深度合成技术治理：基于弱势群体“双重效应”的滞后性破解研究

吴佳妮

西安翻译学院人文与艺术学部 陕西 西安 710105

【摘　要】：老年人与残障人士等弱势群体在深度合成技术普及中面临“赋权不足”与“风险放大”的双重困境，而治理体系的滞后性进一步加剧了其风险暴露与边缘化处境。研究聚焦弱势群体，通过梳理双重效应的现实表征，从法规、技术与协同三个层面剖析治理滞后的加剧机制，并借鉴国际经验，提出构建以“精准适配”为导向的法规补位、技术适配与协同赋权三维治理体系，以推动治理与群体需求动态匹配，缓解其结构性困境。

【关键词】：深度合成技术；弱势群体；技术赋权；精准适配

DOI:10.12417/3041-0630.26.07.078

1 核心概念界定与现状呈现

1.1 核心概念界定

2022 年发布的《互联网信息服务深度合成管理规定》，将深度合成技术定义为借助深度学习、虚拟现实等生成合成类算法，制作文本、图像、音频、视频及虚拟场景等各类信息的技术形态，其最核心的特征是能生成高度真实、肉眼难以辨别的虚拟内容。从研究范围来看，既包括生成式 AI 视频、语音合成，也涵盖文本生成、图像合成等技术形态，这些技术打破了传统内容创作的时空限制，却也因技术和应用门槛的差异，对不同群体产生了差异化影响。

在深度合成技术的社会影响下，弱势群体有着特定的内涵，特指在技术的接触、使用与风险抵御过程中，因自身能力、资源或环境限制处于相对弱势的群体。具体可分为三类，第一类是生理特征受限的群体，比如老年人因年龄增长出现认知能力衰退、技术适应能力弱的问题，视障、听障等残障群体在接触技术时需要依赖辅助形式，却在信息真伪辨别上存在障碍；第二类是数字素养较低的群体，像低收入群体、教育水平薄弱的群体，因缺乏系统的网络使用培训或没有接触网络的能力，在信息辨别上存在明显短板；第三类是信息获取渠道单一的群体，例如偏远地区居民，受地域限制难以接触外界多元信息，对深度合成技术生成的内容易产生认知依赖。这类群体的共同特点是，既难以享受深度合成技术的“技术红利”，又极易成为技术风险的承受者。

1.2 弱势群体双重效应失衡的现状呈现

深度合成技术与弱势群体的互动过程中，形成了技术赋权与风险加剧的“双重效应”，本应是相辅相成的两面，却在现实中出现了严重的失衡问题。技术赋权本是希望通过技术突破弱势群体的生理局限或环境障碍，比如用手语转换、语音合成等功能，帮助听障、视障群体便捷获取信息、参与社会活动，缩小与普通群体的差距；但风险加剧又因为技术被滥用，使得弱势群体更容易成为受害者，比如遭遇伪造语音诈骗、被虚假医疗内容误导等。这种失衡，本质上是技术的赋能价值未能充分落地，而风险防护机制又未能及时跟上，二者形成的难以调和的矛盾，这并非技术发展的必然结果，而是技术分配不公、管理防护体系存在漏洞共同导致。

深度合成技术的赋权价值已有实际落地的案例，在听障群体服务领域，AI 手语视频技术成为信息无障碍建设的重要突破，例如上海嘉定区行政服务中心推出的数字手语人智能服务平台，能通过屏幕手语演示告知听障用户业务办理流程，配合无障碍导航系统实现全流程指引，改变了以往低效的办事模式；甘肃省兰州市、金昌市等多地政务大厅也引入智能手语翻译终端，通过实时转换和远程翻译功能，确保政策解读、办事指南传递零误差。在教育场景中，南京聋人学校借助 AI 工具制作手语数字人辅助教学，配合 AR 眼镜与智慧黑板的语音转文字功能，大幅提升了听障学生接收信息的准确性。针对视障群体，语音合成技术构建起“听觉信息通道”，打破了阅读限制，浙江特殊教育职业学院的 EagleMovie 智能系统，还能将电影画面转化为带情感的解说语音，让视障学生能通过“听”感知光影世界，实现与普通学生的情感同频。

虽然这些应用极大突破了弱势群体在使用智能系统时遭遇的局限，但这些应用仍存弊端，例如工具适配性与场景兼容

作者简介：吴佳妮（1999-）女，汉，陕西省西安市人，硕士研究生，西安翻译学院人文与艺术学部，助教，研究方向：媒介文化研究、文化传播。

项目信息：本论文由西安翻译学院 2026 年度校级科研项目资助，项目名称：《数字向善：深度合成技术与陕西弱势群体权益保护研究》（2026B12）。

性不足，多数 AI 手语系统仅支持标准普通话转化，对地方方言、专业术语的识别准确率不高，在医疗问诊、法律咨询等专业场景中难以发挥作用；语音合成技术则普遍存在“机械音”问题，长时间聆听易引发听觉疲劳，对表格、公式等结构化信息的识别能力也较弱。其次是覆盖范围存在显著落差，以陕西为例，农村地区的 AI 服务覆盖率低但老龄化率高，如安康镇坪县老龄化超全国均值 8 个百分点，72%老年人深陷“数字生存困境”。这也让深度合成技术的赋权效应呈现出“精英化”倾向，没能真正触达最需要帮助的弱势群体。

与赋权侧发展受限形成对比的是深度合成技术风险正集中爆发。语音克隆技术的门槛不断降低，不法分子仅用 15 秒语音片段，就能以低至 5.9 元的成本生成高度逼真的模拟声线，诱导老年人转账汇款。2024 年 11 月湖北孝感就发生了一起 AI 语音诈骗案，不法分子克隆“孙子”的声线骗取老人 2 万元，类似的案例屡见不鲜。2025 年公安部数据显示，第一季度全国 AI 换脸、拟声诈骗案件数量环比激增 45%，其中老年群体受骗占比达 38%，足以见得老年群体已成为技术诈骗的重灾区。

慢性病患者也是技术滥用的主要侵害对象，不法分子通过 AI 技术伪造权威医生形象、合成虚假诊疗视频，将普通食品、保健品包装成包治百病的“神药”，甚至篡改临床治疗方案，误导患者停用正规药物，造成多起病情恶化的案例。2024 年视频号“华仔集结号”通过克隆刘德华声音促销保健品，3 天内销量暴涨 217%，78%的购买者因信任伪造的名人推荐下单，其中不乏大量需要规范用药的慢性病患者。治理体系在技术迭代面前的应对迟缓，让风险防控始终处于被动状态，也进一步加剧了双重效应的失衡程度。

2 治理滞后性对双重效应的加剧机制分析

弱势群体面临的深度合成技术双重效应失衡，核心症结在于法规及技术治理体系不尽完善。法规层面主要体现在弱势群体保护条款的缺失与责任划分的模糊。国内现行的《生成式人工智能服务管理暂行办法》，虽聚焦服务提供者的内容审核、数据安全责任，但并未明确平台对弱势群体的额外注意义务，比如是否需要针对老年人放大合成内容的警示语、增加语音提示。同时责任划分存在明显断层，当残障群体比如视障者因辅助工具缺失而无法识别合成视频的水印标识时，现有条款无法界定“平台未提供适配防护”与“用户认知不足”的责任边界，这也让弱势群体维权时缺乏明确的法律依据，进一步放大了风险效应。

技术治理的滞后，则表现为防护工具的通用化与弱势群体需求的适配性失衡。从技术供给端来看，微软 Video Authenticator、百度文心一言深度合成检测模块等主流深度伪造检测工具，多面向媒体、企业等专业场景，操作门槛较高，老年群体因数字技能不足难以完成复杂操作；且这些工具多以视觉交互为主，未能适配视障群体的信息获取需求，让防护技术对这一群体形同虚设。从技术应用端来看，现有检测工具主要针对高清晰度、完整时长的深度合成内容，而针对弱势群体的诈骗内容，多是低清晰度、碎片化的语音片段，检测准确率大幅下降；同时，这类风险多发生在家庭群、社交软件等私人场景，而现有防护工具多部署于公共平台，私人场景的防护空白让风险无法被及时阻断。技术供给与弱势群体需求的脱节，让他们既无法通过防护工具规避风险，也难以充分享受技术的赋权价值。

3 国际经验借鉴：兼顾弱势群体的治理实践与启示

深度合成技术席卷全球，因此欧盟、美国等其他国家的治理探索，也为我国破解治理滞后性提供了多元的参考思路。

2024 年 8 月 1 日正式生效的欧盟《人工智能法案》，将对弱势群体的针对性保护纳入核心监管框架。法案将利用自然人因年龄、残障等弱点扭曲行为造成伤害的 AI 系统，归为“不可接受风险”类别并全面禁止，违规者将面临高额罚款，直接遏制了针对弱势群体的技术剥削。在风险管控环节，法案要求高风险 AI 系统部署者开展基本权利影响评估，专项分析对老年人、残障群体的潜在危害，并制定相应的缓解方案，未通过评估的技术将被禁止进入欧盟市场。在合规义务方面，法案要求深度合成技术生成的内容添加显著标识，且说明需清晰易懂，同时强制高风险 AI 系统提供适配残障群体的可访问性使用说明。这套监管框架采用欧盟协调、成员国落地的模式，通过分级监管让对弱势群体的保护要求真正落到实处。

美国则选择依托现有民权法规，结合行业实践，重点解决老年人、残障群体面临的深度合成技术风险。在监管层面，联邦贸易委员会加强了对误导老年人的深度合成诈骗广告的执法力度，对相关企业处以高额罚款，并要求平台为面向老年群体的深度合成内容增设醒目警示。在技术研发层面，企业与残障组织的协作更为紧密，谷歌 2019 年启动的 Project Euphonia 项目，联合肌萎缩侧索硬化症患者组织采集独特语音数据，开发适配该群体的深度合成语音技术，同时简化操作界面，满足认知障碍用户的使用需求。在跨境执法方面，2025 年 6 月微软联合日本 JC3、印度 CBI 破获针对日本老年人的跨国 AI 诈骗案，精准打击了自动化的诈骗犯罪网络，既从源头减少了诈骗触达老年人的可能也增强了该群体对 AI 诈骗的防范意识。

欧盟与美国的实践表明，深度合成技术的管控不能依赖一刀切的通用规则，关键是要构建以弱势群体需求为核心的“精准适配”治理体系，将需求前置到治理的全流程，实现治理与群体需求的动态匹配。

4 治理滞后的破解路径：基于“精准适配”的三维构建

针对深度合成技术治理滞后性的核心痛点，需从法规、技术、协同三个维度构建“精准适配”的治理体系，通过专项规制、靶向技术、多元联动，实现治理措施与弱势群体需求的动态匹配，从根本上破解深度合成技术治理滞后的问题。

法规层面的补位，需要从构建以弱势群体为导向的专项规制体系出发，针对现有法规的泛化性缺陷，通过增设专项条款、细化责任边界、配套实施标准，形成“评估、责任、执行”的闭环。具体来说在涉及深度合成技术的产品上线前，要求技术开发者提交专项评估报告，明确不同弱势群体面临的独特风险，以及可验证的风险缓解方案，比如面向视障群体的应用需附加音频水印技术的第三方测试报告，同时建立评估动态更新机制，要求开发者每半年根据技术迭代与群体反馈补充评估，未完成更新或评估不合格的产品暂停服务，避免“一次评估终身适用”的问题。此外要差异化责任分配机制，根据弱势群体的能力差异，划分开发者、平台、服务机构三方的责任。开发者需预留辅助技术接口，比如面向听障群体的平台必须嵌入手语翻译接口；平台需承担“额外注意义务”，比如面向老年人的内容采用大字体与方言语音双模态警示，未履行义务造成损害的需承担主要责任；社区养老、残障服务机构则需协助开展风险告知工作，形成完整的责任链条。

技术层面的适配，关键是打造弱势群体可及的防护工具与联动网络，弥合弱势群体用不了、不会用防护工具的差距。例如面向老年人，可联合运营商研发内置预警功能的老年机，自动识别疑似深度合成的语音，通过实体按键、方言语音触发预警，同时开发支持语音指令的“一键核查 APP”，核查结果以方言语音播报；面向残障群体，则由残联牵头联合企业开发适配工具，为听障群体开发“手语溯源插件”并嵌入主流手语平台，为视障群体开发兼容屏幕阅读器的“音频水印识别 APP”。

在此基础上，还需社区和平台机构进行协同与兜底。在社区层面，于养老服务中心、残障人服务站设立防护工具帮扶点，配备志愿者协助群体安装并进行操作演示，将工具使用指南纳入老年课堂、残障人技能培训。平台机构可以与残联、老龄办达成合作，在应用界面设置用户反馈通道，每月收集弱势群体的使用问题并优先迭代，每季度联合残联、老龄办开展适配测试，邀请不少于 50 名弱势群体代表参与，确保迭代方向贴合实际需求。

5 结语

技术越是追求通用与高效，其治理反而越需要差异与温度。研究通过分析深度合成技术对弱势群体产生的“赋权—风险”双重效应，明确了当前治理体系需构建以“法规、技术与协同”为导向的三维治理体系，推动治理逻辑从被动响应到主动适配。未来研究可进一步开展分群体、分地域的微观调查与实证检验，推动治理设计向差异化、精细化发展。

参考文献：

[1] 宋保振.“数字弱势群体”权利及其法治化保障[J].法律科学(西北政法大学学报),2020,38(06):53-64.

[2] 王也.数字鸿沟与数字弱势群体的国家保护[J].比较法研究,2023,(05):121-137.

[3] 陈琦,闫小童.弥合数字残疾沟:视听障碍群体数字化生存的困境与突破[J].现代传播(中国传媒大学学报),2023,45(09):134-139.

[4] 潘祥辉,李东晓.视听障碍人群信息汲取的传播环境:一个文献综述[J].重庆社会科学,2011,(09):76-81.

[5] 张赐琪.消弭数字鸿沟:美国弱势群体信息权利保护的理论与实践[J].毛泽东邓小平理论研究,2012,(04):98-103+116.