体具身性在场的虚拟声音景观。例如，基于深度神经网络的歌唱语音合成系统能够精准复现目标音色进行歌唱演绎并还原演唱时的情境，甚至生成原作者从未演唱过的全新歌曲。在邓丽君逝世27周年之际，酷狗音乐的阿波罗实验室就利用AI歌声合成技术再现了她的天籁之声，以邓丽君的声线与风格演唱全新歌曲《没有寄出的信》，让听者仿佛置身于真实的演唱会现场；在有声图书的朗读中，AI语音合成技术可以高度还原作者乃至小说中人物的声音，让听者“声临其境”地进入了书中的真实世界。其本质是让用户遗忘了技术中介（如录音机）的存在而产生了真实的临场感，从而促进了用户的沉浸体验。

AI语音技术有着双重解耦能力：一是解耦声源主体，无需发声主体的存在，便可制造出发声者从未言说的话语；二是解耦声音内容与原始语境，为生成的语音附加任意虚拟的环境音效（如嘈杂的车站）与情绪色彩（如急切的求救）。这标志着“声音–身体”同一性的瓦解，即声音彻底脱离其具身性源头与原始语境。历史地看，前工业时代的人际交流依赖共时共地的身体在场。而工业革命后的媒介技术——电话、广播、互联网等虽延伸了声音的时空可达性，但也带来了听觉空间中身体的离场与沉浸感的消退。此外，媒介技术并不能离开发声者而凭空存在，也不能剥离发声的情境，其本质只是一种声音的再现。例如，MP3录音机仅能存储与回放既存声迹，无法创造非实存之声或脱离原始语境的声景。但AI语音技术的出现克服了传统媒介下的“主体限制”与“时空限制”，通过建构高度逼真与沉浸的虚拟声景让个体置身于特定的叙事空间，从而实现身体在场。

然而，这种技术赋能的“身体递归”亦产生了多重伦理与社会风险，近年来影响极大的“AI语音诈骗”就是其重要代表。2019年，首例经媒体广泛报道的AI语音诈骗案中，不法分子成功模仿德国某能源公司CEO的声音，骗取22万欧元。仅在2023年，国内黑灰产业利用AI声音伪造等方式进行欺诈造成的经济损失就高达1149亿元。个体为什么易受AI语音技术的欺诈？技术现象学家唐·伊德（Don Ihde）在《听觉与声音》一书中指出，听觉维度（auditory dimension）从一开始就表现出全身性体验的普遍特征。从现象学上讲，人不仅仅是用耳朵这一个器官在听，而是用整个身体在感知。[7]声音景观以其固有的包围性与环绕性——伊德喻之为区别于视觉“前向锥形场”的“沉浸式球状场域”——作用于听者，引发身体的感官共振。换言之，AI深度伪造让声音能够从所有方向包围与欺诈听者，形成一种带有空间感的沉浸式听觉场。“如果我在一个听觉效果很好的礼堂里听到贝多芬的第九交响曲，我突然发现自己沉浸在四周的声音中。这种声音如此具有穿透力（penetrating），我的整个身体都在回荡……听觉场围绕着听者，而环绕性（surroundability）是声音场域的一个基本特征。”[7]不法分子通过AI语音技术模拟被害人亲朋好友的声音，让虚假的声景以“沉浸式球状场域”充斥被害者的全身感官。受害者沉浸于技术构建的逼真叙事场，其身体在具有“逼真性”与“沉浸感”的虚拟声音情境中被诱导实现非理性确信。

在典型的语音诈骗模式中，诈骗者通过生成高度逼真且蕴含特定情绪的伪造语音，诱导受害者产生情感焦虑与认知偏差，进而做出非理性决策。AI技术所营造的沉浸式声景不仅有着足以混淆真实性的声学仿真，还具有足够的互动性：在诸多语音诈骗案例中，被模仿者往往都能够对答如流，因而诱使聆听者更加信任听感的真实性并维持身体在场。受害者通过身体的听觉经验触摸到了AI声音所置身的那个“彼岸世界”，其身体维度的在场感被技术性劫持，让听者放下了应有的理性与戒备。

此现象可置于更广阔的听觉文化研究与权力理论框架下审视其伦理风险。汤姆·赖斯

（Tom Rice）曾借用福柯在《惩罚与规训》中讨论的“全景敞视监狱”，而提出了“全景声狱”（Panaudicon）的概念：当人被环境中多种多样的、不合时宜的噪声所包围，便构成了一种全方位的听觉监狱。[8] 当人际交往的声音环境被大量难以溯源的、具有欺骗性的AI生成语音即新型“噪音”所渗透时，社会个体被迫陷入一种普遍化的听觉怀疑状态。这种由技术催生的不确定性，侵蚀了人际互动中以声音作为真实性担保的认知基础。当难辨真假的“AI噪音”充斥人与人的交往环境，人们必然以一种怀疑一切的警惕态度来面对外在世界，整个社会将处于一种人人自危、互不信任的紧张状态，在引发公共信任危机的同时加剧人际关系的疏离。

二、AI语音的互动性：社会在场与情感依赖

人机交互是AI语音技术的一个重要应用领域，无论是苹果的“Siri”、小米的“小爱同学”还是AI聊天机器人，都有AI语音合成技术的身影。从二维、抽象的文字交流，到三维空间内富有情感的“听觉互动”，AI语音合成技术的特殊性在于其通过“声音”这一媒介极大地强化了机器作为陪伴者的情绪表达能力。在非面对面交流的媒介时代，个体感受的社会在场感的强弱取决于媒介技术所提供的亲密性与即时性。但是，传统的合成语音过于生硬且互动性不足，易让听众产生较强的疏离感与违和感。机械声音强烈的“冰冷感”依然让人能够随时认识到是在“聆听机器”，这使得人们更愿意将其看作工具、助手，而不是能够分享心事或值得信赖的挚友。

李宽珉认为，当技术用户成功地模仿其他人类或非人类智能时，社会存在就会发生。[4]

或者说，社会在场发生于个体将虚拟的社会存在感知为真实的社会行动者之时。社会在场可以理解为一种人间世的“烟火气”——一种来自同类的、日常的生存性安全感。[9] 因此，尽管社会在场的定义与理解不尽相同，但情感层面的感知则始终重要。微软发布的VALL-E、Meta AI发布的Voicebox等语音大模型仅需要极少的原始语音片段，就能完全模仿他人的语气、语调乃至语言习惯。最重要的是，其还能以具有“同理心”的方式感知说话者的声音情境与情感波动，进而提升使用者社会在场的心理感受。因此，相比于传统机械发音的人机交互场景，AI语音技术不仅仅是一个声音信息的传递工具，而是能够在人机互动中提供多种“情绪价值”与“烟火气”的重要手段。

声音能够负载人丰富的情绪，是个体感知他人情绪与形成自我体验的重要媒介。AI语音依托于数据、算力与算法等优势，为解决传统人机互动中情感表露的不足提供了可能。例如，个体在聆听导航语音时，熟悉的人声会让我们感到温暖和亲近，缓解迷路时的迷茫和无助；在考试失败等情绪低落的场景中，来自亲人或爱慕对象的声音可以给个体带来情绪上的满足感与安全感。他们将基于AI语音的虚拟陪伴者感知为具有社会性的交互对象，由此唤起了“社会在场”的认知体验。[10] AI语音克隆技术不仅能够精准复现特定个体（如亲友或倾慕对象）的声学特征，还能够依据用户的偏好与交互需求，动态调整语气、语调。AI语音以“平静”为零点，能够模拟人类的情绪语气——如快乐、温柔、宠溺、喜爱、惊讶乃至愤慨——从而提高聊天机器人模拟复杂情感表达的能力。

然而，这种具有高度互动性与沉浸感的语音交互并非止步于功能性的情绪支持，而是可能诱发用户形成习惯性且难以剥离的过度情感依赖，削弱个体在真实人际关系中处理冲突、维系情感纽带的能力。麦金泰尔认为，人类对他人的依赖是一个无法回避的事实，“承认依赖性是走向独立性的关键”。[11] 社会性与依赖性是人的重要特征，人在社会中总是需要依赖他人而存在。然而在社会在场中，AI语音被感知为真实的社会行动者，使用者可能陷入一种过度依赖的状态：他们依赖AI语音提供的即时、定制化的情绪回应，并将其视为可靠的情感出口，甚至在现实社交受挫或孤独感加剧时更倾向于向其寻求慰藉乃至心理“治疗”。这种情感上的依赖虽能短期内缓解部分心理压力，但

长期的过度使用将减少个体现实生活中的社会交往需要，降低现实感，淡化对人际关系和同理心的真实感知。[12]

相关研究指出，即便与人互动交流的是一台机器，对亲密度、情感卷入度等维度上的技术革新，亦有可能让使用者将这台机器感知为一个有生命的交流对象。[13]AI语音陪伴所营造的“在场幻象”，实则是将人类最深层的情感需求——被倾听、被理解、被持续回应——转化为可计算、可优化的机械参数，从而使得情感本身被降格为一种可被算法化处理的“数据资源”。约瑟夫·维森鲍姆（Joseph Weizenbaum）曾警告称，认为机器具备真实的情感功能是一种十分危险的错觉。智能系统并非情感主体，也不具有真实的情感能力。将情感互动的价值赋予机器，实则是一种虚伪的情感欺诈。AI语音的情感陪伴机制不仅模糊了真实人际交往与虚拟交往之间的界限，还使得人们放弃在充满不确定性与困难的真实情感关系中寻求承认与成长的努力，转而沉溺于一种无风险、无摩擦但也无真实社会体验的情感消费之中。正如雪莉·特克尔（Sherry Turkle）所言，“我们对科技的期待越来越多，对彼此的期待却越来越少。”[14]当AI语音合成技术通过克隆真人语音来满足使用者的情感陪伴诉求时，每个人的精神世界就只需要作为技术物的虚拟语音而不需要现实的“人”，现实的人对于人而言将成为冗余物，人再无意愿也无兴趣与现实中的人对话并逐渐对现实的人类陪伴产生排斥心理，进而产生过度依赖引发的“交往茧房”现象。

三、AI语音的定制化：自我在场与自我认同

自我在场是指个体对自身存在的直接、即时且非对象化的意识体验。它强调人在实践中保持自我同一性的能力，即“我”作为意识主体始终在场于自身的体验中，而非被客体化或碎片化。李宽珉认为，当技术用户没有感知到虚拟环境中构建的另一个自我的虚拟性时，自我在场就会发生。[4]换言之，自我在场可以理解为个体通过感官或非感官方式，将虚拟对象体验为真实对象的心理状态。Voicebox、ElevenLabs等AI语音大模型具有多样化的语音采样与生成能力，其能够提供个性化、定制化的语音服务来为用户创造一个属于自身的“虚拟声音”。当个体使用AI语音来复制、美化乃至定制自身的虚拟声音并忽视了这种虚拟性时，自我在场便会产生。

AI语音技术产生的自我在场，可以显著提升用户的主观感受并带来一定程度的自我满足：在玩耍电子游戏时，将用户的声音虚拟为游戏主角的声音，可以显著提升玩耍的沉浸感；在观看影视剧时，粉丝可以借助语音技术模仿偶像声音，以此获得一定的自我满足；而在工作、恋爱等社会交往情境，个体可以使用AI语音技术美化与修饰自身声音，从而提升自我表现的效果与增强社会优势。美国得克萨斯大学的一项研究显示：外表漂亮的人在收入上更有优势。声音亦是如此，“洋洋盈耳”“声如洪钟”等成语，以及“如风铃一般清脆”的文学描绘，都体现了人们对悦耳声音的向往。声音好听的人会在工作、学习上具有一种潜在的优势：拥有温柔声线的心理医生更易获得患者信任；声音洪亮有力的演讲者则能更有效地调动听众情绪等等。因此，AI语音技术高度定制化的特征能够让使用者依据不同情境与主体需求“随意修改”自身声音，塑造了足够真实且定制化的自我在场体验。

然而，这种技术赋权使得声音可塑性的边界被极大拓展，构成了对自我认同的潜在挑战。李宽珉指出，当个体长期处于技术媒介所营造的虚拟体验中时，强烈的自我在场感可能会产生某种类型的现实身份混乱，进而引发“身体图式的扭曲”（distortions in body schema）。[2]当AI语音技术使用户能够自由塑造甚至彻底改变其听觉身份时——使用者将虚拟自我（声音）体验为真实自我，“我是谁”这一基本命题是否也随之动摇？声音作为自我表达的重要载体，不仅仅是单纯的信息传递，更承载着个体的社会身份、情感特质与文化归属。当AI语音技术允许个体轻易地将原生声音替换为更符

合“理想自我”的合成音色时，个体将逐渐疏远甚至否定自己原本的身体经验，进而陷入一种“声格分裂”的困境。

这种自我认同的混乱尤其体现在身体异化后的社会互动与适应中。安东尼·吉登斯（Anthony Giddens）认为，作为行动系统与实践模式的“身体”嵌入日常生活的社会互动，是维持连贯的自我认同感的基本途径。[15] 若将声音视为一种身体，那么AI语音技术对声音的美化与定制实则是一场对“身体”的介入与重构。每个人喜欢的声音各不相同，声音美学的多元性本是社会包容的象征。但在自我在场的诱惑下，那个被技术美化后的“我”在社交中获得满足与认可，而那个真实的“我”的声音却被视为需要被修正或隐藏的累赘。长期使用AI虚拟声音可能导致个体对真实自我的接纳能力下降，其真实的、带有个人生活印记的原生声纹便被有意无意地边缘化了，乃至产生一种“声音焦虑”。因此，使用AI语音技术定制自我声音这种看似自由的选择，实则将自身置于技术与他者的审视之下，个体在不断调整自身声音以符合外部期待的过程中迷失了原本的自我。

面对AI语音技术的不同应用场景，我们理应持有不同的态度。例如，有些人因为某次意外而失去发声功能或发声功能受损。AI语音算法则通过学习其原始声音数据建立永久的个人声音库，并采取语音实时转换、佩戴微型发声单元等手段，让患者能够重新拥有正常的语音交流能力。人们在情感上能够普遍接受AI语音用于发声修复是因为它体现了人的“祛弱权”，其可视为一种在受伤条件下以恢复原状为目的修复或治疗行为。而因“工作优势”对声音系统的修改或美化，则更像一种否定自身后的“增强”，其无法诠释过度定制化后引发的自我认同困境。此项技术最根本的使命，不在于合成完美的声音以遮蔽人的本真存在，而在于重建声音的具身锚点，使AI语音成为人之声音的延伸，而非其替代。

四、AI语音合成技术的治理路径

在有关声音伦理的探讨中，一个核心的出发点就是围绕发声者与聆听者之间的互动关系与责任关系的深入研究。[16] 相较聆听者，发声者享有更多的权利，同时也需承担更多的责任。AI语音技术及其使用者作为新的发声者，对自身的发声方式、效果与内容负有重要责任。这种责任体现为对作为聆听者的人的尊重和关照，即任何的发声行为都应基于平等意义上的信息共享与情感交流，而非一种建立在自我利益基础上的诱导、操纵甚至控制。例如，使用AI语音合成技术的聊天机器人就不能以让用户沉迷或成瘾为设计目的，使用AI语音克隆技术的音视频就不能以欺骗聆听者为目的，而应该注重发声者与聆听者之间的良性互动。综上所述，有必要根据人类声音景观的“时空德性”与AI语音的技术特点、交互场景等特征，从算法治理、公众科普、制度创新等维度出发，构建一个技术自治与制度治理精准衔接的伦理治理框架。

1. 添加数字印记，发展敏捷治理的AI语音技术生态

AI语音技术在设计之初的目的，就是为了更好地模仿真人的声音，以降低需要人声配音场景的成本，抑或弥补声音无法再现的缺憾，因此，其设计一开始就是向着准度更高、更像真人、更难以分辨的目标前进的。AI语音合成的相似度一旦越过“奇点”，甚至连家人与亲朋挚友都无法分辨声音真假时，便有可能出现不可逆的消极后果。就像假钞厂生产的假钞连银行都无法分辨时，必将产生剧烈的社会冲击。因此，有必要通过相应的算法、程序或代码的嵌入为AI语音作品添加“数字水印”，增强使用者识别语音真实性的能力。

数字水印技术作为一种非易失性标记，能够在音频内容中嵌入隐蔽的标识信息，使得伪造内容在传播过程中容易被识别和追溯，为打击深度伪造行为提供强有力的技术支持。例如勒·马修（Le Matthew）等人设计的卷积二元分类模型，可用于区分真实世界的语音和AI合成的语音。[17] 该模型的应用可以从算法设计层面为AI音频添加一种可追溯的数字水印，减少

高仿真语音的滥用风险。通过添加数字印记，使用者可以在面对AI语音生成的作品时像微信扫码一样，通过“扫一扫”的方式来识别一段语音是否是AI生成的。技术人员还应不断优化检测模型，提升对AI克隆语音源的鉴定能力。[18]

除了强化AI语音算法的可识别性与可检验性，技术治理的敏捷性还体现在两个方面。一是算法工作人员应根据AI语音多元化的用户情况与娱乐、治疗等不同应用场景进行“个性化设计”。对于应用在聊天机器人、AI复活等场景的语音算法，应采用价值敏感的设计方法在算法层面进行拟真度限制，保持合理的技术韧性，防止用户形成过度的依赖心理。二是算法工作人员应努力确保语音数据来源的多元性与全面性。AI语音大模型依赖大量的原始声音素材进行驱动，并通过针对特定数据集的深度学习训练以实现声音特征的塑造。当训练数据存在显著质量缺陷时，可能导致最终生成的语音内容出现不可控的歧视性偏见或刻板印象等问题。个人偏好的声音可能源于亲密关系或童年记忆中的特定人物，设计者可能有意或无意地将这些声音偏好内化于算法设计之中，从而在不经意间引入了潜在的文化偏置。因此，设计人员在设计和训练语音合成算法时，应充分收集、考量各种文化背景下的人声数据，避免可能存在的数据偏见。

2. 提升媒介素养，构建多元主体共治共享的协同框架

AI语音合成技术的应用、推广与风险治理需要一个公众广泛参与的决策议程。STS研究表明，让缺少相应专长的民众来参与专业技术决策容易导致科学民粹主义与“无知者的暴政”，进而陷入棘手的延伸性困境。[19]针对这一难题，菲利普·基切尔（Philip Kitcher）提出了一种良序科学（well-order science）的协商机制：拥有不同偏好的协商者像茶余饭后的“家庭聚会”一般聚到一起，共同去了解相关的技术细节与他人的偏好，最终做出相应决策。[20]

良序科学理论蕴含三个核心理念：充分代表、理想协商、专家辅导。第一，充分代表要求所有与AI语音技术相关的研发者、使用者、管理者都需要通过代议制的方式参加到协商中来表达自身偏好。第二，理想协商要求参与者相互解释自己的偏好以及偏好的强烈程度。协商者承诺一视同仁地看待其他人的偏好，并能够设身处地理解他人的困境。例如，配音工作者要向其他协商者解释AI语音技术引发的失业风险，而患有先天性发声缺陷的人也要充分陈述自身面临的特殊脆弱性以及对AI语音技术的渴望。第三，专家辅导要求拥有相关专业知识的技术工作者向公众科普AI语音技术的发展前景、技术原理与潜在风险。

基于良序科学的AI语音技术治理既充分尊重了公众意见，又通过有效的信息传播提升了公众认知，进而避免了“无知者暴政”的产生。中国信息通信研究院发布的《人工智能伦理治理研究报告》（2023）就指出应提升各主体的人工智能伦理风险应对能力：通过小说、影视剧、短视频等多种形式引导社会公众认识科技风险，倡导合理使用生成式人工智能等工具。因此，一方面，公众应主动了解AI语音生成的基本原理，积极反馈AI语音的使用经验，并对新闻媒体中的音频内容持有合理的质疑；另一方面，政府部门则应整合科普教育、企业活动、新闻传播等多方面资源提升公众的人工智能商（AIQ），[21]增强公众尤其是中老年群体对AI合成语音内容的辨识能力，倡导“多源验证”的信息处理习惯，为推动多元主体共治共享的协同框架的形成奠定素养基础。

3. 打通治理壁垒，推进AI语音技术的动态预警机制

AI语音合成技术具备成长性、新颖性与不确定性等特征，其在商业领域的巨大潜力已然是不争的事实，但其潜在的社会风险并不明朗，存在多种路径下的不确定性。因此，在技术治理中应坚持预警原则（precautionary principle）以应对其不确定性的价值冲突。预警原则提倡一个积极、民主、开放的原则——不是消极的应付，而是主动寻求更加完善的决策方式和标准。[22]

基于预警原则的AI语音伦理治理应采取以下治理策略：第一，倡导审慎发展，鉴于AI语

音合成技术可能存在的成瘾性、类真性等风险，其发展应受到合理的限制；第二，重视风险与收益的权衡，AI语音在情感治疗、发声修复等方面都有重要的价值，应立足善治基准，采取“谨慎推进”的策略；第三，推崇事前预防胜于事后补救的理念，治理者应制定有关AI音频技术的行业准则与法律法规，明确声音权的法律边界并细化民法典中声音权的适用范围，制定声音侵权的认定标准与赔偿机制，将技术风险防患于未然；第四，坚持安全优先的原则，防止伦理风险的“消极性误判”，应根据AI语音的不同应用场景建立分级分类的管理体系，限制如新闻、医疗等敏感场景的应用；第五，建立积极、动态的技术审查和反馈机制。相关部门应该对AI语音的发展和应用情况予以实时监测，为相关伦理问题的解决提供恰当的缓冲程序，避免技术失控。预警原则下的技术治理应立足共享、共治的理念，明确AI音频企业、学术科研机构、行业协会以及用户等多元利益相关者的责任分工，构建起一套能在技术的提供方、应用方与使用者之间有效共享信息、经验与责任的共治体系。

结　语

AI语音合成技术使人的声音突破了时空限制，从听觉维度强化了人的“在场”体验，其作用包括塑造沉浸式的声音体验、促进人机情感交流、满足个体定制化的声音需求以及帮助聋哑人“重获新声”等等。然而，AI语音对个体听音环境的颠覆式改变，将产生语音欺诈、过度情感依赖与自我认同混乱等伦理与社会问题。如何规避其可能造成的不利影响，激发出对人类有利的一面？重要的是采取一种审慎、进取的立场，将哲学、传播学、声学等跨学科研究整合为兼具技术理性与文化价值的听觉文化分析工具，并根据AI语音合成技术的不同应用场景制定出切实可行的治理策略。

声音不仅仅是一个简单的数字信号，更承载了人丰富的情感依托，蕴含了多重的“生命寓意”与“人文价值”。乔尔·贝克曼（Joel Beckerman）指出，“每个声音都可能唤醒一段记忆，引发一连串的情绪反应，甚至可能在一瞬间左右我们的选择，改变我们的情绪。”[23]

因此，AI语音合成技术的伦理治理应该从人声的情感负载与社会价值出发，将被动的伦理诉求转化变为主动的价值嵌入，确保AI算法在生成语音的过程中遵循人类的道德和伦理标准，实现人与机器在发声者与聆听者角色上的价值对齐。以此将关涉AI语音的伦理原则落实到具体的制度与行动方面，避免技术发展脱离安全边界。政府部门应联合多方治理主体共同营造一个具有德性的“声”态环境，在法治和德治的双重保障下为AI语音合成技术的良性发展提供顶层导向与持续动力，以确保其能够为人类带来更多的福祉。

［参 考 文 献］

[1] Caldwell, E. 'Advances in AI-based Voice Synthesis'[J]. *International Journal of Artificial Intelligence and Machine Learning*, 2014, 14(1): 381–388.

[2] Hutiri, W., Papakyriakopoulos, O., Xiang, A. 'Not My Voice! A Taxonomy of Ethical and Safety Harms of Speech Generators'[A], Binns, R. (Ed.) *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*[C], New York: Association for Computing Machinery, 2024, 359–376.

[3] 季亚茹、张丽. 人工智能时代生成式AI语音伦理问题研究[J]. 视听，2024,（7）：151–154.

[4] Lee, K. M. 'Presence, Explicated'[J]. *Communication Theory*, 2004, 14(1): 27–50.

[5] Buteau, E., Lee, J. 'Hey Alexa, Why do We Use Voice Assistants? The Driving Factors of Voice Assistant Technology Use'[J]. *Communication Research Reports*, 2021, 38(5): 336–345.

[6] Lombard, M., Ditton, T. 'At the Heart of It All: The Concept of Presence'[J]. *Journal of Computer-Mediated Communication*, 1997, 3(2): JCMC321.

[7] Ihde, D. *Listening and Voice: Phenomenologies of Sound*[M]. Second Edition, Albany: State University of New York Press, 2007, 44–76.

[8] Rice, T. 'Soundselves: An Acoustemology of Sound and Self in the Edinburgh Royal Infirmary'[J]. *Anthropology Today*, 2003, 19(4): 4–9.

[9] 邓建国. 我们何以身临其境？——人机传播中社会在

场感的建构与挑战 [J]. 新闻与写作, 2022,（10）: 17–28.
[10] 唐瑞蔓、冯博博. 聆听AI: 智能语音的声音景观考察 [J]. 西南民族大学学报（人文社会科学版）, 2025, 46（3）: 151–158.
[11] 麦金泰尔. 依赖性的动物: 人类为什么需要德性 [M]. 刘玮 译, 南京: 译林出版社, 2013, 70.
[12] Shimada, K. 'The Role of Artificial Intelligence in Mental Health'[J]. *AI and Psychology*, 2023, 43(5): 1119–1127.
[13] 王天娇. 社会临场理论的三个内生性问题 [J]. 国际新闻界, 2011, 33（6）: 46–51.
[14] 雪梨·特克尔. 群体性孤独 [M]. 周奎、刘菁荆 译, 杭州: 浙江人民出版社, 2018, 297.
[15] 安东尼·吉登斯. 现代性与自我认同 [M]. 赵旭东、方文 译, 北京: 三联书店, 1998, 111.
[16] 孙琦、李雪枫. 时空德性: 声音景观的伦理规约 [J]. 编辑之友, 2021,（5）: 69–75.
[17] Le, M., Apoorv, V., Bowen, S., et al. 'Voicebox: Text-guided Multilingual Universal Speech Generation at Scale'[J]. *Advances in Neural Information Processing Systems*, 2023, 2306(15687): 1–30.
[18] 王学光、诸珺文、张爱新. 一种三维度基于改进MFCC特征模型的AI克隆语音源鉴定方法 [J]. 计算机科学, 2023, 50（11）: 177–184.
[19] 邱德胜、罗译泓. 科技决策的延伸性问题何以解决？——兼议柯林斯的专长决策理论 [J]. 自然辩证法研究, 2023, 39（2）: 101–107.
[20] Kitcher, P. *Science in a Democratic Society*[M]. New York: Prometheus Books, 2011, 114–115.
[21] 解学芳、陈思函. 基于深度合成技术的文化科技伦理风险识别与善治机制 [J]. 南昌大学学报（人文社会科学版）, 2024, 55（4）: 53–65.
[22] Tickner, J., Raffensperger, C., Myers, N. *The Precautionary Principle in Action: A Handbook*[M]. Windsor ND: Science and Environmental Health Network, 1999, 4–5.
[23] 乔尔·贝克曼、泰勒·格雷. 音爆: 声音的场景影响力 [M]. 郭雪 译, 北京: 北京联合出版公司, 2016, 9.

[责任编辑　李斌]

JDN